Ethics without errors: Universal Moral Error Theory and the objection from loss

Ethics Without Errors: Universal Moral Error Theory and the Objection From Loss

Andrew Jonathan Dines

Submitted for the degree of PhD in Philosophy

Birkbeck College, University of London

I hereby declare that the work presented in this thesis is entirely my own.

Abstract

According to Universal Moral Error Theory (UMET), all moral propositions are false – or, on some formulations of the theory, lack a truth value. In recent years, something of a niche literature has sprung up, which addresses the question of what do with moral discourse if UMET is true. This is the "What Next?" Question, also known as the "Now What?" Problem. The "Now What?" Problem looks to be hypothetical in structure: either UMET is true or it isn't, and if it is true, then a workable solution to the "Now What?" Problem would appear to be desirable. However, this hypothetical structure is somewhat misleading: some proposed solutions to the "Now What?" Problem, I argue, are of dubious consistency with their own error theoretical motivations, and this points towards a powerful objection to moral error theory. So the "Now What?" Problem really is a *problem* for UMET, and its difficulty has been underestimated by some commentators.

The objection to which all this points is the objection from loss. This objection is essentially is that ethical discourse is practically indispensable, and for that reason cannot be targeted by an error theory of universal scope. This is different to David Enoch's indispensability argument, which aims to demonstrate the existence of robustly real metanormative properties. The objection from loss does not establish realism. All it establishes is that UMET is false. The success of the objection from loss is to be explained by a tension within the error theory to which the shortcomings of some proposed solutions to the "Now What?" Problem point – namely, the tension between the need for UMET-proponents to convict morality a particular concrete error (or a collection of concrete errors), and the presentation of this version of moral error theory as having universal scope.

Contents

Abbreviations Used in This Thesis

AfO: Argument from Obscurity

AfR: Argument from Reasons

EPM: The Epistemic Parity Version of the Moral Moorean Argument against UMET

GNET: Global Normative Error Theory

G-NWP: The version of the 'Now What?' Problem arising from GNET

LMET: Local Moral Error Theory

L-NWP: The version of the 'Now What?' Problem arising from an LMET

MET: Moral Error Theory (of unspecified strength, such that UMET and any given LMET are both instances of MET)

NICR: Non-institutional categorical reason

NWP: 'Now What?' Problem (also known as: 'What Next?' Question)

SEW: Scepticism about the External World

UMET: Universal Moral Error Theory

U-NWP: The version of the 'Now What?' Problem arising from UMET

Introduction

§0.

According to universal moral error theory (henceforth: UMET), our moral thought and talk is systematically infected with error. Furthermore, according to this theory, the error is so pervasive that we involve ourselves in error when we make any moral assertion whatsoever, even one as uncontroversial as "it is morally wrong to torture human babies merely for fun". Recently a niche literature has sprung up concerning the problem that I will call the "Now What?" Problem for UMET (henceforth: U-NWP). This is the problem of what to do with moral discourse if UMET is true. The literature has the appearance of an in-house dispute between error theorists: many of the contributors are UMET-proponents engaged in the project of formulating policy proposals concerning how to proceed now that the error involved in moral discourse has come to light.

In this thesis, I will at times engage with the U-NWP literature. However, this will not be in the service of trying to identify which of the various policy proposals offered in the literature are the most promising. Instead, my engagement with the U-NWP literature will help me to make an objection to UMET. This is the objection from loss. The objection from loss, in its simplest form, is as follows.

> **Objection from Loss – Simplest Version**
> [1] If UMET is true, then there is some sense in which it is illegitimate to adopt an ethical point of view.
> [2] It is not the case that there is some sense in which it is illegitimate to adopt an ethical point of view.
> [3] Therefore: UMET is not true.

It should be noted that this 'Simplest Version' is *deceptively* simple. In particular, premise 2 can afford to be so strong – claiming not just that *there is some sense in which it is legitimate* to adopt an ethical point of view but that *there is no sense in which it is illegitimate* to do so – because of the specific ways that the terms "moral" and "ethical" will be used in this thesis (see §1.1.4.1).

The reason why engaging with the U-NWP will help me press this objection is that the contributors to the U-NWP literature appear to endorse the first premise whilst assuming (on the basis of UMET) that the second premise is false. Thus, I share with contributors to the U-NWP literature a view that UMET, if true, potentially comes with serious consequences for first order ethical thought and talk. I, however, think that gaining a clear idea of what those consequences would be shows that UMET cannot be quite correct, and I think this sometimes shows up when looking at the details of U-NWP policy proposals.

Relatedly, I think that when we are alive to the variety of possible ways in which we can adopt an ethical point of view, then some U-NWP policy proposals are of questionable compatibility with their own error-theoretical motivations. This is certainly the case with the two U-NWP policy proposals I discuss in detail, abolitionism and substitutionalism. My discussion of these two policy proposals thus sets the stage for the objection from loss, which I press at length in the final chapter.

Here, then, is how I proceed in what follows. In the first chapter, I discuss arguments for UMET. What I am to establish is that UMET is worth taking seriously. I do this by discussing two major strands of suspicion we might have about morality, namely, the obscurity of the notion of moral obligation and the phenomenon of persistent disagreement concerning foundational propositions about morality. In the second chapter, I bolster the claim that UMET is worth taking seriously by rebutting what might initially seem a good objection to it: that it implausibly implies that even those moral propositions we take to be obviously true, such as that it is morally wrong to torture human infants merely for fun, are in fact false.

Next, I set about inquiring as to the practical consequences of adopting UMET. As mentioned above, this takes me into the realms of considering a couple of U-NWP policy proposals. Abolitionism is considered and rejected in Chapter 3. Substitutionalism is considered and rejected in Chapter 4. In both cases, the basis on which the policy proposal is rejected is that the proposal is of dubious compatibility with its own error-theoretic motivations. The discussion in Chapters 3 and 4 sets the stage for the discussion of the argument against UMET in Chapter 5. Here, the objection from loss is pressed.

In sum, then, this thesis presents an argument against UMET, and discusses some U-NWP policy proposals *en route* in a way that lays the foundations for the argument against UMET. The overall aim is not to solve U-NWP, but to dissolve it.

Chapter 1:
Arguments for UMET

§1.0

Different forms of moral error theory (henceforth: MET) are possible (Lillehammer 2004, 2013). A distinction that will be important to everything to come in this thesis is that between UMET and various forms of local moral error theory (henceforth: LMET). UMET claims that all moral propositions are false – or, in some formulations, lack a truth value. LMETs are versions of MET that claim only that a particular region (or a limited number of particular regions) of what I shall call (in §1.1.5 below) the ethical web are infected with error. For instance, a thoroughgoing utilitarian might adopt an LMET targeting the notion of moral rights.

This thesis is an extended argument against UMET. There would be no point devoting an entire thesis to an argument against a view that had nothing going for it in the first place. Does UMET initially score enough "plausibility points" for such a project to be worthwhile?[1] I think so – although the occurrence of the word "initially" in the previous sentence is crucial. I am fairly confident that UMET is false, and my confidence is based entirely on the objection to UMET that I will pressing over the course of this thesis. So, in the final analysis, I don't think that UMET scores very well on plausibility points at all. I think nearly every other metaethical theory scores better. Nonetheless, in this first chapter, I will be arguing for the following claim:

> **Plausibility:** There are arguments that together present a strong *initial* case for UMET. These arguments have sufficient persuasive force to merit our taking UMET seriously.

This claim is one of three that are important in this chapter. The other two claims are these:

[1] I have taken the term "plausibility points" from Enoch (2011).

> **Consequence:** If *Plausibility* is true, then it is reasonable to inquire as to the consequences that UMET might have for ethical thought and talk as it currently exists in society.
>
> **Partial Escape:** Even if UMET is not true, we will still need to accommodate the insights contained in the arguments referred to in *Plausibility*.

Partial Escape is connected to the distinction mentioned above, i.e. the distinction between UMET and LMETs. What I will eventually suggest (in this thesis as whole, not in this chapter) is that many arguments often used in the literature to support UMET in fact support versions of LMET instead.

This chapter falls into three parts, corresponding to the three claims to be defended. I start with *Plausibility*.

UMET's Plausibility

§1.1.0: Stands of Suspicion

In my view, there are many grounds for suspicion about morality. It is not that there is some single master argument that clearly and incontrovertibly demonstrates that moral thought and talk is committed to some single, determinate error. (When does that ever happen in philosophy?) Instead it is that there are various grounds for thinking that morality *cannot be quite what it appears*.[2] UMET looks plausible when these disparate and largely independent stands of suspicion are gathered together. Perhaps no strand individually would be enough to establish UMET as the most plausible. But as the mysteries keep accumulating, morality's credentials can seem to be eroded to the point that we might begin to think that UMET is more likely to be true than not, or (more weakly) that UMET is more likely to be true than any other metaethical view.

Discussion in this thesis will revolve around two main strands of suspicion. One concerns the obscurity of the notion of moral obligation. The other concerns doubts arising from the phenomenon of persistent disagreement – more specifically, the failure of philosophers to reach consensus concerning any foundational proposition concerning

[2] The italicised part of the previous sentence echoes Williams (1985: p149). Of course, Williams was not a UMET-proponent.

morality. I discuss each of these concerns in the following two sections of the present part of this chapter. However, it should be appreciated that these two strands of suspicion are not the only ones we might reasonably entertain, and thus ***Plausibility*** might well receive additional support from some argumentative strands that I shall be ignoring. Suspicions about obligation and persistent disagreement just happen to be the two strands in which I am most interested.[3,4]

§1.1.1: The Obscurity of Obligation – Joyce's Argument from Reasons

Let's start with what I said above is the most important strand of suspicion about morality: the obscurity that appears to be inherent in the notion of obligation. The obscurity is brought out by Joyce's (2001) argument from reasons (henceforth: AfR), which he presents as an argument for UMET. In the present section, I offer an exposition of this argument.

---

[3] One notable strand of suspicion that I am ignoring focuses on technical issues in metaphysics concerning location problems and reduction (Jackson 1994). Streumer's (2017) argument for GNET articulates these sorts of suspicions. Streumer recycles Jackson's reduction argument (Jackson 1998 – the argument here is an application of the general strategy in Kim 1993) and derives from it a more circumspect conclusion than Jackson drew. Whereas Jackson takes the argument to show that normative properties are necessarily coextensive with descriptive properties, and are thus identical to them, Streumer emphasises that what it actually shows is only that *if there are any normative properties* they are identical to descriptive properties. However, Streumer also argues that if there are normative properties, they are *not* identical to descriptive properties. The problem that reductive realists face, Streumer argues, is that of explaining in virtue of what a particular normative predicate ascribes a particular descriptive property. Jackson thinks that this can be done by offering a "network analysis" of normative terms (Jackson 1992; Jackson & Petit 1995, 1996). Streumer, however, finds fault with such attempts, concluding that reductive realists cannot explain what determines why a given descriptive property can be ascribed using *this* normative predicate rather than *that*.

For discussion of the reduction argument see: Williamson 2001; Shafer-Landau 2003, pp. 89–98; McNaughton and Rawling 2003; Dancy 2004, 2005; Majors 2005; FitzPatrick 2008, pp. 198–201; Kramer 2009, pp. 207–12; Plantinga 2010; Suikkanen 2010; Brown 2011; Enoch 2011, pp. 137–40; Dunaway 2015.

For discussion of the network analysis argument, see: van Roojen 1996; Yablo 2000, pp. 16–19; Zangwill 2000; Schroeter and Schroeter 2009; Horgan and Timmons 2009.

[4] Another source of suspicion that morality cannot quite be all that it appears comes from reflection on human agency. Personally, I am strongly inclined to find views highly sceptical of free will (e.g. Pereboom 2001, 2014; Strawson 1986, 1994; Honderich 1988; Leiter 2005; Wegner 2002) far more plausible than compatibilist or libertarian alternatives. Of course, the literature on this topic is huge and I shall make no attempt to argue the case here.

However, it might be worth considering momentarily what the impact on morality's credentials would be if a view in the rough neighbourhood of mine were found to be correct. Would it lead to UMET? Some, apparently, have thought so. Susan Wolf thinks that the adoption of such a view requires one to "stop thinking in terms of what ought not to be. We would have to stop thinking in terms that would allow the possibility that some lives and projects are better than others." (Wolf 1981: p. 386.) Others have similarly claimed that scepticism about free will and moral responsibility commits us to something much like UMET (Copleston 1965: p. 488; Murphy 1988: p. 400; Hintz 1958; Rychlak (1979); Smilansky 2000, 2005 – however see Smilansky 1994). Most free will sceptics disagree with Wolf on this point (e.g., Pereboom 2001, 2009, 2014) and so do I.

§1.1.1.1: A recipe for Error Theoretical Arguments

Here is one way to generate an error theory about an area of discourse: first, identify a non-negotiable platitude or set of such platitudes that governs the meaning of a term that is central to the discourse. This is the conceptual step in the argument for error theory. Then, demonstrate that there is nothing in the world of which the non-negotiable platitudes are true. This is the substantive step. AfR can be condensed down to a two-premise argument of this structure.

Some examples will help. Joyce gives the example of *tapu* discourse. "Tapu", Joyce tells us, is the name for the concept of a type of forbidden-ness utilised in some Polynesian cultures:

> "Tapu" centrally implicates a kind of uncleanness or pollution that may reside in objects, may pass to humans through contact, may then be transmitted to others like a contagion, and which may be cancelled through certain ritual activities, usually involving washing. This is not a concept we employ, although one may find something similar in ancient Roman and Greek texts.
>
> (Joyce 2001: p. 1.)

How to generate an error theory of tapu discourse? Suppose we grant that the three qualities mentioned in the above quote of things that are tapu are non-negotiable platitudes that govern the meaning of what it is for something to be tapu. That is, if T is the property of being tapu, and P, C, R are the properties relating to passing through contact, contagiousness, and ritual cancellation, then we can say: Tx iff Px & Cx and Rx. This is the conceptual stage of our error-theoretical argument targeting tapu discourse.

The substantive stage of our argument is simply to note that $\neg\exists x$ st Px & Cx & Rx. So, nothing is tapu. So, tapu discourse is thoroughly infected with error, and no claims of the form "x is tapu" are true.

Other examples Joyce discusses are: an error theory of witch discourse (non-negotiable platitude: x is a witch only if x has magical powers); an error theory of phlogiston discourse (non-negotiable platitude: x is phlogiston only if x is stored in bodies and x is released during combustion and soot mainly comprises x).

Sometimes, of course, it can be the case that users of a certain discourse make a fundamental error of presupposition concerning the central term(s) of that discourse, and yet we are not tempted to draw an error theoretical conclusion about that discourse. An obvious example is motion: nowadays our concept of motion is relational, but in earlier times motion was conceptualised in absolute terms. A candidate platitude here would be: "*x* has moved in the past four hours only if *x* has an absolute position *p** that is different from the absolute position *p* that *x* had four hours ago". This embodies a presupposition error, but we do not endorse an error-theory of motion discourse. That is, we do not think that this platitude is non-negotiable. Another example is water discourse: water was formerly thought to be an element, but we don't suppose that "water is an element" functions as a non-negotiable commitment of water discourse. Those who endorsed this incorrect platitude in days of old were still talking about the same stuff that we talk about, and very often correctly asserted things about water, such as that water can be used to extinguish fire and to satiate thirst.

How then do we distinguish non-negotiable platitudes, which suitably combined with an eliminativist substantive argumentative step have the power to generate an error theory about a given discourse, from negotiable platitudes which do not? Part of Joyce's answer is to explicate the notion of non-negotiability in terms of counterfactual interpretation. Suppose we found some community using some term A, and the use of A was governed by only two of the three platitudes concerning phlogiston considered above. Translating their term A into English, would we use the word "phlogiston"? Another part of his answer, more relevant to our purposes, is to consider what the *point* of the discourse in question is, and ask ourselves: if the platitude or set of platitudes were to be discarded, would the discourse that remained still serve the same purpose?

> The distinction [between discourses which can survive the diagnosis of an erroneous presupposition and those which cannot] must revolve around how important is the "fault" that we discover to the discourse in question. In the case of witches, for example, the whole point (one might say) of having a witch discourse was to refer to women with *supernatural* powers. To discover that no human has supernatural powers is to render the whole discourse pointless. [...] By comparison, the point of having a "motion discourse" was to refer to the change in position of objects in space over

> space and time. There was never a particular need to refer specifically to *absolute* motion.
>
> (Joyce 2001: p. 96.)

In summary, then, one way of generating an argument for an error theory of some discourse D is to identify some non-negotiable commitment or set of commitments of D, and then show that said commitment(s) are erroneous. AfR is an argument of this type for UMET. Let us now turn to look at the non-negotiable commitment of morality that Joyce identifies at the conceptual stage of AfR.

§1.1.1.2: The Conceptual Step in Joyce's Argument

AfR targets the notion of moral obligation. The conclusion to be derived is that there are no such things as moral obligations. So, in line with the two-step argumentative procedure outlined in §1.1.1.0, we need a conceptual claim of the form O*xa* only if P*xa* (where *x* is a proposed course of action of a given agent *a*). The conceptual claim Joyce offers could be stated as: *x* is a morally obligatory course of action for *a* only if *a* has a "non-institutional desire-transcendent" (Joyce 2011: p523) reason to perform act *x*. In this subsection, we will look at what these non-institutional desire-transcendent reasons (henceforth: NICRs) would have to be.

It is common to draw a distinction between categorical and hypothetical imperatives. Hypothetical imperatives are those which apply to an agent only as a result of the agent's having some desire or end. The imperative "read at least a little French every single day" may apply to me in virtue of my being engaged in an ongoing project of trying to become fluent in French. This and other related imperatives apply to me because if I do not comply with them, I am unlikely to become fluent in French – something I very much desire. If I lost my desire to become fluent in French, these imperatives would cease to apply to me. Categorical imperatives are not like this. They apply to agents regardless of their desires or interests.

Now, moral reasons clearly fall on the categorical side of this divide, but that alone is not what makes them problematic. For there are a species of categorical reasons of which we can make perfect sense, and these are institutional reasons. Joyce gives the example of the unwilling competitor in gladiatorial combat. The rules of gladiatorial combat

prohibit, say, the throwing of sand into the opponent's eyes. Now consider Celadus, the unwilling participant, who, during a fight that is not going his way, faces a choice: either throw sand into his opponent's eyes, in contravention of the rules, or die in combat. Does Celadus have a reason not to throw sand in his opponent's eyes?

Well, he has one type of reason: he has an *institutional* categorical reason. As a simple matter of fact, he *is* a competitor in gladiatorial combat, and simply because of this, the rules apply to him. In this institutional sense, he has a reason not to throw sand. But, as an *unwilling* participant, why should he care about the rules? Obviously, he shouldn't. So, even though there exists an institutional categorical reason for him not to throw the sand, still there is no *normative* reason for him not to throw the sand – that is, no reason that *should* motivate him not to throw the sand. From the point of view of practical reason, he has every reason to throw the sand. Thus, Joyce wants to say that Celadus' normative reasons point in the opposite direction to the institutional categorical reasons that apply to him in virtue of his role as a gladiator.

Now, what about moral reasons? These, Joyce argues, purport to be both categorical and non-institutional. Compare the institutional categorical imperative "do not throw sand in your opponent's eyes" with the moral imperative "do not secure job promotions by murdering your rivals". In the case of the institutional categorical imperative, we can make perfect sense of Celadus' rejecting the imperative, we can understand (and indeed *agree with*) his claim that the categorical reason he has not to throw sand simply does not ground the least normative reason for him not to so do. Joyce's point is that it is crucial to the functioning of moral discourse that the moral imperative "do not secure job promotions by murdering your rivals" cannot be escaped in this way. Faced with someone who claimed that the categorical moral reason not to kill rivals didn't ground a normative reason *for them* to abstain from murder, arguing that they just didn't care about morality and valued the advancement of their career more than the continued existence of their rivals, the believer in morality will not take the normativity of the reason to have evaporated in the way that the normativity of institutional reasons can evaporate. On the contrary, because moral reasons are precisely those things that are inescapable, the moralist will say that categorical moral reasons always generate normative reasons. This

is what Joyce thinks the moral believer qua moral believer is committed to. Thus, the conceptual stage of AfR amounts to the claim that an agent stands in the relation of moral obligation to a course of action *x* if and only if there is an NICR that, all by itself, constitutes a normative reason for the agent to perform action *x*.

§1.1.1.3: The Substantial Step in Joyce's Argument

With NICRs identified as a presupposition of moral obligation claims, the question arises whether any sense can be made of such reasons. Can we say of some agent A that he has a (normative, i.e. non-institutional) reason not to φ in circumstances C *simply in virtue of the fact that φ-ing in C is morally prohibited*?

To investigate this question, Joyce turns his attention to the nature of practical rationality itself. The view of practical rationality that he favours is one that we could (inaccurately) label Humean.[5] Joyce's view can be simply stated once we have in hand a crucial distinction between "subjective reasons" and "objective reasons". An agent has an objective reason to φ if φ-ing will *in fact* promote some element in what Williams (1981: Chapter 8) calls the agent's "subjective motivational set" (and abbreviates, as I will, to "S"). The agent has a subjective reason to φ if they are *rationally justified in believing* that φ-ing will promote some element in their S. Joyce's account of practical rationality is this (here I paraphrase rather than quote):

> **Joyce-PR:** An agent is practically rational to the extent that they are guided by their subjective reasons.

[5] Joyce calls his view "non-Humean" because it allows for the assessment and revision of ends, not just the identification of means. I say that we can inaccurately call his view Humean because it is instrumentalist. Millgram (1995) notes that in the literature, "Humean" and "instrumentalist" are used as synonyms. This is a mistake, says Millgram, because Hume's actual view was not that practical rationality is limited to identifying the means to our ends, but that there is no such thing as practical rationality. Joyce (2001: p. 57) also attributes this view to Hume, as does Korsgaard (1997). However, Joyce's attribution of the sceptical view to Hume is based only on the claim that Hume thinks that ends are present desires, and that present desires never fail to motivate. This makes practical irrationality impossible. Millgram's reconstruction of Hume's argument takes the form of an argument by elimination: reasoning concerns either relations of abstract ideas or relations of objects drawn from experience. In neither case is the reasoning practical. So, there is no practical rationality. Korsgaard rejects Humean scepticism on the grounds that its cost is "nothing short of the loss of personal identity" (p. 254). Although Millgram thinks that scepticism is "philosophically interesting and important" and "should be a reference point in the discussion of practical reasoning", I am not sure that he rejects Korsgaard's diagnosis of the cost of scepticism when he writes: "One is either extremely fortunate or unfortunately complacent if one has not had bleak mornings during which it seems suddenly clear that purported reasoning about action is nothing more than empty posturing" (p88).

A good way to think about practical rationality is to think about how an account of it might account for the phenomenon of practical irrationality. The teenager has a long-standing desire to do well in her GCSE exams, but finds herself often caving into the desire to play video games when she knows that in order to achieve her aim of exam success, she ought to be revising. The dieter has a long-standing goal of attaining a healthy weight, but finds himself caving into temptation when he walks past the kebab shop on the way home from work. The procrastinator knows that the deadline is approaching and that if he does not begin work *now* his final product will be less good than it would be otherwise, but his negative emotions get the better of him and he launches into an unnecessary spring clean which he irrationally feels is a prerequisite to starting work. Once the spring clean is done, he feels that the task is unmanageable in the remaining time available, and procrastinates further. What do all these examples have in common? Joyce's answer: that the desires that these practically irrational agents frustrate (to do well in exams, to achieve a healthy weight, to complete the assignment well) are in a sense *more considered* than the desires on which the agents actually act (to experience the reward mechanisms of the video game, to feast on greasy donner meat, to alleviate negative emotions somehow associated with the task at hand).

This leads Joyce to adopt an account of practical rationality highly similar to that offered by Smith (1994). Smith starts from the following platitude: "what we have normative reason to do is what we would desire that we do if we were fully rational" (1994: p. 150). In our examples, if the teenager were fully rational she would desire that she studied rather than played video games, if the dieter were fully rational he would desire that he proceeded home and cooked a healthy dinner rather than bought dinner from the kebab shop, and if the procrastinator were fully rational he would desire that he made a prompt start on his work project. Can the platitude be turned into an analysis? Yes, answers Smith, because "the platitude is related to a whole host of platitudes about *advice*" (p. 151). This leads him to the following account of practical rationality, which here I paraphrase rather than quote:

> **Smith-PR:** We are rational to the extent that we act in accordance with our normative reasons. In any circumstances C, we have normative reason to do what a fully rational counterpart of ourselves would want us to do in C.

Our procrastinator – let us call him Jonathan – finds himself in circumstances in which if he does not make a prompt start on his project then his finished product will not be as good as it should be. Call Jonathan's fully rational counterpart Jonathan+. What Jonathan+ wants for Jonathan in these circumstances is that Jonathan makes a prompt start. So, what Jonathan has most normative reason to do is to make a prompt start. He does not do this, led astray by negative feelings of anxiety and self-doubt, and it is in this that his practical irrationality consists.

As Joyce sees it, *Smith-PR* is essentially correct. What it provides is not an alternative to *Joyce-PR* but an explication of the notion of a reason to which *Joyce-PR* appeals. So far, in explicating the notion of normative reasons, we have only said that an agent has a subjective (objective) reason to φ if they are rationally justified in believing (if it really is the case that) φ-ing will promote some element in their S. But we need to offer more than this if we are to explain what practical irrationality is: Jonathan, after all, does have *some* element in his S that will be promoted by his procrastinating, namely, the desire to avoid the negative feelings of anxiety and self-doubt through which he must persevere for however long it takes him to lose them in absorption in the task at hand. But this doesn't give him a normative reason to continue to procrastinate (or, even if it does, it is a normative reason that is strongly outweighed by the normative reason to make a prompt start, namely that his finished product will be better if he does). So the course of action Jonathan has *most* normative reason to follow needs to be not just any course of action that will promote *some* element of S, but *the* course of action that recommends itself, given the totality of his S and correct deliberation from it. The idea of an ideally rational counterpart captures this idea.

Joyce and Smith's views of practical rationality are very much alike, then. There is, however, a key difference: Smith commits himself to a convergence claim that Joyce rejects. This is a difference with important consequences: because Smith thinks there are reasons to be optimistic about convergence, he is led to a form of moral rationalism;

because Joyce thinks that the convergence claim is very unlikely to be true, he is led to UMET. I discuss this is more detail in the following two subsections, but Joyce's side on this debate also receives support from the discussion in §1.1.2, concerning the implications of entrenched disagreement. If Smith is right that rational agents will converge in the way that we are assuming is sufficient for morality, we might expect to see some movement towards such consensus amongst those who have been engaged full-time in the project of identifying moral truths for decades. No such consensus appears to be in the offing.

For the time being, let us note a slightly less important difference between Joyce and Smith. This is that Smith, but not Joyce, imagines the perfectly rational counterpart of the agent to be in possession of full factual information – including that to which the actual agent has no epistemic access. Thus, we can think of Smith as replacing the word "subjective" in *Joyce-PR* with the word "objective". Joyce dislikes this move, for the following reason: suppose that some fact F – one that I could not possibly know – gives me an overriding objective reason right now to φ. If my perfectly rational counterpart is, in addition to being the perfect deliberator, in possession of the relevant information concerning fact F, then what my counterpart will want for me is that I φ right now. Thus, I am practically irrational to the extent that I do not φ. But how could this be, when I have no access to the information? We should suppose for the sake of this example that the fact F is an unlikely truth: as a matter of fact it would be epistemically irrational for me to believe F. If it is practically irrational for me not to φ, then I can be rationally faulted for not φ-ing. But how can I be faulted for failing to act on the basis of information I have no rational business believing? Hence Joyce thinks practical rationality has to do with subjective, not objective, reasons.

I am inclined to side with Joyce rather than Smith on this point. However, nothing really turns on this issue, for Joyce thinks that he can get his conclusion – that no room can be made for NICRs on *Joyce-PR* – whether we are working with subjective or objective reasons.[6] Start with *Joyce-PR* precisely as quoted above. Here, normative reasons are

[6] Noordhof (1999) argues that the requirement that the agent has full information causes problems for Smith's rationalist project.

relativised to the agent's S. But the whole point of categorical reasons is that they apply regardless of the contents of the agent's S. In order to be categorical, the NICR would have to stand independently of the agent's S, but in order to be normative the reason would have to be relativised to the agent's S. These are incompatible requirements, and so NICRs are impossible. And since NICRs are a non-negotiable commitment of morality, morality is error-infected.

Does anything change if we plug objective reasons into *Joyce-PR* instead of subjective reasons? Not really, because these are still relativised to the agent's S. Consider Williams' famous example: an agent believes (rationally) that the liquid in the glass on the table in front of him is gin, but instead is poison. He has a normative objective reason not the drink the liquid, despite not knowing and having no way of knowing that this is the case. Still there is no categoricity here, though, because the normativity of the objective reason derives from the contents of the agent's S: he desires not to be poisoned. (We can imagine making changes to the agent's S such that he comes to have an objective normative reason to drink the liquid.)

To summarise this subsection, then: the substantial premise of Joyce's AfR is that there is no sense to be made of the NICRs that the conceptual premise identifies as a non-negotiable commitment of morality. This is because the correct account of practical rationality relativises normative reasons to an agent's S, but NICRs are precisely the sort of normative reasons that are supposed to apply categorically.

§1.1.1.4: Might The Desires of All Rational Agents Converge?

As mentioned in the previous subsection, there is an important difference between Smith and Joyce in that Smith thinks that the desires of all rational agents will converge, so that everyone's rational counterparts will desire identical things for them – at least, where there are no differences in circumstances.[7]

[7] It is at least a complication and perhaps a fudge on Smith's part that Smith admits that differences in preferences can relevantly count as differences in circumstances. You and I are both looking for somewhere to eat in Beckenham one Friday evening. I prefer spicy food to plain; you prefer plain food to spicy. That the Kathmandu Masala does a spicy vindaloo provides a reason for me to eat there, but it doesn't provide you with such a reason. You might be better off heading to the George Inn for a traditional pub dinner. My perfectly rational counterpart wants that I go to the Kathmandu, yours wants that you go to the George. However my preference for spicy foods is part of my circumstances, yours for plain is part of yours. So on Smith's view there

The important thing is that if such convergence does in fact obtain, and if the convergence is towards the right kind of actions, and if we make one key assumption[8], then we have all we need to make sense of NICRs in terms of Joyce's own account of practical rationality. For suppose that the desires of all rational agents converge on, say, returning the wallet to the passer-by who has just dropped it in the circumstances that you witness his dropping it. Then, when I find myself in the situation of witnessing the passer-by drop his wallet, do I have an NICR to return it to him? It seems I do: my reason is normative in that what my rational counterpart (like everyone else's) will want for me in these circumstances is that I return the wallet. My reason will be categorical in the sense that no matter what motivational set I am starting with, I will be able to reason soundly towards accepting that I have this reason. So, if Smith can make good on his claim that we should expect the desires of all rational agents to converge, it seems that we should expect there to be sense to be made of NICRs after all. Getting clear about why Joyce thinks this project must fail will deepen our understanding of AfR.

What sort of convergence would we need in order to make sense of NICRs? Certainly, convergence on some claims *p, q, r…* (that an agent has reason to return the wallet) need to be perfect (in the sense that those agents who judge that ~*p* can be rationally faulted). But what *scope* of convergence is required – do all agents need to make all the same reasons-claims, or just some? I think the relevant notion is of perfect convergence on *some* reasons-claims, namely those with moral content. If Stanley thinks, concerning the question of how to spend his leisure time, taking into account all his relevant circumstance (including his preferences – see fn. 7), that what he has most reason to do is to devote most of his spare time to hiking, whilst Gideon thinks that, taking into account all Stanley's relevant circumstances, what Stanley has most reason to do is to devote most

is no divergence between the desires of our rational counterparts. Me+ wants that I-go-to-the-Kathmandu-in-the-circumstances-that-I-prefer-spicy-food; You+ want that you-go-to-the-George-in-the-circumstances-that-you-prefer-plain-food. No divergence here. In the main text, I try to ignore this complication. However, see fn. 11.

[8] The assumption is that *rational convergence with respect to reasons for action would be sufficient for morality, provided the convergence was towards reasons for the right kinds of action.* In making this assumption, I am obviously discounting without argument views on which convergence is neither necessary nor sufficient for morality, like Shafer-Landeau (2003). Regardless of what there is to be said for and against views like this, in the context of this chapter it should be conceded both to the defender and the opponent of UMET that convergence towards the right sort of actions would suffice for morality.

of his spare time to practicing martial arts, perhaps this is not especially bad news to someone hoping to make sense of NICRs on a theory of practical rationality like Joyce's or Smiths.[9] What would be disastrous would be if divergence like this threatened distinctly moral reasons – if, in the returning-the-wallet example, Gideon and Stanley found themselves in fairly similar circumstances,[10] and Stanley judged that his normative reasons favoured returning the wallet, whilst Gideon judged that his normative reasons favoured keeping the wallet for himself, but Gideon could not be rationally faulted for his so judging.[11]

Smith is optimistic that convergence of the required type is likely. His main thought appears to be that on the whole, moral argument tends to elicit agreement:

> [T]he empirical fact that moral argument tends to elicit the agreement of our fellows gives us reason to believe that there will be a convergence in our desires under conditions of full rationality. For the best explanation of that tendency is our convergence upon a set of extremely unobvious *a priori* moral truths. And the truth of these unobvious *a priori* moral truths requires, in turn, a convergence in the desires that fully rational creatures would have.
>
> (Smith: 1994: p. 187)

Now the reaction of a certain type of reader at this point will be to baulk. Isn't there a very well-known and widely discussed argument for UMET that proceeds from precisely the opposite observation, that moral discourse is characterised most notably by terminally entrenched *dis*agreement (see §1.1.2)? This argument starts from the observation of irresolvable disagreement and offers an argument to the best explanation: namely, that

[9] Smith thinks that to avoid an error theory of normative reasons, convergence on even this sort of reasons-claim is required. In relaxing this requirement, clearly I am making it easier to avoid ***Plausibility***.

[10] "Fairly similar" rather than "exactly similar" because what the defender of NICRs needs to guard against is that differences in Stanley and Gideon's S can make a normative difference to whether or not to return the wallet. But Smith counts the agents' preferences as part of the circumstances.

[11] Sobel (1999) notes a difference between tastes or preferences on the one hand (like my preference for spicy food in the discussion in fn. 7) and ideals on the other hand. He objects to Smith on the grounds that for convergence of the sort we are interested in here, all agents will have to agree both on: (i) what counts as a mere taste or preference and what counts as an ideal, and (ii) their ideals. This, he objects, makes no room for personal ideals – those things that we want for ourselves come what may, regardless of any future changes in our mere preferences, but which we do not see as binding on other agents. But I think Smith could easily deal with this. All that would be required would be for agents to agree on: (i) what counts as a mere taste, what counts as a personal ideal, and what counts as a moral ideal, and (ii) their moral ideals. Moral ideals could involve such things as a commitment to respecting others' personal ideals more robustly than respecting their preferences.

there is no truth to be had in moral matters, and that is why disputes never come to any end. But Smith starts by taking it as established "empirical fact" that moral argument tends to produce agreement rather than disagreement. Why the optimism?

His grounds are threefold. First, "alongside such entrenched disagreements […] we also find massive areas of entrenched agreement" (p. 188). As evidence of this, Smith adduces the fact that we have and use thick ethical concepts. A practice of using such terms as "brutal", "duplicitous", "mean" and "treacherous" requires consensus that acts that are *descriptively* such that they can be described with such adjectives are to be evaluated negatively.[12] Likewise, a practice of using such terms as "honest", "courageous", and "public spirited" requires consensus that acts characterised in these ways are to be evaluated positively. Second, "when we look at current areas of entrenched disagreement, we must remember that in the past similarly entrenched disagreements were removed *inter alia* via a process of moral argument" (ibid). Here Smith asks us to consider such issues as slavery, workers' rights, women's rights and democracy. Third, and finally, Smith points out that much moral disagreement can be explained by irrationality on the part of some or all of the disputants. He puts much moral disagreement down to people's (irrational) tendency to base their moral opinions on "the directives of a religious authority rather than as the result of the exercise of their own free thought in concert with their fellows" (p. 188-9).

The conclusion that Smith reaches, then, is this:

> We should […] be quite optimistic about the possibility of an agreement about what is right and wrong being reached under more idealised conditions of reflection and discussion. We might eventually become pessimistic, of course. Our epistemic situation might deteriorate, widespread disagreements might emerge, disagreements that seem both unresolvable and inexplicable. And if that were to happen, then we might well quite justifiably come to think that Mackie was right after all, that there are no moral facts, though there would still be room for doubt. The point is simply that this *is not* our current epistemic situation.
>
> (Smith 1994: p. 189.)

[12] Here, Smith seems to presuppose that the normative element of thick concepts are part of their meaning, but this is disputed by some (e.g. Vayrynen 2013). However, many commentators agree with Smith here (e.g., Roberts 2011, 2013, Kirchin 2017), and I shall assume he is right.

Notice that the resulting view is a form of moral rationalism: moral failings are failings of practical rationality. As Joyce sees it, some form of moral rationalism along these lines is indeed the moralist's best hope for avoiding UMET. However, according to Joyce, moral rationalism of this sort will not work. The trouble is that the rationalist faces a dilemma: on one horn of the dilemma the rationalist alienates the agent from their reasons, on the other reasons are relative to agents, thus the rationalist project of finding convergence in the desires of all rational agents is sunk.

§1.1.1.5: The Rationalist's Dilemma

Bear in mind that in our account of practical rationality, we are after an account of an agents' *normative* reasons. If R genuinely is a reason for A to φ, then it should be the case that if A asks us "why should I φ?" and we reply "because R", this answer forestalls a response of "so what?" or "and why should I care about *that*?". It is this that the device of the rational counterpart in ***Smith-PR*** achieves. Suppose I think that R constitutes an overwhelming normative reason for Barry to φ, but Barry is currently unmotivated to φ. I say to Barry: "you ought to φ". Barry replies: "why?" I give a reason, R. Perhaps R is that it will make Barry's next door neighbour happy. Barry asks: "and why should I care about that?" Barry has twice asked me for a reason, he has thus communicated to me that he is "in the business of accepting reasons" (Joyce 2001: p. 83), that he is willing to weigh up reasons for and against his φ-ing, and act on what appears to be the best reason or set of reasons. Simply by virtue of this, Barry is committed to taking the desires of Barry+ as reasons. If it really is the case that what Barry+ would want for Barry at this moment in time is that Barry φs, and that Barry+ wants this precisely because Barry's φ-ing will make Barry's neighbour happy, then if I can get Barry to see that this is the case, Barry cannot (meaningfully, legitimately, without incoherence) say: "yes, I see that Barry+ would want me to φ because R, but why should I care about what Barry+ wants?". This is because what Barry+ wants for Barry just is what Barry would want if he deliberated correctly. So the "so what?" question becomes: "yes, I see that if I deliberated correctly I would want to φ because R, but still, why does R constitute a reason for me to φ?", and

this has the obvious ring of incoherence about it. In asking for reasons, an agent enters into the project of deliberation.

But this unintelligibility of a "so what?" response depends on there being the right kind of link between the agent and their idealised counterpart. If we instead made the idealised counterpart exactly like Firth's (1952) *ideal observer*, for example, then plausibly the link is severed: Firth's ideal observer is supposed to be not just ideally rational but *disinterested and dispassionate*. But when Barry asks me why he should φ, he "in no (obvious) way implies an allegiance to dispassionateness" (Joyce 2001: p83). If it turns out that in saying that R constitutes a reason for Barry to φ, I mean not that Barry+ but that a *Firthian* ideal observer of Barry would want that Barry φs because R, it seems perfectly intelligible that Barry could reply: "yes, I know that this ideal observer would want that I φ because R, but what is that to me?".

It is the requirement that the idealised counterpart of the agent needs to be suitably connected to the actual agent that generates the dilemma, which is this. On the one hand, we can try to specify the idealised counterpart of the agent in such a way as to keep the link to the actual agent fairly close. What we would most probably need to do here is to start with the agent's individual subjective motivational set, complete with all its idiosyncrasies, and then add *only* faultless deliberative capacities. In this case, there is no reason to expect convergence between agents. On the other hand, if the idealised counterpart is specified in such a way as to secure convergence, then the link between the actual agent and the idealised counterpart is severed, such that the agent is alienated from the reasons provided, with the consequence that these no longer appear to be *normative* reasons.

Consider the first horn of the dilemma. Here, we are specifying the idealised counterpart of the agent in such a way that we start with actual agent's individual subjective motivational set, and add only faultless deliberative capacities. The (possibly quite idiosyncratic) contents of this individual subjective motivational set are bound to make a lot of difference to what normative reasons the agent has. In my advice to Barry, I said that R constitutes a reason for him to φ because *Barry+* would want that Barry φs in the circumstances. What if, instead, I had said that R constitutes a reason for him to φ because

Larry+ would want that Barry φs in the circumstances, where this is taken to mean: starting from Larry's subjective motivational set, what Larry has most reason to do is to see to it that Barry φs? Then a "so what?" response on Barry's part would be not only intelligible but (presumably) entirely legitimate! What Larry+ wants for Barry in these circumstances will be what perfect deliberation would recommend when it starts from Larry's subjective motivational set, not Barry's. Why would Larry's subjective motivational set ground reasons for Barry? Larry might even be both cruel and an enemy of Barry, and might want for Barry that Barry φs for reason R simply because Barry's φ-ing for R would lead to Barry's suffering a gruesome and painful injury.

Consider now the second horn of the dilemma. Here, we are not allowing the details concerning the contents of the agent's subjective motivational set to make much of a difference: the idealised version of the agent is something like a Firthian ideal observer. Let's call this observer Harry. Again, in my advice to Barry, I said that R constitutes a reason for him to φ because Barry+ would want that Barry φs in the circumstances. What if I had instead said that R constitutes a reason for him to φ because *Harry* would want that Barry φs? Presumably, Barry will want to know who this Harry character is. I tell Barry: "Harry is perfectly rational – he is a Firthian ideal observer", and intend by that to have provided evidence that we *all* have reason to do as Harry wants us to do. Arguably – and certainly in Joyce's opinion – a "so what?" response from Barry remains both intelligible and legitimate.

Specify the idealised counterpart such that convergence is achieved, then, and you alienate the agent from their normative reasons. On the other horn of the dilemma, where the link between the idealised counterpart and the actual agent is kept fairly tight, we should expect normative reasons to be relative to the agent. But this means that there is no guaranteed convergence in agents' normative reasons, and *a fortiori* no convergence towards reasons to return the wallet, to honour the commitment, to help others in need, and so on. [13]

[13]Although Joyce takes Smith as his principle opponent, there is another view in this neighbourhood that might rescue the idea that all agents who fail to act in the way that morality requires are guilty of practical irrationality. This is Schroeder's (2007) hypotheticalism. According to hypotheticalism, if an agent has a reason *r* to perform

The upshot of the rationalist's dilemma appears to be this. For any putative NICR, it seems likely that a plausible counter-example can be constructed. The counterexample will take the form of imagining a rational agent whose subjective motivational set is such that the agent's rational counterpart would not desire that the agent be motivated by the putative NICR. But this means that there is no convergence of rational agents. Smith's moral rationalism is sunk.

§1.1.1.6: Improving Joyce's Argument

As we have seen, AfR is based on an instrumental view of practical reason. Suppose we think that this is a vulnerability in the argument. Can the argument be freed from this not uncontroversial commitment? Perhaps.

In a chapter immediately following the one in which he discusses the rationalist's dilemma, Joyce offers further support for AfR by defending from various criticisms the famous argument Williams (1981: chapter 8) makes about external reasons claims. The idea here is that claims of the form 'A morally must φ' must be like external reasons claims in that they will not be falsified by the absence of a relevant element in A's subjective motivational set. Joyce thus defends Williams' argument against external reasons claims, and thereby takes himself to have further supported the idea that there is no sense to be made of reasons that are at once normative and categorical.[14]

I am not going to offer a detailed exposition of this part of Joyce's argument. Instead I wish to make a single observation about it. This part of the argument, no less than the part concerning the rationalist's dilemma, is based on the instrumental view of reason. This is because Joyce takes the instrumental view of reason to be central to the concerns of the

---

some action, then what explains why the agent has this reason is that they have some desire, the object of which will be promoted by the action.

Whilst obviously similar to Joyce's account, Schroder thinks his can deliver agent-neutral reasons, and thus accommodate moral reasons. However, on Schroeder's view, agent-neutral reasons are relativised to the activity in question: "The right kind of reasons involved in any activity are the ones that the people involved in that activity have, *because* they are involved in that activity" (p. 135). But this suggests that moral reasons are relativised to the activity of, say, *conforming ones behaviour to moral standards*. This makes morality look once again like an institution. But the whole point of the conceptual stage of AfR is that morality presents itself not just as another institution but something with inescapable authority.

[14] Strictly speaking, Joyce disagrees with Williams that external reasons claims must all be false. Joyce thinks that there might be true external reasons claims, but that these claims will not be about *normative* reasons. He thinks that there are no true external reasons claims about normative reasons.

argument Williams presents in 'Internal and External Reasons'.[15] However, it is arguable that the instrumental view of reason has little or nothing to do with the Williams argument. For instance, on Finlay's (2009: p. 15) reading of Williams, an upshot is that "at least some deliberation takes the form of reasoning to a conclusion about what one has reason to do" and that "to deliberate is sometimes to reason to an explanation of what one would do if one deliberated soundly". This, Finlay notes, is a strikingly different view of practical rationality to instrumentalism.

Finlay is not the only or the first person to have made such observations about Williams' argument, and the finer details of Williams exegesis need not detain us. The little that has been said already is enough for me to make my suggestion as to how AfR could be improved. Adopting a reading of Williams' argument like Finlay's, Joyce could take the argument against external reasons to show that no sense can be made of NICRs even if instrumentalism is not true. Then AfR could start from the obviously true disjunction that either instrumentalism is the correct account of practical rationality, or it isn't. If instrumentalism is the correct account of practical rationality, then anyone who wants to make sense of NICRs will have to face the rationalist dilemma and explain why agents are not alienated from their normative reasons. On the other hand, if instrumentalism is not the correct account of practical rationality, then anyone who wants to make sense of NICRs will have to explain either how there can be external reasons, or explain how there could be internal reasons that all agents share at all times, regardless of the peculiarities of their subjective motivational sets. This represents an improvement in the dialectical force of Joyce's available arguments in support of UMET.

§1.1.2: The Implications of Moral Disagreement

A second strand of suspicion that morality cannot quite be all that it appears to be comes from the phenomenon of moral disagreement. As previously remarked, this strand is connected to the concerns about the obscurity of the notion of moral obligation. Persistent

[15] For example, Joyce (2001: p109) says that merely believing an external reasons statement about oneself cannot motivate one to act on it because "[a] very familiar and eminently sensible view says that in order to explain an action the belief must couple with desires". Thus, on Joyce's reading, Williams' argument against external reasons claims is an argument for instrumentalism.

disagreement seems to present something of a problem for Smith's optimism about convergence. But this strand also stands independently, and provides further reason to suspect that morality cannot quite be all that it appears to be.

Broadly speaking, there are appear to be at least four main contexts in which the phenomenon of irresolvable disagreement tends to be recruited for argumentative purposes by metaethicists: (i) it is used to attempt to establish some form of cultural relativism (e.g. Brandt 1954; for criticism see Moody-Adams 1997); (ii) it is used to undermine moral realism, without targeting other non-MET metaethical views, like expressivism (e.g. Loeb 1998, where the target is naturalistic moral realism of the sort proposed by Boyd 1988); (iii) it is used epistemologically, to undermine claims of moral knowledge (e.g., McGrath 2008); (iv) it is used to establish some form of MET. In attempting to assemble a case for ***Plausibility***, my interest in arguments based on persistent moral disagreement should be restricted to (iv); indeed, it should be restricted to the sub-case of (iv) in which the form of MET established is UMET.

However, as does nearly every other commentator when it comes to this topic, I shall start (albeit briefly) with what Mackie called his argument from relativity, despite the fact that Mackie seems to use this argument to undermine realism rather than to establish UMET directly.[16] Mackie begins:

> The argument from relativity has as its premiss the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community.
>
> (Mackie 1977: p. 36.)

He takes the existence of such variation "indirectly [to] support second-order subjectivism" because it makes it "difficult to treat those judgements as apprehensions of objective truths" (ibid). This is because, unlike scientific disagreement where the best explanation for its existence is that it results from "from speculative inferences or

[16] That is how Loeb (1998) treats Mackie's argument. More specifically, he sees it as specifically undermining intuitionistic moral realism. I think that what is happening here is a common occurrence in Mackie's discussion: for him, there is little distance between disproving realism and establishing UMET, because he thinks the presuppositions of the ordinary believer in morality include all the commitments of realism. Here, this includes a commitment to intuitionism.

explanatory hypotheses based on inadequate evidence" (ibid), the best explanation of persistent moral disagreement is that differing moral opinions "are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values" (p. 37).

Now I am not going to pause to consider the merits of this version of the argument, because it is not the version of the argument that I want to claim exerts pressure in moving us towards ***Plausibility***. All I want is to draw attention to two distinctions, one of which is explicitly mentioned by Mackie in the above quotation. The first distinction is between synchronic and diachronic moral disagreement. I am opposed to the institution of capital punishment (see §5.2.8), and thus disagreement exists between me and contemporary supporters of capital punishment. This is an example of synchronic disagreement. But there is also moral disagreement of the diachronic type: in depressingly recent human history, for example, support for slavery was the dominant moral opinion. Thus, there exists disagreement between current British and American society on the one hand and British and American society as it was in the early nineteenth century on the other concerning the morality of slavery. This is an example of diachronic disagreement. This distinction is mentioned by Mackie in the quotation above when he mentions variation in moral codes both "from one period to another" and "within a complex community".

The second distinction is between, on one hand, the sundry disagreements that occur in everyday life concerning those specific, concrete moral opinions that are relevant to deciding what to do with respect to some particular problem or the other, and on the other hand, disagreements that occur amongst philosophers concerning *foundational propositions about morality* (Leiter 2014). These are the sorts of propositions – such as the proposition that the wrong-making feature of an action is that it negatively affects aggregate utility – about which philosophers argue.

With these distinctions in place, we can note the following: Mackie's argument from relativity takes as its premise an observation concerning the persistence and irresolvability of both synchronic and diachronic disagreement concerning specific, concrete moral opinions. This is not obviously the best place to look. A stronger push in

the direction of accepting ***Plausibility*** might come from noting the *failure of philosophers to reach consensus* concerning even a single foundational propositions about morality. Here it is neither synchronic nor diachronic disagreement per se that is at issue, and yet both are relevant; the phenomenon to be explained is the failure of philosophers to show any signs of moving towards consensus concerning the truth value of any foundational proposition about morality. According to Leiter, Nietzsche offers an argument for UMET based on the observation that philosophers have failed to converge on any such propositions. As he (Leiter) says, this argument is of independent philosophical interest, and the remainder of this section is devoted to a consideration of it.

The following passage is helpful to Leiter in attributing the present version of the argument from disagreement to Nietzsche:

> It is a very remarkable moment: the Sophists verge upon the first critique of morality: –they juxtapose the multiplicity (the geographical relativity) of the moral value judgements; –they let it be known that every morality can be dialectically justified; i.e., they divine that all attempts to give reasons for morality are necessarily *sophistical* – a proposition later proved on the grand scale by the ancient philosophers, from Plato onwards (down to Kant); –they postulate the first truth that a "morality-in-itself", a "good in itself" do not exist, that it is a swindle to talk of "truth" in this field. (WP 428.)

Of course, my present concerns have nothing to do with Nietzsche scholarship, and so I shall ignore the question of whether this passage, whilst from the *Nachlass*, is nonetheless representative of Nietzsche's view. The question is whether a decent argument can be extracted from it.

The important claims here are the closely related: (A) that "every morality can be dialectically justified"; (B) that "all attempts to give reasons for morality are necessarily sophistical". As Leiter says, the word "sophistical" here "is obviously meant to have the pejorative connation that the *apparent* dialectical justification does not, in fact, secure the truth of the moral propositions so justified" (p.137). Thus (A) is to be understood as the claim that every morality can have the *appearance* of being dialectically justified – this appearance is to be understood as misleading. But why, Leiter goes on to ask, should the

work of philosophers "from Plato onwards (down to Kant)" *prove* (B) to be correct? Answer: Nietzsche makes an abductive inference to the best explanation. Here's Leiter:

> The best explanation for the *existence* of incompatible moral philosophies providing dialectical justifications for conflicting moral truths is that (1) it is possible to construct *apparent* dialectical justifications for such moral truths, because (2) given the diversity of psychological needs of persons (including philosophers), it is always possible to find people for whom the premises of these dialectical justifications seem plausible and attractive, and (3) there are no objective moral facts offering an obstacle to the philosopher satisfying his psychological needs in this way.
>
> (Leiter 2014: p. 138.)

This quotation, whilst brief, does in fact lay out entirely the bare bones of the Leiter-Nietzsche argument (as I shall henceforth call it). It seems to me a reasonably persuasive argument, but since the disagreement strand of the argument from obscurity is the less important, subsidiary strand, I shall be brief here and emphasise just two points.

First, the restriction of the present argument to failure of consensus on foundational moral propositions within academic moral philosophy represents an improvement over Mackie's argument from relativity insofar as the present argument is immune to some objections that the argument from relativity naturally faces. For the following are all commonly made objections to the argument from relativity: (i) most moral disagreement is not really intractable, but reflects factual disagreement about, e.g. what consequences will actually follow from which actions – factual disagreement that exists against the background of agreement on basic moral principles; (ii) moral disagreements about basic moral principles are resolvable in principle; (iii) where moral disagreement about basic moral principles is not resolvable, this is best explained by cognitive defects, irrationality or bias on the part of one or more parties to the dispute; (iv) where moral disagreement about basic moral principles is not resolvable and not explicable in terms of cognitive defects and the like, it is due to differences in background theory. But whatever these objections have going for them in their original context, they do not appear to have any force against the Leiter-Nietzsche argument. (i) doesn't seem plausible as an explanation of the entrenched disagreement between Kantians and utilitarians over the wrong-making feature of actions. What, after all, could it be that one party knows that the other doesn't?

No plausible candidates suggest themselves. Whether (ii) is plausible or not when applied to disagreements about concrete moral opinions, it seems absurdly optimistic when applied to disagreement over foundational propositions about morality. With regard to (iii), Leiter notes that "one must appreciate how strange it is in response to the Nietzschean argument appealing to disagreement among moral philosophers across millennia. Are we really to believe that hyper-rational and reflective moral philosophers, whose lives, in most cases, are devoted to systematic reflection of philosophical questions […] have reached no substantial agreement on any foundational moral principle because of ignorance, irrationality or partiality?" (p. 141-2.) Finally, (iv) is an obvious failure as a response to the Leiter-Nietzsche argument because the background disagreement is the very thing to be explained.[17]

Second, the most obvious objection to the Leiter-Nietzsche argument can be easily countered. This is the objection that the argument has proved too much. This objection can be made in one of two ways. In what we could call the "whole-discipline" version of the objection, the complaint is that the failure of philosophers to reach agreement on any foundational moral proposition is unremarkable by the standards of the discipline as a whole. There is likewise no convergence on any of the foundational propositions about which the debates rage in epistemology or metaphysics either. This observation could be recruited as the basis of a "companions in guilt" argument to the effect that, although it is not currently known which of the various foundational moral propositions are true, some of them are. But why resist the other natural conclusion? Leiter points out that Nietzsche himself seems happy to embrace it, citing BGE 6 in which Nietzsche says that "what every great philosophy so far has been" is nothing more than "the personal confession of its author and a kind of involuntary memoir", and appears to view the details of the epistemological and metaphysical systems as parasitic on the moral aims of their authors. Might we have good reasons to follow Nietzsche here? We certainly might if we agree with Van Inwagen (2008: p. 10) that in metaphysics "there is no information and there

[17] Everything I have said in this paragraph is discussed at greater length in Leiter's paper.

are no facts to be learned besides information and facts about what certain people think, or once thought, concerning various metaphysical questions".[18]

The second way in which the objection might be put is what we might call the "self-refutation" version. The conclusion the Leiter-Nietzsche argument aims at is that *there are no moral truths*. But most philosophers will not conclude, on the basis of the Leiter-Nietzsche argument, that this is so. So why not say that, by dint of this meta-disagreement, there is no fact of the matter as to whether we should infer that there are no moral truths on the basis of the Leiter-Nietzsche argument? Answer: we need to consider what the *best* explanation for the persistence of disagreement is. In the case of disagreement over foundational moral propositions, quite possibly the best explanation is that one Leiter gives in the quotation above. In the case of disagreement over whether to infer that there are no moral facts on the basis of the Leiter-Nietzsche argument, other explanations suggest themselves. Leiter offers the following candidate:

> Surely one possibility – dare I say the most likely possibility? – is that those who are professionally invested in normative moral theory as a serious, cognitive discipline [...] will resist, with any dialectical tricks at their disposal, the possibility that their entire livelihood is predicated on the existence of ethnographically bounded sociological and psychological artifacts.
>
> (Leiter 2014: p. 148.)

It seems to me that the explanation that Leiter offers here as to why there is persistent disagreement over whether we should infer, on the basis of the Leiter-Nietzsche argument, that there are no moral truths is eminently plausible. At least it seems more plausible than the explanation that there is no fact of the matter as to whether we should so conclude. And certainly, the explanation that there are no moral truths is a better explanation of the persistence of disagreement concerning foundational truths about morality than the explanation that there is no fact of the matter as to whether we should conclude that there are no moral truths on the basis of the Leiter-Nietzsche argument is of the disagreement over that, given the availability of the alternative explanation

[18] Similarly pessimistic views about the possibility of philosophical knowledge are expressed in Van Inwagen (2006) and Lycan (2013).

provided in the above quotation. So, it doesn't seem that Leiter-Nietzsche argument is guilty of self-refutation.

§1.1.3: The Argument from Obscurity for UMET

Putting both strands of suspicion together, we can formulate a stronger argument for UMET that combines their force. This is the argument from obscurity:

Argument From Obscurity

(1) There appear to be many confusions involved in moral thought. For example:
   a. Moral judgements *routinely* (but perhaps not always) appear to express claims about the existence of NICRs, but there is no sense to be made of such reasons;
   b. Moral believers often take themselves to have sound reasons for endorsing their moral beliefs in the face of competition from rival moral beliefs in a way that commits them to believing that it is possible to establish which moral beliefs are correct on the basis of the standard procedures of moral inquiry, but the failure of philosophers to achieve consensus on even a single foundational proposition about morality suggests that this faith in these procedures is misplaced.

(2) If moral thought is confused in the way that (1) outlines, then it is obscure what S believes when S believes that *z*, where *z* is a moral proposition.

(3) If S believes that it is obscure what S believes when S believes that *p*, then S should not believe that *p*.

(4) It is obscure what I believe when I believe *z*, where *z* is a moral proposition. (From (1) and (2).)

(5) I should not believe *z*, where *z* is a moral proposition. (From (3) and (4).)

Since the conclusion of this argument is that I should not believe *z*, where *z* is any moral proposition, this argument, if successful, establishes UMET, rather than some version of LMET. But where is the proper place to draw the boundary between UMET and LMET? The time has come to say a little about this. I shall therefore close the discussion of *Plausibility* with a section devoted to this question.

§1.1.4: UMET and LMETs; Ethics and Morality

Think back to AfR. Whilst I am using it as just one strand in the (hopefully) more persuasive AfO, Joyce offers it for consideration as an argument that, all by itself, establishes UMET. But it might strike us that there is something of a mismatch between the focus of the premises of AfR on a seemingly rather narrow moral concept – viz., that

of moral obligation – and the wide scope of the theory that it is used to support. Grant for the sake of argument that nobody is ever under a moral obligation. Grant that all talk of moral obligations is incoherent, because it presupposes that sense can be made of NICRs, and it can't. Now suppose that I say that capital punishment is wrong. Do I necessarily involve myself in error? I certainly do so if what I mean when I say that capital punishment is wrong is that the state is under a moral obligation not to execute criminals. But what if I mean only that a state of affairs in which all states outlaw capital punishment *is morally preferable* to a state of affairs in which some states make provision in their criminal codes for such punishment? Have I involved myself in error? In what sense do I commit myself to there being sense to be made of NICRs when I make this claim? Or, to switch examples, what if I evince disapproval of the Croydon cat killer by saying that they are *cruel*? It is even less clear now that I commit myself to some claim invoking NICRs. So how does Joyce get from here to UMET, rather than just an LMET targeting the notion of moral obligation?

Joyce's answer is presented over the course of a mere two pages (pp. 175-6). Three main parts to his response can be identified: (i) the notion of obligation underwrites most ethical concepts; (ii) non-infected regions of the ethical will not be recognisable as a moral discourse; (iii) there may well be other errors in the ethical besides the commitment to NICRs. I certainly agree with (iii). After all, the whole reason I prefer AfO to AfR as the argument to support ***Plausibility*** is that I think that there are a collection of reasons to be suspicious about morality. But however many errors we diagnose in morality, we still need a way to decide whether the conclusion to which the diagnoses point is UMET or a collection of LMETs. What I want to do in this section is to consider (i) and (ii), in order to draw a (stipulative) distinction between ethics and morality, and to understand exactly how a MET could be an instance of *U*MET.

§1.1.4.1: Contagion Spread

The first main part of Joyce's response to the concern about the specificity of obligation is that the infected notion of obligation underwrites many other ethical notions, so that

the infection spreads from this infected central notion to almost everything we recognise as ethical thought and talk:

> If there are no inescapable moral obligations, for instance, then there will be no inviolable claim rights (and *claim* rights are the central currency of ordinary rights-based discourse). Similarly, talk of virtues and vices generally implies the existence of obligations. Virtues are often thought of as character traits that one is obligated to cultivate […]. [A] virtuous agent is taken to be one who is, *inter alia*, sensitive to, and acts in accordance with her moral obligations.
>
> (Joyce 2001: p. 175.)

Similarly with regards to thick ethical concepts:

> But what is *evil* if not something we ought not do or be? A "thick" evaluative may term have a comprehensible descriptive component, but it also necessarily has an evaluative component, and this evaluative component demands explanation here no less than it does for a "thin" evaluative term like "good". My claim is not that all thick evaluative terms are suspect, but that those at the heart of *moral* discourse are.
>
> (p. 176.)

The last sentence of this quotation is interesting. If there are thick evaluative terms that are not "at the heart of *moral* discourse", and are thus presumably exempt from the attribution of error, then in what way are they evaluative, and is this way ethical? Several commentators have drawn a distinction between the ethical and the moral on which the ethical is a wider category than the moral, with the moral constituting part of the ethical (Darwall 2018, Lillehammer 2013, Kramer 2009). In the context of discussing MET, the obvious stipulative terminology would be to define morality as *that which is targeted by a given argument for MET*. In taking different targets (the concept of natural rights, the concept of blameworthiness, the concept of inescapable authority), different METs thus utilise different conceptions of morality. Ethical evaluation, meanwhile, would be any evaluation that relates to the serving or frustrating of the interests of others (cf. Greene 2002: p. 19-20).

The question now is: how much of the evaluative does a MET-proponent have to conceive of as being morality (where the word "morality" is now a pejorative) before they count as a *U*MET-proponent? What definitely needs to be resisted here is defining UMET

in such a way that the UMET-proponent is committed to seeing the whole of the ethical realm as coextensive with the moral. In Chapter 3 I will be arguing that abolitionists tend to involve themselves in self-defeat when they attempt to advertise the advantages of their policy-proposals. But clearly if UMET is defined as the view that the whole of the ethical is coextensive with the moral, then this argument becomes too easy to make. Of course the advertisement of the advantages of their policy proposal is going to involve them in the realm of the ethical. The question in Chapter 3 will be whether the involvement in the ethical is so great as to attract a charge of inconsistency. I will argue that abolitionist proposals to be found in the literature currently do this. This will form the beginnings of the objection to UMET to be pressed in this thesis, namely, the objection from loss. But it would obviously be cheating to saddle the abolitionist *qua* UMET-proponent with such an extreme commitment at this stage. That would reduce the argument to come to an uninteresting *ignoratio elenchi*.

However, an easy answer suggests itself, and a particularly good way to see the merits of this answer is via reflection on Joyce's second response to the concern about the specificity of obligation. Let us now turn to that, then.

§1.1.4.2: Dialectical Insufficiency of Uninfected Nodes

The second part of Joyce's response to the present concern is as follows:

> It is not necessary for me to claim that absolutely every piece of recognisably moral language implies the validity of categorical imperatives, only that a sufficient portion of them do, such that if we were to eliminate categorical imperatives and all that imply them from the discourse, whatever remained would no longer be recognisable as – could not play the role of – a *moral* discourse.
>
> (ibid.)

In other words: it's not just that the notion of moral obligation infects many moral concepts, it's also that those moral concepts that are not touched by AfR are insufficient to ground the *system* of morality:

> Moral discourse, in other words, is a house of cards, and the card at the centre bottom has "categorical imperative" written on it.
>
> (p. 177.)

A better image than a house of cards might be a web of justification. Ethical discourse, Joyce could say, is a web whose central node is labelled "categorical imperative". From this central node, the contagion spreads along many of the connections between the ethical judgements in the web, such that much or most of the web is infected with error. This allows the possibility that there are individual judgements in the web that survive. (My disapproving judgement that the actions of the Croydon cat killer were cruel looks like a prime candidate.) Does the critic who wishes to turn the present concern – that the notion of obligation is too specific to ground the case for UMET – into a decisive objection emerge victorious simply because we have identified a small class of surviving ethical objections?

Not necessarily. The first quotation in the present subsection could be elaborated. "True," Joyce could say, "my AfR does nothing to impugn your disapproving judgement that the actions of the Croydon cat killer were cruel, and it does nothing to undermine a myriad of other similar ethical judgements. But suppose that judgements like these were *all* that survived. Could they ground anything *recognisable* as ethical discourse? I think the answer to this is that they couldn't. And so we are led to a universal MET: moral discourse is practically the whole of ethical discourse and it is sunk."

The point I am driving at here is that plausibly a MET ceases to be an LMET and becomes full-blown UMET when the damage inflicted to the ethical web by knocking out all nodes infected with error is sufficient to undermine the social practice of ethical discourse. This is an important claim of mine that I will return to at various points in the thesis, so it would be wise to give it a name at this juncture:

> **UMET Condition:** It is necessary and sufficient for a MET to amount to UMET that the ethical web constructible from the uninfected nodes belonging to the old error-infected web is insufficient to serve as the basis on which to continue a social practice of ethical discourse.

This, then, gives us a method of sorting UMET from other METs. But under what conditions do we have a MET at all? Where is the line between MET and simply rejecting a collection of moral views? Consider a Christian moralist who thinks that certain acts

can be sinful, but not dispositions or vices.[19] Does this count as LMET? A natural thought here is that the answer to this question is "no", and that this is because a MET must reject a collection of moral claims *based on the rejection of a concept that governs the standards to which those claims are intended to be responsive*. Someone who wanted to count the view of the Christian moralist currently under consideration and who accepted this condition would have to say that the Christian moralist counts as an LMET-proponent because he rejects the concept **SINFUL DISPOSITION** that some of his fellow Christian moralists use to ground some of their normative judgements. However, the concept **SINFUL DISPOSITION** is obviously just as decomposable as the concept **CUP FILLED WITH COFFEE**.[20] Just as the latter decomposes into the concepts **COFFEE**, **CUP**, and **FILLED**, so the former decomposes into **DISPOSITION** and **SIN**. And the moralist currently under consideration does not reject the concept **SIN**. So, he does not count as an LMET-proponent. He would count, however, if he rejected the concept **SIN** entirely. For in this case he rejects all normative claims concerning which acts, thoughts, intentions, dispositions, vices, etc. are sinful, and he does this because he rejects the concept (*viz*., **SIN**) that governs the standards to which users of sin-discourse suppose themselves to be in the business of responding.

This gives us:

> **MET Condition:** It is necessary and sufficient for a rejection of a collection of moral claims to amount to MET that the rejection of the moral claims is based on the rejection of a concept or collection of concepts that governs the standards to which those claims are intended to be responsive.

Putting *MET Condition* and *UMET Condition* together, then, all the following count as LMETs, because they fulfil *MET Condition* but do not fulfil *UMET Condition*: Bentham's (1792) discounting of natural rights as "nonsense upon stilts"; any ethical perspective that dispenses with talk concerning desert on grounds of scepticism about moral responsibility

[19] McClusky (2017) attributes such a view to Aquinas.

[20] I aim not to assume much here about the correct account of concepts. In particular, I am not assuming an atomistic view. All I am assuming is that proponents of views like, for example, the prototypical view of concepts owe us an explanation of why complex concepts *appear* to decompose.

but which is committed to continuing to think in terms of "what ought not to be" (Wolf 1981: p. 386 – see fn. 4); the rejection of the concept **SIN**.

UMET Condition seems to capture what it is that UMET-proponents themselves take to be distinctive about their view. For it makes best sense of the interest that UMET-proponents take in U-NWP. It is to this problem and its connection with what is to come in the thesis to which I turn in the next part of the present chapter.

Consequence

§1.2.0: The "Now What?" Problem for UMET

If UMET is true, then what should the consequences be for moral thought and talk? Should morality simply be jettisoned? Would that even be possible? And if morality ought not be jettisoned, then how could it be saved? What should we do if UMET is true?

These are the sorts of question that *Consequence* claims are reasonable to ask if *Plausibility* is true. And indeed, over the past decade or two, something of a niche literature has sprung up devoted to the question of what to do if UMET is true. This question has been called the "Now What?" Problem (by Lutz 2014) and the "What Next?" Question (by Isserow 2017). Henceforth I shall use Lutz's term and abbreviate it to NWP. Of course, two types of NWP are possible. For any instance of LMET, we face an L-NWP. An error theory targeting natural rights leads to the question: shall we retain talk of natural rights in ethical discourse? It is hard to see any reason why this L-NWP would not be answered in the negative, as Bentham answers it. L-NWPs, then, tend not to be very interesting. U-NWP, on the other hand, looks like a bigger problem. For it concerns nothing less than the continuation or discontinuation of the practice currently recognisable as ethical discourse itself.

There are four main policy options to be found in the NWP-literature. Conservationism (Olson 2014; Isserow 2017, 2019) recommends retaining belief in morality, error-infected as it is, on the grounds that moral discourse is too useful to be jettisoned, and nothing short of full retention can secure those benefits. The main problem with conservationism

seems to be that insofar as it recommends retaining what it supposes to be false beliefs, it is epistemically unattractive. Revolutionary fictionalism recommends making a fiction of morality.[21] This comes in two forms: force fictionalism (Joyce 2001: chapter 7, 2005) and content fictionalism (Nolan, Restall and West 2005, Nolan 2005). The main problem with fictionalism appears to be that it cannot adequately accommodate moral disagreement (Olson 2011a). Abolitionism recommends jettisoning morality – jettisoning the entire practice that we currently recognise as ethical discourse, that is – and is the topic of Chapter 3. Revisionism recommends revising moral discourse so that it is no longer committed to error. Substitutionalism, which will be discussed in Chapter 4, is a form of revisionism.

§1.2.1: Who should be Interested in U-NWP?

Many of the contributors to the U-NWP literature are in fact UMET-proponents. As a result, it has the appearance of an in-house dispute. The impression a casual reader with no commitment to UMET but a general interest in metaethics might get from reading a few of these papers is that there is nothing in this literature to interest anyone who has not (yet) arrived at UMET. Only after we have arrived at UMET does U-NWP rear its head, we might think. And this idea might be reinforced by a casual glance at a few of these papers in which error theorists point out all the disadvantages of fellow error theorists' policy proposals and then advertise the advantages of their own.

But error theorists have two motivations to find a solution to U-NWP, not just one. Their first motivation – the one responsible for the "in-house dispute" appearance of the literature currently – is indeed to formulate a practical plan for responding to the truth of UMET when, as a matter of psychological fact, most of those who have argued for UMET have still found the practice of ethical thought and discourse worth saving, or at least difficult to give up. Their second motivation (though one that is occasionally ignored by contributors) is to defuse a potential objection to UMET – namely, the objection to UMET that I will press in the course of this thesis.

[21] Revolutionary fictionalism should not be confused with hermeneutic fictional (Kalderon 2005). This is not a U-NWP policy option, but is instead a rival metaethical view to UMET.

Here, for instance, is Lutz (2014), whose policy proposal will be the topic of Chapter 4:

> [A] good solution to the 'Now What' Problem serves an important role in rebutting an objection to the error theory. Many people consider the error theory to be an entirely unacceptable metaethical theory because they feel as though they must be giving up something essential in accepting it. A good solution to the 'Now What' Problem will provide a framework wherein error theorists are not forced to give up much (or anything) of importance. And thus one important source of resistance to adopting an error theory about morality can be undermined.
>
> (Lutz 2014: p. 352.)

We could dub the objection that Lutz has in mind here "The Objection from Loss". In its most powerful guise, the objection from loss starts from the claim that ethics is indispensable, and from here draws the conclusion that something is amiss with UMET.[22] This is something that is not especially well emphasised in Lutz's way of putting the objection, that UMET entails that "we must be giving up something essential in accepting it". Phrased this way, we might mistake this for the objection that we very much *want* UMET to be false, and so it has to be. Obviously, this is not a convincing argument, and I doubt it is what Lutz has in mind. Instead, the best version of the objection from loss would go something like this:

> **Objection from Loss (first pass)**:
> Ethics is practically indispensable, so even if it were found to be error-infected it would have to be rescued or reinvented. We couldn't simply get rid of it. After rescuing or reinventing ethics, we would end up with *something*, some discourse or set of commitments D. In order to do what is required of it, D would have to be broadly recognisable as roughly continuous with the old ethics. So, D is both non-error-infected and broadly recognisable as roughly continuous with the old ethics. This means that a universal error theory cannot be true of the old ethics. Only the parts excised in moving from the old ethics to D can be error-infected. But in that case, the correct error theory of the old ethics is only a version of LMET.

When the objection from loss is seen this way, the dialectically defensive significance of U-NWP for the error theorist becomes clearer. Lutz is right that the UMET-proponent

[22] Of course, Enoch (2011) appeals to indispensability concerns to establish his robust realism. The connection and differences between his argument and the objection from loss will be considered in Chapter 5.

needs to show that accepting UMET does not mean "giving up something essential": if UMET does mean this, then it will be vulnerable to the objection from loss. Lutz thinks that UMET can escape this objection because he is optimistic that a good solution to U-NWP can be found. I think that in this, he is mistaken.

§1.2.2: A Simple Rationale for *Consequence*

In my view, the objection from loss is probably the most formidable objection to UMET that there is. In fact, I think that in the final analysis it is decisive, and I shall be arguing as much in Chapter 5. But in order to see how decisive it is, we need to turn our attention to the available NWP policy options, and assess them for cogency. That is what I will be doing in Chapters 3 and 4. Thus, the rationale for *Consequence*, quite simply, is this. With *Plausibility* established, we naturally want to know more. We want to know if we should accept UMET. Because the objection from loss looks so promising, in assessing further UMET's plausibility we will want to inquire as to what the consequences for moral thought and talk would be if UMET were found to be true. That is, we should do as *Consequence* says we should do.

A Note Concerning Conservationism and Revolutionary Fictionalism

I shall not be assessing conservationism or fictionalism in this thesis. This is because I do not think that doing so is necessary to get to the conclusion that I seek to establish, viz., that the objection from loss is decisive against UMET. What I am keen to emphasise in my discussion of abolitionism and substitutionalism in Chapters 3 and 4 is the tendency of some U-NWP policy proposals to be, on closer inspection, of dubious compatibility with their own error-theoretical motivations. This tendency suggests to me that the objection from loss successfully undermines UMET. Not every U-NWP policy proposal needs to conform to this pattern in order for the tendency of some of them to do so to point towards the objection from loss. For the most striking result is that abolitionism in particular is affected by this problem. There is something distinctly odd, I will suggest in §5.0, about convicting an entire discourse of error if that discourse *could not* be given up. That substitutionalism also conforms to this tendency just amplifies the suspicion that all is not well with UMET.

So I do not need to show, make no attempt to show, and indeed almost certainly could not show, that conservationism and fictionalism also conform to this tendency. I am happy to assume that both proposals are indeed entirely consistent with UMET. In fact, since conservationists and fictionalists both tend to support their proposals by emphasising that morality is too useful to be jettisoned, there is a degree of similarity between their concerns and mine. I will claim (in Chapter 5) that *the ethical project is intrinsically indispensable*. This is not a million miles away from the fictionalist's and conservationist's claim that morality is too useful to be given up. So, both the fictionalist/conservationist and I think that engagement in the ethical project must continue. The difference is that because the conservationist/fictionalist thinks that the error-infection in moral thought and talk is so pervasive that we should accept a MET satisfying *UMET Condition*, they think that rescuing the ethical project means rescuing morality. I, on the other hand, think that the objection from loss shows that a MET satisfying *UMET Condition* is too extreme to warrant our acceptance. The most we can accept is a theory satisfying *MET Condition*. Any such theory will be an instance of LMET, not UMET.[23]

Partial Escape

§1.3

The rationale for the final claim to be established in this chapter – *Partial Escape* – is simple. The argument from obscurity pointed to two major mysteries that seem to make it obscure what we mean when we make moral pronouncements: the obscurity of obligation and blame, and the lack of consensus concerning foundational propositions about morality. If we find the objection from loss convincing, this will not immediately remove these obscurities. We will need to demystify, or perhaps conclude that there are

[23] Notice that this does not mean that the conclusion of this thesis will be disjunctive, to the effect that either UMET is to be rejected or conservationism or fictionalism must be the preferred U-NWP solution. The conclusion is just that UMET is to be rejected. As will become clear in Chapter 5, those UMET-proponents who favour fictionalism or conservationism as their U-NWP policy still cannot endorse my claim that the ethical project is *intrinsically indispensable* (see §5.12.). For this term has a specific meaning, and so the claim goes beyond fictionalists' and conservationists' claim about usefulness.

certain things that we *appeared* to be committed to in moral thought and talk, but never in fact were. If we can at all help it, we should avoid being to committed to that which we recognise to be obscure. (This the rationale behind premise (3) in AfO.)

If what we want from a metaethical theory is an account of what we are doing when we judge ethically, and an account of why we are doing something other than utilising an error-infected discourse or buying into obscure commitments that cannot even be articulated, then the sorts of obscurities that form the material of the argument from obscurity cannot just be ignored. Even if we are convinced by the objection from loss that UMET is not true, we need some sort of account of why those seeming obscurities are nothing of the sort, or how ethics could be purged of them.

I do not think that the prospects for successfully executing the first of these tasks are rosy. I think the obscurities canvassed in this chapter are real obscurities. In that case, we are stuck with some LMETs that cannot reasonably be rejected. I shall try to show, however, that there is ample reason for optimism concerning our ability to navigate our way around these LMETs and actually *do some ethics without error*.

An Objection

§1.4.0

Over the course of §1.1.1, I gave an exposition of Joyce's AfR. Part of this argument is the presupposition premise that the notion of moral obligation presupposes the notion of NICRs. One prominent critic of this presupposition premise is Stephen Finlay. In this part of the present chapter, I want to consider in some detail Finlay's criticisms of this presupposition premise, and Joyce's response to it. The conclusion I aim to establish is:

> *Stalemate*: The exchange between Joyce and Finlay ultimately terminates in a stalemate, with no easy method of determining who is right concerning the presupposition premise of AfR. Both views seem plausible.

There are two reasons why it is important to discuss and respond to this objection. First, if Finlay is right, then the main strand of AfO is undermined, and thus the case for *Plausibility* is significantly weakened. However, if I can successfully establish *Stalemate*,

then the damage done by this objection to case for *Plausibility* can be significantly limited. Enough of the case for *Plausibility* survives that it remains worthwhile investigating the practical consequences of UMET's adoption, as I will doing from Chapter 3 onwards. Second, some of the argument against substitutionalism to be pressed in Chapter 4 amounts to a repurposing of Finlay's argument. I will claim greater force for these sorts of considerations in that context than I think they have in the context in which Finlay makes them. So, it will be useful to have an appreciation of the argument in its original context ahead of time.

§1.4.1: Finlay's Criticism of Joyce's Presupposition Premise — Uncharitable Interpretation

Finlay (2008) argues that even if ordinary users of moral discourse are committed to the existence of NICRs as Joyce says they are, it is a mistake to think that this commitment infects the meanings of moral terms. Just as an erroneous commitment to absolute motion does not involve a speaker in error when they say that a boat has moved, so a mistaken view about the authority of morality does not involve a speaker in error when they say that murder is morally wrong.

For the purposes of comparison with the moral case, we have four non-controversial examples of erroneous suppositions in play at this point, two of which scupper their respective discourses, and two of which do not. The error theoretical examples are witch-discourse and phlogiston-discourse; here the erroneous assumptions infect the meanings of the terms 'witch' and 'phlogiston' such that statements like 'Mary is a witch' and 'phlogiston is being released here' are wrong. The non-error-theoretical examples are motion and water. I discussed the example of motion in §1.1.1.1.

With respect to water, the erroneous supposition is that water is an element; this does not prevent speakers making true statements to the effect that water satiates thirst, is vital for life, and can be found flowing in streams. The question is: on which side of the divide does morality fall?

We know Joyce's answer, of course. The answer is that we have to consider what the *point* of the discourse in question is. In the case of witches, the whole point was to refer

to women with supernatural powers. In the case of motion, the commitment to absolute motion was incidental to the main point of the discourse. With morality, the *whole point* is to talk about inescapably normative reasons that do not depend on agents' desires. So morality falls on the side of witches and phlogiston and not on the side of water and motion.

Finlay sensibly proposes that to answer the question, we need to turn to essential application conditions. He makes the following claim about essential application conditions on the concept **MORALLY WRONG**:

> *Relational Moral Application*: the term "morally wrong" is applied to actions that frustrate certain ends or violate certain standards.

He then says:

> After all, the absolutist about motion makes (what look like) substantially the same first-order motion judgements as the rest of us, and the absolutist about morality makes (what look like) the same first-order moral judgements as the rest of us. Assuming there is no genuine absolute motion, or genuine absolute moral properties, the absolutist's judgements could not be responsive to these fictional properties. Rather, his judgements about motion are responsive to his sensitivity to motion relative to particular frameworks, and his judgements about moral wrongness are responsive to his sensitivity to the relation of actions to certain moral standards or ends. It is because of this that we rightly attribute relational moral and motion concepts and terms even to the absolutist, and justifiably claim that he misunderstands his own language and thought.
>
> (Finlay 2008: p. 365.)

However, this does not get us all the way to answer to the question concerning whether morality belongs with witch-discourse and phlogiston discourse or motion-discourse and water-discourse. For the quoted argument is vulnerable to an obvious objection, in response to which Finlay needs to apply a fix. The objection comes from the inconvenient fact that people did once use/apply the term "witch". The application of the concept witch presumably did track some sort of real property, such as being a woman whose enemies had suffered illness and/or misfortune. So, couldn't the argument simply be re-run with the concept witch, so as to avoid an error theory of witches? We'd get this:

> Assuming that there are not genuinely any women with supernatural powers, the believer in witches could not be responding to *this* property in making his witch judgements. […] Rather witch judgements are responsive to the sensitivity to a complicated disjunctive property including, for example, the property of being a woman whose enemies have suffered illness and misfortune. […] It is because of this that we rightly attribute [the] concept of *witch* that have realisations in the actual world even to people whose theory of witches … construes these things such that they have no actual realisations.
>
> (p. 366.)

The fix is to distinguish between application of concepts on the basis of sensitivity to that which we take to be *evidence* of their instantiation, and application of concepts on the basis of sensitivity to that which we take to be *constitutive* of their application.[24] Competent users of the term "witch", Finlay says, will easily be able to conceive of situations in which all the usual evidence of witch-hood obtains, and thus points to a given woman's being a witch, but in which she is nonetheless not a witch. Likewise, they will easily be able to imagine the converse situation, in which despite the absence of evidence a given woman is indeed a witch.

At this point, the AfR-proponent might think that they can pull the same move with respect to morality. Can't the standard user of moral terms imagine a scenario in which all the standard evidence that something is morally wrong obtains, but in which that thing is not morally wrong? Or vice versa? "These moves," Finlay writes, "are not nearly as plausible as in the case of witches and phlogiston. Consider that the 'evidence' in each case [motion and morality] is the relational property that the relational theory identifies with morality or motion" (p. 366). This last claim merits a name:

> *Evidence:* the evidence on the basis of which the standard user of moral terms infers that "morally wrong" applies to a given action is that the action frustrates certain ends or violates certain standards.

It is now that we get the charge of gratuitous uncharity:

[24] Some might have doubts as to whether this is the correct way to draw the contrast. Finlay seems to have distinguished between an epistemic claim and a metaphysical claim. This is legitimate because ***Relational Moral Application*** is supposed to capture *essential* application conditions. It is thus a metaphysical claim: it tells us what it is for an action to be "morally wrong". Likewise, that the concept **WITCH** is to be applied to women with supernatural powers is a metaphysical claim; it tells us what it is for a person to be a witch. This is to be distinguished from claims telling us about reliable ways to infer that someone is a witch.

> [E]rror theory now seems gratuitously uncharitable: it claims that ordinary judgements of moral wrongness [...] track awareness of actual value relative to moral standards [...], but are nonetheless systematically false because they take the real thing to be merely evidence of fantastical counterparts.
>
> (p. 366.)

§1.4.2: Joyce's Unsuccessful Reply

Joyce complains that ***Evidence*** is not correct. It is not that φ-ing violates certain standards or thwarts certain ends that is the basis on which the standard user of moral terms infers that φ-ing is morally wrong; it is on the basis of such things as the causing of suffering, the presence of selfish motives, etc. Parity is thus restored between the witch case and the morality case, as Joyce takes the table in Fig 1 to demonstrate.

	The evidence	The (possibly) instantiated property	The uninstantiated property
WITCHES	e.g., testimony about a woman's suspicious behaviour, accusations from upstanding citizens, coerced confessions, etc.	'a complicated disjunctive property including, e.g., the property of being a woman whose enemies have suffered illness and misfortune' [366]	being a woman who can cast spells and has entered into a pact with the devil
MORALITY	e.g., the presence of suffering and selfish intentions, etc.	non-absolute relational property pertaining to the action's bearing on the speaker's ends	absolute moral property

Fig 1: Joyce's comparison of witch-discourse and moral-discourse (Joyce 2011: p. 532).

However, this reply ultimately fails, because the "concrete evidence" (as Finlay calls the moral evidence in the first column of Joyce's table) is so disparate and diverse. The only thing unifying such concrete evidence, Finlay says, is that "it is evidence for value relative to various salient ends of concern" (2011: p546). Thus, the error theorist "cannot plausibly deny that relational value commonly […] plays a significant evidential role" (ibid). This is not the case with witch-judgements, where the instantiated property in the second

column is gruesomely disjunctive, such that it "isn't plausible that ordinary people's witch judgements are tracking such an uninteresting, gerrymandered property, and we can safely presume that they rather infer from the concrete evidence the interesting but uninstantiated property of being a witch" (ibid).

§1.4.3: A Better Reply?

Joyce takes the defence of ***Evidence*** to be the "pivotal moment in Finlay's argument". But we only arrive at ***Evidence*** if we first accept ***Relational Moral Application***. Joyce does appear to reject ***Relational Moral Application***, too. The trouble is that this rejection is glossed over rather quickly:

> Finlay claims that just as 'the absolutist about motion makes (what look like) substantially the same first-order motion judgements as the rest of us [who are sensible motion relativists], … the absolutist about morality makes (what look like) the same first-order moral judgements as the rest of us'. But this is precisely the claim about which I am doubtful. And while it may be true that I haven't said enough to ground this doubt persuasively, by the same token Finlay hasn't said enough simply to help himself to the claim.
>
> (Joyce 2011: p. 530.)

The claim that the absolutist about morality makes the same first-order moral judgements is offered by Finlay as evidence to support ***Relational Moral Application***. I take it that here Joyce aims to reject ***Relational Moral Application***, instead thinking it is part of the essential application conditions of the term "morally wrong" that any agent has a NICR not to do that to which the term is applied. But Joyce says nothing more than this about ***Relational Moral Application***. I think that this is where we should concentrate.

To support Joyce's claim that absolutists (here meaning those who presuppose the inescapable authority of morality) make different first-order judgements to relativists (here meaning those who make no such presupposition), we need examples of one or both of two types. Helpful first would be examples in which a moral judgement is made, even though none of the standard evidence obtains. The phenomenon of moral condemnation based on pure prejudice and/or irrational feelings of disgust or contempt plausibly provides such an example, but a better example might be Kant's famous insistence that the last murderer of a soon to be defunct society be put to death. Here the whole point of

Kant's example was that even setting the murderer free would violate no important ends. Still, Kant sees it as required to put the criminal to death. Clearly, the only thing that can account for this judgement is Kant's enthusiasm for the idea of inescapable authority.

The second type of example that would help Joyce to substantiate his claim that absolutists make different judgements would be those in which all the standard evidence that a course of conduct violates ends held dear to the speaker obtain, but there is no moral condemnation. Popular opinions concerning the punishment of criminals, easily observable in the letters pages of tabloid newspapers, plausibly fit into this category. Even when in possession of highly reliable evidence that overly harsh punishments have a negative impact on recidivism rates, such that we should be confident that more humane conditions in prisons, for instance, would have an overall positive impact on society as a whole (and not just the lives of prisoners), readers of such publications still support the impositions of great harms on those who commit crime, out of a sense of desert. It is hard to explain why people would hold such obviously incorrect ethical opinions[25] without appealing to their sense that their unattractive retributive sentiments ground some sort of inescapable normativity.

§1.4.4: Finlay's Response to The Better Reply

Finlay anticipates the reply I have just given:

> The error theorist will likely resist the claim that the essential application conditions for ordinary moral concepts are relational in the way I have suggested. Mackie and Joyce both take the view that these conditions include the presence of genuinely practical reasons […] They suggest, in other words, that competent use of moral concepts entails the following: if a monster like Fred West were (perhaps *per impossible*) to have no genuine reasons that made it irrational for him to perform his crimes of child abuse, rape, and murder, then those acts could not coherently be considered morally wrong for him to perform.
>
> (Finlay 2008: p. 367.)

And he gives very short shrift to it, replying immediately:

[25] That is – obviously incorrect by the standards governing the term "morally wrong" on Finlay's relational analysis. I also happen to think that these ethical opinions are obviously incorrect, simpliciter. But in the current context, that is not the relevant issue.

> I think we should deny that the ordinary concept of moral wrongness entails any such thing. Even if we grant that moral judges usually assume that everyone has (and necessarily must have) genuine conclusive reasons to avoid morally wrong actions, we should not suppose that they would withhold hypothetical moral appraisals of the actions of an agent who lacked such reasons. Our moral concepts allow us coherently to contemplate the figure of the rational villain, and also coherently contemplate the question 'ought I rationally act as I morally ought?' These considerations seem to me decisive against the error theory.
>
> (ibid.)

It is extremely odd that Joyce doesn't push back harder than he does here. After all, the figure of the rational villain was a major source of inspiration in AfR. A major reason for thinking that the non-Humean instrumentalist account of practical reason cannot accommodate NICRs is precisely that once we relativise practical reason to ends in the manner of that account, NICRs can find no purchase on the suitably rational villain. So the consideration that Finlay finds decisive *against* UMET – that "our moral concepts allow us coherently to contemplate the figure of the rational villain, and also coherently contemplate the question 'ought I rationally act as I morally ought?'" – is actually a major part of Joyce's argument *for* UMET.

What seems to be happening, then, is that Joyce and Finlay differ in their view of the implications of our ability to contemplate the character of the rational villain. Joyce says: to claim that A morally ought not φ entails that A has a reason not to φ. So, it is incoherent to say that A morally ought not φ but that A might have no reason not to φ. Finlay says: to claim that A morally ought not φ entails that A has a reason not to φ relative to some moral aim/perspective. This is consistent with the possibility that A morally ought not φ and yet have no reason not to φ. On which side does common sense fall? Finlay will say that common sense falls on his side. But Joyce will say that common sense falls on his side: most standard users of moral discourse will say that a character like West does have a reason not to engage in acts of rape and child abuse: namely, a moral reason. Joyce will say that it takes reflection on the nature of practical rationality, of the sort Joyce provides at length in *The Myth of Morality*, to see that there is no sense to be made of such reasons.

§1.4.5: Another Reply

Of course, the reply I formulated in §1.4.3 is very much the orthodox UMET response. It is a defence of the presupposition premise of AfR, and so aims to reject *Relational Moral Application* by offering an a non-relational account of the essential application conditions of **MORALLY WRONG**. But there is another way to reject *Relational Moral Application*, and that is to say that **MORALLY WRONG** simply *has* no essential application conditions. This is the sort of response that seems best for the UMET-proponent who has arrived at error theory on the basis of AfO instead of a more traditional argument following the "recipe" outlined in §1.1.1.1. That would certainly reduce the effectiveness of Finlay's critique as a critique of *UMET*. But the topic here is more specifically the dialectic between Joyce and Finlay: I am attempting to establish *Stalemate*. So I shall pass over this response to Finlay here. However, it is in fact the response I think a UMET-proponent would be best advised to make.

§1.4.6: Conclusion

Finlay thinks that the AfR-proponent is gratuitously uncharitable in interpreting standard users of moral discourse. But we have seen that the AfR-proponent has some resources to counter this criticism. It is not clear (to me, at least) who has the upper hand in this debate. Therefore, with respect to the tenability of *Relational Moral Application*, which Finlay accepts and Joyce rejects, I am inclined to accept:

> *Stalemate*: The exchange between Joyce and Finlay ultimately terminates in a stalemate, with no easy method of determining who is right concerning the presupposition premise of AfR. Both views seem plausible.

However, we shall see in Chapter 4 that *Stalemate* does not apply when Finlay's arguments are redeployed in a new context: that of assessing the substitutionalist's prospects for escaping the problem affecting their proposal that I call the problem of shrinking space. In that context, Finlay's argument will be seen to have greater power than it has been seen to have in the present context.

Summary and Preview

§1.5

In this chapter, I have given a reasonably detailed exposition of AfR, and suggested an improvement that can be made to it to remove its vulnerability to objections to the instrumental view of practical reason on which it is based. AfR has helped to establish *Plausibility*; the main way it has done this is not by functioning as a stand-alone argument but as a strand in the more powerful but messier AfO.

I have established *Consequence* on the grounds that the concerns of the U-NWP literature should be of interest not just to error theorists, but to those looking for good methods of stress-testing UMET. For it is possible that a particularly powerful objection to UMET is the objection from loss.

Finally, even if UMET is eventually fatally undermined by the objection from loss, this fact alone does little to dispel the air of mystery around those characteristics of moral discourse focused on in the argument from obscurity. Even after UMET is dispensed with, we will still need to resolve these mysteries somehow. This has established *Partial Escape*.

It is time to make a start on the project of inquiring as to the effects on moral thought and talk UMET would have if found to be true. The project of inquiring as to the *practical* effects begins in Chapter 3. In the next chapter, though, I consider perhaps the most basic *epistemic* consequence: if convinced of UMET, we would have to start disbelieving even the most obvious moral propositions, such as that it is morally wrong to torture human babies merely for entertainment or pleasure. This, we might think, is such a cost to UMET in terms of plausibility points that UMET is sunk immediately. I shall argue in the next chapter that this is not the case. UMET not quite so easy as that to dispatch. We need to turn our attention to the practical consequences.

Chapter 2:
Moral Moorean Arguments

§2.0

In the previous chapter, I offered the argument from obscurity as support for ***Plausibility***. This claim, recall, concerned the *initial* case for UMET, not the level of confidence we should invest in UMET in the final analysis. So even with AfO in hand, we may still wonder if UMET is really worth taking all that seriously. Perhaps the strong initial case for UMET can be easily defeated?

One way in which we might think that the case for UMET can simply be dismissed is by means of a moral Moorean argument. Moral Moorean arguments take the same strategy that Moore famously applied to arguing against scepticism about the external world (henceforth: SEW) and apply it to arguing against UMET. Moore argued against the sceptic about the external world that scepticism was obviously false because it implied that he (Moore) did not have hands.[26] But (pointing at one hand with the other) *look! I clearly do have hands!* Likewise, moral Moorean argument target UMET because UMET implies that it is not the case (say) that it is morally reprehensible to torture human babies merely for fun. But it is obvious that it *is* morally reprehensible to torture human babies merely for fun. The parallel structure of the arguments is obvious.[27,28]

In this chapter I am going to discuss and reject the possibility of dismissing UMET in this way. The relevance of this discussion to the overall project as outlined in §0 is

[26] Of course, what SEW in its most familiar form implies is that Moore cannot *know* that he has hands. I ignore this complication. The version of SEW most analogous to UMET is indeed the Berkeleian claim that there *is* no external world.

[27] Bamborough (1979) uses a moral Moorean argument of this sort.

[28] The formulation in the text is that UMET implies that *it is not the case that* it is morally wrong to torture human babies. Does UMET also imply that torturing human babies is *not wrong*, and does that imply that on UMET torturing human babies is *morally permissible*? That depends on what we say about the Doppelganger problem and the reinforced Doppelganger problem (Pigden 2007). Dworkin (2011: chapter 3) thinks that this saddles the UMET-proponent with moral commitments, resulting in an incoherent view; Kramer (2017) agrees and emphasises the distasteful nature of the moral opinions the UMET-proponent commits themselves to. Olson (2014: chapter 1) thinks that there is no conceptual entailment from "x is not wrong" to "x is morally permissible"; he thinks that the entailment is only a matter of conversational implicature.

twofold. First, in rejecting the moral Moorean arguments, I will be bolstering the case for *Plausibility* made in the previous chapter. If UMET can withstand the force of this highly intuitive objection, then the initial case for UMET is strengthened. In line with *Consequence*, we thus acquire further motivation to inquire as to the consequences of UMET's adoption, as I shall be doing in Chapters 3-5.

Second, this chapter serves a disambiguation purpose. Both the Moorean argument targeting SEW and the one targeting UMET follow the Moorean Schema:

Moorean Schema
1. If S, then not-M
2. M
3. Therefore: not-S.

Compare this with the objection from loss as advertised in §0. There I boiled the objection down to the following form: if UMET is true, then there must be some sense in which it is illegitimate to adopt the ethical point of view; there is no sense in which it is illegitimate to adopt the ethical point of view; therefore, UMET cannot be true. The schema of this argument is:

Objection from Loss – Basic Schema
1. If UMET, then L
2. Not-L
3. Therefore: not-UMET.

Although the position of the negation operator has moved, this is still a highly similar schema, and the crucial premise, that there is no sense in which it is illegitimate to adopt the ethical point of view, might at first sight appear to be the same sort of claim as that made by the proponent of the moral Moorean argument. However, it is not. As shall become clear in the discussion of the present chapter, premise (2) in the moral Moorean argument concerns epistemic rationality. The point is that we *ought to be more confident* that baby-torture is morally reprehensible than we ought to be that any of the premises of AfO are correct. In Chapter 5, it will become clear that the crucial premise of the objection from loss, i.e. that there is no sense in which it is illegitimate to adopt the ethical point of view, does not function in this way. (To advertise ahead of time: in the objection from loss, the focus is on practical rationality, not epistemic rationality.)

This chapter comprises two parts. In the first, I discuss a case put forward by McPherson (2009) against the effectiveness of Moral Moorean Arguments. As we shall see, McPherson is interested in the strength of Moorean Arguments in the moral context *relative to their strength in their original context* of responding to SEW. McPherson's conclusion is that in the moral context the effectiveness of the Moorean Argument is reduced. This conclusion depends on testing our intuitive sense that UMET must be mistaken against a number of epistemic indicators. Here, McPherson argues, our intuitive sense that UMET must be mistaken does not fare as well as our intuitive sense that scepticism about the external world must be mistaken.

In the second part I consider a potential problem for McPherson's analysis. Some commentators have argued that error-theoretical arguments against moral reasons must generalise to target epistemic reasons to just the same extent. I spend the second section defending McPherson's analysis against this threat by resisting the claim that error-theoretical arguments against moral reasons must generalise to target equally epistemic reasons.

Moral Moorean Arguments

§2.1.0: McPherson's Epistemic Indicators Strategy

As already noted, Moorean Arguments take the following form:

Moorean Schema:
1. If S then not-M
2. M
3. Therefore: not-S

(S is the sceptical thesis, i.e. SEW or UMET.[29] M is the Moorean Premise, i.e. the premise that *I have hands* in the SEW case, or the premise concerning the wrongness of torture in the moral case.)

[29] McPherson calls it the "revisionary thesis", in the sense that it requires our revising our pre-theoretical intuitions. It would be confusing to retain his terminology here because "revisionism" is already the name of a U-NWP policy option (as mentioned in §1.2.0). It is *also* the name of a metaethical view in competition with UMET – this form of revisionism will be discussed in Chapter 4.

When it comes to (2) in the context of the argument against the external world sceptic, McPherson explains, the point Moore makes is one about relative confidence. He is just more confident that M is true than in any one of the components of the case the external world sceptic makes to argue his case. McPherson takes the point not to concern Moore's degree of subjective confidence, but instead to concern indirect evidence. This is because to think the point concerned subjective confidence only would be to read Moore as endorsing a norm to the effect that you should apportion your belief to your confidence. This, however, is an "extremely dubious epistemic principle" (p. 5). Taking the point to concern indirect evidence makes for a better interpretation of the Moorean argument form:

> Insofar as one is epistemically virtuous, one's confidence in a claim will tend to track the quality of the evidence that one has for the truth of that claim. Thus, confidence may serve as defeasible indirect evidence of the truth of a claim.
>
> (McPherson 2009 p. 5.)

What ought we say about the nature of this indirect evidence? McPherson suggests that the best way to understand the epistemic force of the Moorean strategy against the external world sceptic is in terms of the "indirect evidence that we possess concerning the relative epistemic quality of the relevant Moorean and revisionary claims" (p. 6). To this end, he proposes five indicators of epistemic quality. (He is clear that this is not intended as a comprehensive list; other factors might be relevant, but the five he offers seems to him to be the most important.) These are:

> i. Relative confidence in the Moorean and sceptical theses
> ii. Prevalence of philosophically naïve proponents of scepticism
> iii. Extent and nature of the reorganisation of our beliefs required by scepticism.
> iv. Relative consilience of the Moorean premise and the scepticism with our epistemic paradigms.
> v. Vulnerability of the Moorean premise to debunking explanations.
>
> (pp. 6-7.)

What McPherson is going to do next is: (1) apply these criteria to the anti-SEW Moorean argument and show that it does well as measured against these criteria; (2) apply these

criteria to the Moral Moorean Argument and attempt to show that it performs much less well than the original anti-SEW version; (3) consider the question of how far from the ideal of the anti-SEW version a given instance of a Moorean Argument can fall before it becomes unconvincing. From here, he will be in position to evaluate the prospects of moral Moorean arguments. He thinks the prospects of success for such arguments are dim. But at the outset we might have some questions about these epistemic indicators.

§2.1.1: The Relevance of the Epistemic Indicators

First, what is the relationship between the five indicators? Do they each pull equal weight, or are some more important than others? Second, where do these indicators come from? What is their rationale? Why treat these criteria as standards to which the supporter of the Moral Moorean Argument is to be held?

McPherson doesn't say much about the first of our questions. To be fair, given that his aim is just to show that Moral Moorean Arguments have *less* force than anti-SEW arguments, he perhaps doesn't need to. As we shall see shortly, his strategy is to argue that the anti-SEW argument does very well against *every* criterion, and that there are some criteria on which its moral analogue fairs poorly, so regardless of how the criteria are weighted his argument for the conclusion that moral analogues are less effective will go through. In what follows, I more or less ignore the issue of relative weighting, and treat the criteria as simply additive. (In §2.1.2 I will speak of keeping score, with one point available for each version of the argument as measured against each criterion. McPherson does not do this explicitly, but as I read him, doing so is very much in the spirit of his argument.)

Why hold the would-be Moorean to these criteria? McPherson says little about this and here I would like to bolster his case a little. I shall provide a brief discussion of each indicator in turn. However, I will be leaving indicator (ii) until last, because with respect to this indicator, I think some people might have some doubts, and I shall try to accommodate those doubts. So, I shall go through the indicators that can be simply and amply defended first.

Indicator (i): Relative confidence

In *Taking Morality Seriously*, David Enoch provides two arguments which, taken together, are intended to establish that Robust Realism scores sufficiently highly on plausibility points that it should be our preferred metaethical theory. The arguments in question are the argument from the moral implications of objectivity and the argument from the deliberative indispensability of irreducible normative truths. The details of the arguments are not important for current purposes; the important thing is just that, like moral Moorean arguments, both these arguments contain normative premises. In the case of the argument from the moral implications of objectivity, the normative premise is moral – it concerns a principle governing when it is and is not appropriate to be impartial – and the conclusion is roughly that metaethical theories that do not accommodate the objective truth of certain moral propositions are mistaken. Since UMET is such a metaethical theory, it is targeted by this argument.

So, the argument from the moral implications of objectivity is much like a moral Moorean argument in that both purport to show that UMET is false, both contain a moral premise that the UMET-proponent will have to deny, and that both could be rejected by the UMET-proponent for precisely this reason on the grounds that the argument is objectionably question-begging. In §5.2.2 of his book, Enoch considers this potential objection. Here is part of his response:

> We all know that one person's modus ponens is another's modus tollens. Given a valid argument of this sort, then, when should you take it as a modus ponens (and proceed to accept its initially plausible conclusion), and when should you take it as a modus tollens (and proceed to reject at least one of its initially plausible premises)? I would have loved to have more to say in reply to this question, but all I have is as follows: If you are (justifiably) more confident, prior to thinking about the relevant argument, in the conjunction of the premises than in the negation of the conclusion, treat the argument as an instance of *modus ponens*. If you are (justifiably) more confident, prior to thinking about the argument, in the denial of the conclusion than in the conjunction of the premises, treat it as a modus tollens. This is just a long way of saying that given a set of propositions you believe, and a proof of their inconsistency, the one to be tossed aside is the one in which you are (justifiably) least confident.
>
> (Enoch 2011: p. 118.)

Fairly obviously, this just *is* an appeal to indicator (i) – and if Enoch's reliance on a normative premise in each of his more sophisticated arguments for Robust Realism requires him to appeal to indicator (i), then it seems obvious that the advocate of the simpler moral Moorean argument, for whom the appeal to a normative premise constitutes the entirety of the case against UMET, will likely also have something like indicator (i) in mind. So, indicator (i) presents a legitimate standard against which to test the relative performance of the anti-SEW and the moral versions of the Moorean arguments.

Indicator (iii)

An aspect of Enoch's discussion of the concern that his arguments might be objectionably question-begging that I suppressed in the discussion of indicator (i) above is this: in addition to explaining why his argument is not objectionably question-begging, he is attempting to explain why the availability of the simpler moral Moorean arguments do not simply render his more sophisticated arguments redundant. If deploying a moral premise in an argument against UMET is legitimate, then why bother with the argument from the moral implications of objectivity at all? Why not rely entirely on moral Moorean arguments? In answer to this question, Enoch asks us to imagine a Moorean argument directed at mathematical error theory. Fitting such an argument into our Moorean Schema, it would have to run: (1) If mathematical error theory is true, then it is not the case that 1+1=2. (2) It is obviously *is* the case that 1+1=2. (3) Therefore: mathematical error theory is not true.

Here is what Enoch says about this version of the Moorean Argument:

> It is hard to get excited about this argument. This is so, I think, because any remotely plausible error theory would have to explain why it is OK to continue doing mathematics … even though there are no mathematical objects. […] The [mathematical] error theorist, in other words, allows us to say things very close to the premise of that argument [i.e. that 1+1=2] … so the argument loses much of its force.
>
> (Enoch 2011: p. 118.)

This looks like an appeal to indicator (iii). The point is that the mathematical error theorist does not commit herself to a major reorganisation of her belief system. The UMET-

proponent, on the other hand, does so commit herself: she has to throw out all her moral beliefs. This represents a much greater reorganisation of her beliefs than the reorganisation of beliefs required of the mathematical error theorist.

What Enoch is making here is a relative claim of much the same sort as will be at issue in the discussion of McPherson's argument to come. McPherson is comparing the anti-SEW Moorean argument to the moral Moorean argument, and will find that the moral Moorean argument is less interesting than the anti-SEW version. One reason for this will be that the moral version performs less well than the anti-SEW version against indicator (iii). Here, Enoch is comparing the moral Moorean argument with the mathematical Moorean argument, and finds that the mathematical version is less interesting than the moral version – and his reason appears to be that the moral version performs better against indicator (iii) than the mathematical version.

In this, Enoch must surely be right. Compare: Sarah is a mathematical error theorist. She is reading Wiles' proof of Fermat's Last Theorem, a sophisticated and complicated 128-page proof. Her sceptical commitment does not prevent her from "buying in" to the train of thought expressed in the proof. A sense in which she can accept it as a proof survives her scepticism. Morag is a moral error theorist. She is reading Marquis' (1989) argument against abortion. Absent some non-abolitionist solution to U-NWP, her sceptical commitment *does* prevent her from "buying in" to the train of thought expressed in the argument. This is because a crucial claim in Marquis' case against abortion is that what explains why it would be morally wrong for you to murder me is that in so doing you would be depriving me of my future, and thus inflicting on me one of the greatest losses possible. As Marquis puts it: "killing someone is wrong, primarily because the killing inflicts (one of) the greatest possible losses on the victim" (Marquis 1989: p. 49). But of course, you cannot accept this claim if you are UMET-proponent. And thus, you cannot "buy into" Marquis' train of thought in the way that a mathematical error theorist can "buy into" Wiles' train of thought despite their commitment to mathematical error theory. And it seems that similar remarks will apply to any moral argument any commentator could possibly provide in support of any moral view whatsoever (recall *UMET Condition* from §1.1.4).

In discussing the previous indicator, my point was that anyone advancing the moral Moorean argument is highly likely *already to be explicitly committed* to the validity of the indicator. Here, the claim is that anyone advancing the moral Moorean argument *ought* to be committed to the validity of the indicator. The rational case for being so committed comes from reflecting, as Enoch does, on the relative performance of the moral Moorean argument and the analogous version targeting mathematical error theory, and considering what best explains why the version targeting mathematical error theory performs less well than the moral version.

Indicator (iv)

In §2.2 we will be considering the claim that if Joyce's Argument from Reasons successfully shows moral normativity to be conceptually problematic, then it generalises to impugn epistemic normativity too. This is a claim that forms the basis of another argument against UMET that we might think of as being very closely related to the simple Moorean Argument. This is the Argument from Analogy (Cowie 2019), also known as the Argument from Epistemic Reasons (Rowland 2013). In the form in which we will be considering it, this argument says that if the Argument from Reasons for UMET is sound, then an analogous argument targeting epistemic reasons is also sound, but an epistemic theory about reasons cannot be true, and so the Argument from Reasons is not sound. Why can epistemic error theory not be true? We shall consider these arguments in more detail in §2.2, but for the time being notice that indicator (iv) does an excellent job of articulating the obvious initial unattractiveness of epistemic error theory. Two basic complaints against epistemic error theory are that if it is true, then there can be no arguments for anything and no possibility of epistemic merits and de-merits (Cuneo 2007, chapter 4). Why would it be so undesirable to hold that there could be no arguments for anything or that there were no epistemic merits or demerits? The explanation offered by indicator (iv) is that once we find ourselves committed to this position, we have to abandon all our pre-theoretical epistemic paradigms, and this is simply too high a price to pay. Thus, we have plenty of independent motivation to take indicator (iv) seriously, and to regard it as a legitimate indicator against which to test the relative performance of the anti-SEW and anti-UMET versions of Moorean arguments.

Indicator (v)

Mention of "debunking explanations" in the metaethical literature today usually brings to mind first and foremost evolutionary debunking arguments, on which there is a vast literature. Whether or not EDAs can show anything interesting, we would do well to note with respect to indicator (v) that the type of debunking explanation *does not need to be evolutionary*. Once this is recognised, I think the motivation for indicator (v) is obvious. People do often offer debunking explanations of others' moral beliefs intended to decrease our confidence in those beliefs. Indeed, offering debunking explanations of incorrect moral opinions (e.g. resulting from cognitive inattention, vested interest) is one of the ways in which Enoch (2011: chapter 8) explains moral disagreement.

Indicator (ii)

Finally we come to the one indicator with which some might take issue.

When we come to apply the generic indicators to the anti-SEW and anti-UMET versions of the argument below, the claim with respect to indicator (ii) applied to the moral version of the argument will be that belief in UMET is not limited to philosophers, that many of us are acquainted with individuals who think that moralising is in some sense incompatible with a rugged practical realism about the way the world really operates. A defender of the moral Moorean argument might argue: why should the fact that there are philosophically naïve proponents of the view that there are no moral truths provide indirect evidence against the truth of an obvious and uncontroversial moral truth? Recall that McPherson's point is that *insofar as one is epistemically virtuous*, confidence can serve as defeasible indirect evidence of truth. The idea with respect to indicator (ii) must be to appeal to some epistemic virtue of humility. The thought, perhaps, is that if you find that your philosophical reasoning leaves you endorsing a view that *nobody else* holds, then you ought to check your reasoning carefully, for it is likely that you have made a mistake somewhere. If, on the other hand, there are others who endorse your conclusion, you have less justifiable cause to be suspicious of your own reasoning.

The moral Moorean could object to indicator (ii) in the present context, then, by arguing that the existence of philosophically naïve individuals who would deny the moral Moorean premise – which, recall, is to be an utterly obvious and uncontroversial moral

proposition such as "it is morally wrong to torture human babies just for fun" – is not something that should reduce our confidence in the moral Moorean premise any more than the existence of people who believe in astrology should reduce our confidence in our belief that astrology is incorrect. When someone professes belief in astrology, we take this as evidence of their failure to meet certain epistemic standards, not as a reason to reduce our confidence in our own beliefs. Likewise, the moral Moorean could say, when we encounter someone who claims that morality is all nonsense, we should take this to show that not everything is as it should be with the agent, not that we should revise down our confidence in even the moral propositions we take to be most obvious.[30]

I am inclined to say that considerations such as these are not especially convincing, and that indicator (ii) thus remains one we should consider legitimate. However, because discussing the issue at length would involve settling the issues related to the Doppelganger problem (Pigden 2007), and because limitations of space prevent this, I shall proceed instead in the following way. When reviewing McPherson's discussion of the relative performance of moral Moorean arguments and the anti-SEW version below, I shall consider the question of whether or not his conclusion if affected if we remove indicator (ii) and concentrate solely on the others, which we have seen are well-motivated. I shall suggest that the conclusion is unaffected. If this is correct, then there is no need to defend criterion (ii).

[30] Recall the earlier mention of the Doppelganger problem. Kramer, in emphasising the way in which UMET "teems with disconcerting moral implications" (2017: p190), emphasises forcefully something very close to the present point. "For instance," he writes, "according to the error theory, 'It is not the case that the Khmer Rouge acted wrongly when they tortured and murdered hundreds of thousands of their Cambodian countrymen' and 'It is not the case that a man who rapes and tortures and murders a woman has thereby done something morally wrong' and 'It is not the case that members of the Ku Klax Klan are morally obligated to refrain from lynching African-Americans' are straightforwardly true. Even if it were somehow true that there is no domain in which 'not morally impermissible' entails 'morally permissible', the former predicate obviously entails itself. Thus, the proponents of error theory have saddled themselves with an array of decidedly unpalatable conclusions on matters of morality, even if they forbear from any affirmative uses of moral predicates in framing those conclusions" (p. 190-191). If we agree with Kramer that UMET-proponents have indeed "saddled themselves with an array of decidedly unpalatable conclusions on matters of morality", then the natural conclusion to draw about philosophically naïve proponents of UMET is that they have either failed to realise that they have so saddled themselves, or that they are lamentably unconcerned that they have done so. In either case, it is hard to see how the fact that such philosophically naïve proponents of UMET exist does much to deprive the Moorean premise of the moral Moorean argument of dialectical force. So, indicator (ii) appears to be on shaky ground.

§2.1.2: The Application of the Epistemic Indicators

This section will be split into five subsections. In each subsection, the indicators will be applied either to a specific Moorean argument, in order to assess its dialectical power against the sceptical thesis it targets, or to a comparison between two Moorean arguments, to assess their relative merits.

§2.1.2.1: SEW vs. The Anti-SEW Moorean Argument

The first job is to apply the epistemic indicators to the original Moorean argument, which targeted SEW. Here, the first indicator strongly favours the Moorean over the sceptic: anybody can surely be more confident in the proposition that they have hands than they can have in SEW. Taking this as *indirect* evidence for (2) in *Moorean Schema*, McPherson says, explains why appealing purely to subjective confidence seems at first sight to explain the epistemic force of the argument, and also why on closer inspection such an explanation is wanting. (i) gives me very limited evidence that (2) in Moorean Schema is true. This evidence could be undercut.

Next, consider the disputed indicator (ii). This, McPherson says, also strongly favours the Moorean over SEW, since nobody outside of a philosophy seminar room has taken SEW seriously. (On reflection, this is unlikely to be true. Surely there have been people who, as a result of watching the film *The Matrix* for example, have been moved to take SEW seriously. However, we could still say, quite plausibly, that SEW has probably tempted *far fewer* philosophically naïve individuals than has UMET.)

Regarding indicator (iii), McPherson claims that SEW would force upon us a huge reorganisation of belief. Abandoning belief in the external world "calls into question virtually all our factual judgements about the world". It is a similar story with criterion (iv): SEW "threatens to undermine most of our ordinary epistemic paradigms concerning the day-to-day management of belief, and also the status of our best scientific theories as methodological paradigms, at least on the assumption that those theories purport to describe elements of the external world" (p. 10). Lastly, my belief that I have hands doesn't look especially vulnerable to any obvious form of debunking arguments.

In sum, then, every single indicator favours the Moorean over the external-world sceptic. What is the upshot of this? It is that the epistemic indicators approach offers "a natural explanation of the appeal of the canonical Moorean arguments: these arguments exemplify a nearly-best case scenario relative to the five indicators" (p. 12). And this is so whether or not we include (ii). So we have:

> **Conclusion 2.1.2a**: If the anti-SEW Moorean argument is successful, the epistemic indicators explain why.

Next, McPherson considers how well the moral Moorean argument performs when measured against the gold standard established by the anti-SEW version.

§2.1.2.2: Anti-SEW Moorean Argument vs. Anti-UMET Moorean Argument

With regards to the first indicator, we have a similar story as we had with the canonical version: we can be much more confident that it is morally wrong to torture human babies merely for fun than we can be that UMET is true. However, thereafter, the comparison between the anti-SEW argument and the anti-UMET argument doesn't look favourable to the moral Moorean. The main claim relating to indicator (ii) was mentioned above: that UMET has adherents outside of philosophy, that some people see belief in morality as incompatible with a realistic conception of actual human interaction. In addition, McPherson observes that most moral propositions are considerably less obviously true than the claim that torture for fun is wrong. Perhaps the consensus in this case is accidental? Next consider indicator (iii). Here, McPherson argues that the moral Moorean argument fares badly relative to its anti-SEW counterpart because UMET "requires sweeping revision to a merely local part of our belief system: our moral beliefs, and beliefs about morality" (p. 11). Relatedly, for indicator (iv): UMET "fails to threaten our epistemic paradigms as deeply as global scepticism" since "beliefs about our perceptual access to medium-sized dry goods, the legitimacy of induction and the deliverances of physics would all presumably remain unaffected by the acceptance of metaethical error theory" (p. 12). Lastly, indicator (v): in the case of moral beliefs, many debunking explanations of why we have moral beliefs are possible. Two of McPherson's examples: an explanation of the genesis of moral judgement through emotional responses might be

taken to undermine the validity of the judgements; some moral judgements might be explained in terms of distorting psychological pressures.

With the indicators now all applied to both versions of the argument, let us count the score. Indicator (i): both the anti-SEW and anti-UMET did equally well. Each score a point. Indicators (ii) through (v): a clear victory for the anti-SEW version. So, if we are including indicator (ii) in our analysis, the score is 5-1 to the anti-SEW version. If we are leaving indicator (ii) out, then the score is 4-1 to the anti-SEW version. Either way, it is a clear victory for the anti-SEW version. Thus, McPherson has his conclusion: moral Moorean arguments do not fare as well when assessed against the general epistemic indicators as does the original argument against SEW.

> **Conclusion 2.1.2b**: Moral Moorean Arguments do not score as well against the epistemic indicators as the anti-SEW version does. In particular, moral Moorean arguments are less persuasive with respect to indicators (ii), (iii), (iv) and (v).

§2.1.2.3: Scepticism About solidity vs. Moorean Argument Targeting such Scepticism

How far short of the gold standard set by the anti-SEW version can a Moorean argument fall and yet still remain compelling? Certainly, *some* Moorean arguments fail. McPherson gives the following example of a clearly unsuccessful Moorean argument:

> **Brick**
> 1. According to contemporary physical theory, this brick is not solid, but is rather mostly made up of empty space.
> 2. But [raps knuckles on brick for emphasis] this brick is solid as could be. No empty space here.
> 3. Contemporary physical theory is mistaken.
>
> (pp. 14-15.)

The indicators are of use in explaining why Brick fails. Consider indicator (iii): the revision to our beliefs required by contemporary physical theory is very local, affecting only our views about solidity. Also, the epistemic paradigms of indicator (iv) favour the revisionist thesis: "since Galileo we have learned to accord the results of increasingly robust experimental methodology epistemic priority over our intuitive interpretations of physical phenomena" (p. 15). Furthermore, a debunking explanation of the candidate

Moorean premise in Brick is available, viz that the proponent of Brick has conflated the properties of being solid and having no visible holes.

This gets us:

> **Conclusion 2.1.2c**: *Brick* fails. It does so because indicators (ii), (iii), and (iv) all favour the sceptic about solidity over the Moorean.

§2.1.2.4: Mereological Nihilism vs. Moorean Argument Targeting Mereological Nihilism

Now McPherson turns his attention to a slightly less clear-cut case: an argument against mereological nihilism, the candidate Moorean premise in this instance being that a certain table is a complex object:

> **Table**
> [1] According to mereological nihilism, the only material objects that exist are simples: objects with no parts.
> [2] This table exists, and it is pretty clearly a complex material object: the legs even detach.
> [3] Therefore: mereological nihilism is false.
>
> (p. 18.)

With respect to the first two indicators, ***Table*** seems to fare well: most of us can find objects that we are more confident are complex than we are that mereological nihilism is true, and the view that simples are the only material objects that exist is not one with an impressive following outside of philosophical circles. Nonetheless, ***Table*** seems also to fail, and the remaining three indicators seem to do a good job of explaining why. When it comes to indicator (iii), mereological nihilism doesn't appear to require much in the way of revision; this is because there is an easy way to translate our intuitive talk of complex objects into something acceptable to the mereological nihilist: instead of talking of a complex object F (the table), we talk of simples arranged F-wise (objects arranged table-wise). Given the availability of this translation, ***Table*** also seems to fare poorly with respect to indicator (iv): there is little or no threat to our existing epistemic paradigms. Scientific enquiry can go on much as before. Lastly, there is a debunking explanation on offer in the case of ***Table***. This is that it is not clear how well we *understand* the Moorean premise of the argument and the revisionist alternative. At issue in the Moorean argument against mereological nihilism is how the Moorean premise that *there is a complex object*,

namely a table, here fares against the revisionist thesis that *there is not a complex object here*. McPherson says: "we are certainly confident that there are tables around *rather than*, say, cleverly disguised Martians or nothing at all. Perhaps we mistake this contrastive confidence for confidence that there are tables *simpliciter*, because we are simply not used to thinking about mereological hypotheses" (p. 16-17).

So, now we have:

> **Conclusion 2.1.2d**: *Table* fails. It does so because indicators (ii), (iii), and (iv) all favour the mereological nihilist about solidity over the Moorean.

§2.1.2.5: The Fate of the Moral Moorean Argument

Conclusion 2.1.2b was not that moral Moorean arguments fail. It was that they are less persuasive than the anti-SEW version. But *Conclusion 2.1.2d* was that *Table* fails. This is an important conclusion, McPherson says, because *Table* "bears important similarities to our moral Moorean argument" – by which he means, of course, that the very same factors that deprive *Table* of dialectical force against the mereological nihilist also limit the moral Moorean argument's dialectical force compared with that of the anti-SEW version. These parallels, McPherson argues, "suggest poor prospects for Moorean arguments in ethics" because "the very features which explained the epistemic inadequacy of *Table* are also in place in the moral case" (p. 19).

Thus, we have:

> **Conclusion 2.1.2e:** The prospects are poor for moral Moorean Arguments.

Even now, we do not quite have the conclusion that moral Moorean arguments definitely fail. Perhaps they come close enough to the gold standard set by the canonical argument after all. But the prospects appear dim, and this is enough for the strategy of rejecting UMET at the outset based on a moral Moorean argument to fail. As I said in §2.0, this provides indirect support for *Plausibility*. And so, in line with *Consequence*, there are good reasons to turn to the project in which I shall be engaged in Chapters 3-5 – that of inquiring as to what follows for ethical discourse if UMET is true.

The Argument from Analogy

§2.2.0: The Threat to the Generic Indicators Strategy from Parity

Here is our current situation. As a result of considering McPherson's assessment of the performance of Moorean Arguments against four or five epistemic indicators, we have arrived at the conclusion that Moral Moorean Arguments are less persuasive than their anti-SEW counterparts. But now consider the following claim:

> **Parity:** If the Argument from Reasons for UMET is sound, then an analogous argument for the epistemic error theory is also sound.

Parity is one of two premises in the Argument from Analogy against UMET (Cowie 2019). The other premise in the argument from analogy is that epistemic error theory is not true. This latter premise is supported by arguments to the effect that epistemic error theory has such undesirable, counter-intuitive and obviously false consequences that it cannot be true (Cuneo 2007). For example, it is argued that if epistemic error theory is true then there can be no epistemic merits or demerits – most people's belief that the Earth is significantly more 6,000 years old is no better than the belief of a supporter of Christian Science that the earth is only around 6,000 years old. If it is the case both that Parity is true and that epistemic error theory is not true, then the result is of course that UMET is not true. There is debate in the literature about both premises. Olson (2011, 2014) accepts *Parity* but resists the argument from analogy by arguing that epistemic error theory does not have the undesirable consequences that it is commonly thought to have. Cowie's (2019) case against the argument from analogy focuses mainly on attacking Parity. This is the response I will defend here.

The argument from analogy presents a serious challenge to the epistemic indicators strategy discussed in the previous part of the present chapter. This is because the argument from analogy can be boiled down to a moral Moorean argument that seems to score very well indeed on all five indicators.

§2.2.1: The Argument from Analogy as a Moral Moorean Argument

One of the Argument from Analogy's defenders is Rowland (2013). His presents the following version of the argument:

> **The Argument from Epistemic Reasons**
> 1. According to the moral error theory, there are no categorical normative reasons.
> 2. If there are no categorical normative reasons, then there are no epistemic reasons for belief.
> 3. But there are epistemic reasons for belief.
> 4. So, there are categorical normative reasons.
> 5. So the error theory is false.
>
> (Rowland 2013: p. 1.)

(Here, premise 2 is clearly a version of Parity.)

Rowland uses the following argument to argue for premise 3:

> (A) If *S* knows that *p*, then there is some epistemic justification for believing *p*.
> (B) If there is epistemic justification for believing *p*, then there is an epistemic reason for believing *p*.
> (C) Therefore: If *S* knows *p*, then there is an epistemic reason to believe *p*.
>
> (Rowland 2013: p. 14.)

Now, generalise (C) and take the contrapositive. We get: If there are no epistemic reasons, then nobody knows anything. If we accept both this and Parity, then we will have the result that if the Argument from Reasons for UMET is sound, then nobody knows anything. Rowland's argument can thus be put in the form of a Moral Moorean Argument, with some obvious instance of knowledge playing the role of the Moorean Premise:

> **Moral Moorean Argument – Epistemic Parity Version**
> 1. If UMET is true then it is not the case that, as I think something, I can know that there is thought going on.
> 2. Clearly, I can know, as I think something, that there is thought going on.
> 3. Therefore: UMET is not true.

Of the possibility of reducing his argument to a Moorean Argument in this way, Rowland says: "if my way of articulating the argument from epistemic reasons makes it into a Moore-style argument against the error theory, this is not a problem for my argument" (p. 17). The reason he gives is that his version is not vulnerable to the objections to Moral

Moorean Arguments that others have put forward. This includes the objection from McPherson that we have been considering. The idea here is that if we consider how well the moral Moorean argument in its epistemic parity guise performs against McPherson's epistemic indicators, we will find that the epistemic parity version of the moral Moorean argument (henceforth: EPM) actually fares better than the original anti-SEW version. Let's now review Rowland's assessment on this score. (Indicator (i), I assume, remains a draw with both the anti-SEW argument and EPM scoring a point.)

Indicator (ii): Recall that McPherson notes that some philosophically naïve people are attracted to moral scepticism, but fewer are attracted to EWS. Presumably, fewer still are attracted to epistemic error theory and to the claim that you cannot know, as you think something, that there is thought occurring. So EPM does better than the anti-SEW argument against indicator (ii). This is not a draw, and the point is awarded to EPM.

***Indicator (iii)*:** It does seem that accepting Epistemic Error Theory would require quite radical and sweeping changes to our beliefs, because currently, "a great deal and wide variety of our beliefs are based on what we regard as knowledge", (Rowland 2013: p. 20). Take for example the belief, formed upon counting your change, that the shop assistant has given you the correct change. This belief must surely be based on the belief that you have some elementary knowledge of arithmetic. Again, EPM outperforms the original argument, and wins the point.

***Indicator (iv)*:** We tend to think that scientists have methods of acquiring knowledge, that mathematical theorems can be proven, etc. Epistemic Error Theory casts all of this aside. Once again, a point is awarded to EPM over the original anti-EWS argument.

***Indicator (v)*:** No debunking explanation seems forthcoming for the belief that, as I think something, I can know that there is thought occurring. Here, it is a draw; in the absence of any debunking arguments on either side, let's say that no points are scored in this round.

The final score, then, is 4-1 to EPM, or 3-1 if we are dropping indicator (ii). In either case, it is a clear victory to EPM over the original Moorean Argument. But that argument, recall, was the standard against which the Moral Moorean Argument was to be measured.

So we seem to have precisely the opposite to *Conclusion 2.1.2b* and should therefore reject *Conclusion 2.1.2e*.

§2.2.2: Is Parity True?

The Argument from Analogy in general, Rowland's Argument from Epistemic Reasons in particular, and the recasting of the argument from epistemic reasons as EPM more specifically still, are obviously only as persuasive as *Parity*. If *Parity* is not correct, then the Argument from Analogy isn't successful in any of these forms, and the threat to *Conclusion 2.1.2e* is neutralised. In this sub-section, I will follow Cowie in arguing that *Parity* is not correct.

Rowland makes his case for Parity by first noting that moral reasons for action and epistemic reasons for belief are alike in being categorical:

> It seems that the fact that there are dinosaur bones around is a reason for everyone to believe that dinosaurs once roamed the earth, regardless of whether they want to believe this or not. In general, two agents in the same epistemic situation – that is, with the same evidence and background beliefs – seem to have [identical reasons R to believe p for any proposition p, with possible differences in personal goals making no difference to R]. So, it seems that the moral error theory's *ontological component*, the claim that there are no categorical reasons, entails that there are no epistemic reasons for belief as well as no moral reasons for action.
>
> (Rowland 2013: p. 3.)

At the outset, we would do well to distinguish more sharply than Rowland does in this paragraph between categoricity and genuine normativity. As we saw in §1.1 when discussing Joyce's Argument from Reasons for UMET, it is the latter that Joyce takes to be the problematic commitment. Joyce's view, recall, can easily accommodate categorical reasons of the institutional variety – there was a categorical but institutional reason for Celadus not to throw sand in his opponents' eyes. Thus, whilst it is true that if two agents A and B have identical evidence and background beliefs, then if A has an epistemic reason to believe *p*, B has the exact same reason, and whilst it is true that B's having this epistemic reason does not depend on the details of B's personal goals and interests, still all of this is compatible with epistemic reasons being a species of institutional reason. Compare:

> (ER): According to the institution of epistemic rationality, the evidence is such that Henry has a reason R_1 to believe *p*. This reason is categorical. It applies to Henry in virtue of the standards of reasonable belief. Regardless of how detrimental to Henry's practical interests belief that *p* would be – perhaps he is much happier believing ~p – the fact remains that reason R_1 to believe *p* applies to Henry.
>
> (GR): According to the rules of gladiatorial combat, Celadus has a reason R_2 not to throw sand in his opponent's eyes. This reason is categorical. It applies to Celadus in virtue of the standards of gladiatorial combat. (It is true that he is an unwilling competitor in gladiatorial combat, but that does not change the fact that he *is* a competitor and so the rules *apply* to him.) Regardless of how detrimental to Celadus' practical interests it would be to refrain from throwing sand into his opponent's eyes – he will die if he doesn't – the fact remains that reason R_2 not to throw sand applies to Celadus.
>
> (MR): Julie has a moral reason R_3 to rescue the child who has fallen in the water. This reason is categorical. It applies to Julie in virtue of the child's need. No matter how detrimental to Julie's interests intervening would be – perhaps doing so would make her late for a crucial appointment and significantly damage her career, even in the long term – the fact remains that reason R_3 applies to Julie.

R_2 and R_3 are alike in being categorical, but differ in that R_3, according to believers in morality, is genuinely normative. Thus, R_3 is a target of Joyce's Argument from Reasons whilst R_2 is not. What Rowland points out in my most recent quotation of him is that R_3 and R_1 are alike in being categorical. This is presented as a motivation for Parity. But if R_3 are R_2 are alike in being categorical, and UMET targets R_3 without targeting R_2, then the fact that R_3 and R_1 are alike in being categorical does not provide a motivation for Parity. For all that has been said so far, R_1 and R_2 could be alike in being categorical but merely institutional.

We should consider, then, whether a convincing case against Parity can be made along these lines. Cowie (2019: chapter 3) presents such a case. He is concerned to establish the following "basic rationale":

> **Basic Rationale**: Epistemic judgements, unlike moral judgements, are normative only in the sense that judgements within etiquette, fashion, sports and games, and the law are normative.
>
> (Cowie 2013: p. 57.)

Cowie sees *Basic Rationale* as grounding the following view:

> **Epistemic Institutionalism**: Epistemic reasons for belief are a species of merely institutional reason.
>
> (p. 59.)

In committing himself to epistemic reasons being a species of merely institutional reason, Cowie is not depriving himself of the ability to say that epistemic reasons are categorical. Recall the example of Celadus' having an institutional reason not to throw the sand. Although this reason was not normative for Celadus, it was nonetheless categorical: there was no escaping the fact that the reason applied to him.

Cowie's argument for epistemic institutionalism involves identifying two ways in which the characteristics of institutional reasons and genuinely normative reasons diverge, and arguing in each case that epistemic reasons have the characteristic associated with institutional reasons, and thus differ in this respect from moral reasons. Thus, Parity is resisted. The two dimensions on which moral and epistemic reasons are to be compared concern: (i) the ordinary explanation of why we should form our beliefs in accordance with our epistemic reasons; (ii) our reasons (or lack thereof) to adopt beliefs in trivial truths.

§2.2.2.1: Ordinary Explanation

Let's begin by considering institutional reasons. Often we really ought to act in accordance with our institutional reasons; they are genuinely normative. Where an institutional reason is genuinely normative, what explains its being so? Cowie draws our attention to two properties of ordinary explanations of the normativity of institutional reasons: one negative, one positive. The negative property is that in providing explanations of why we ought to follow an institutional reason, in those instances in which it really is the case that we should, we do not explain the normativity in terms of reasons that belong to the institution itself. If someone asks me why they should act in accordance with the law, it will be no explanation at all to offer *legal* reasons for compliance. The positive property is that "we *would* explain why, if it all, one ought to follow such norms in terms of other norms" (p. 61).The obvious places to look are to prudential reasons or to moral reasons. Thus, we might reply to the question "why obey the law?" by mentioning the potential costs and ramifications of non-compliance (being sued or

criminally prosecuted), and/or by appealing to moral reasons (perhaps concerning the moral desirability of a state of affairs in which there is a well-functioning legal system that is generally obeyed, perhaps concerning the moral grounding of the particular law in question).

Cowie's claim is that epistemic reasons are institutional in just this way. Suppose I possess some evidence, *e*, that strongly points towards *p* over *q*, where *p* and *q* are two hypotheses I have been considering. Why does *e* constitute a normative reason for me to believe *p* and not *q*? Cowie's answer: forming my beliefs in this way, making sure that I have warranted beliefs, is a good way to get what I want. (A good way, that is, to get *anything* that I happen to want. It is quite a general truth that warranted beliefs tend to equip a person for success in any endeavour better than unwarranted ones.) As Cowie puts it:

> It's not just that we *could* cite the practical utility of following epistemic norms in explaining why, if at all, one ought to follow them. It's that this seems an overwhelmingly obvious thing to do. It would be really strange *not* to do so.
>
> (Cowie 2013: p. 63.)

These considerations lead us to the following claim of Cowie's:

> **Ordinary Explanation (epistemic)**: The ordinary explanation of why, if at all, one ought to follow epistemic norms is the same as the ordinary explanation of why, if at all, one ought to follow the norms of etiquette, fashion, sports and games, and the law. It is that there is a moral or prudential reason to do so.
>
> (p. 66.)

2.2.2.2: Trivial Truths

Institutional reasons are the sorts of things that do not generate normative reasons in the absence of moral or prudential reasons to follow them. Cowie thinks that epistemic reasons can be seen to be institutional in this way by reflecting on cases of trivial truths – i.e., insignificant truths that are wholly irrelevant to our concerns, such that instrumental rationality makes no demands on us to believe these truths even when we have evidence for them. He appeals to the following example from Leite:

> I'm standing next to a door at a convention centre. I idly notice that all of the many people I've seen come out of the door have been accompanied by dogs. I am certainly not being irrational if fail to form the belief that the next person to come out of the door is likely to be accompanied by a dog. More generally, as I go through my day, I gain all sorts of evidence supporting all sorts of beliefs. But I don't form most of them, and it hardly seems plausible that I have any reason to do so, given that they are about matters of complete indifference to me.
>
> (Leite 2007: p. 458)

Insofar as the agent in this example possesses evidence that the next person to walk out of the door will have a dog with them, they have an epistemic reason to believe that that the next person to walk out of the door will have a dog with them. But does the agent have a normative reason to form this belief? Cowie thinks not. So, we have an epistemic reason that does not generate a normative reason. In this sense, epistemic reasons are merely institutional. Hence:

> **Trivial Truths (epistemic)**: Epistemic norms, much as the norms of etiquette, fashion, sports and games, and the law, are the kinds of things that, unless there is a moral or prudential reason to follow them, don't generate genuinely normative reasons.
>
> (Cowie 2019: p. 66.)

2.2.2.3: Institutionalism and Instrumentalism

Many will find Cowie's argument for *Trivial Truths (Epistemic)* rather swift. In the paper from which the dog convention example is taken, Leite was responding to Kelly's (2003) criticisms of epistemic instrumentalism. Are epistemic instrumentalism and institutionalism the same view? Is institutionalism a species of instrumentalism? I don't think this taxonomic question has an easy answer (see fn. 33). Certainly Cowie himself defends instrumentalism as well as institutionalism.[31] Still, the important issue here is less the precise nature of the connection between instrumentalism and institutionalism, and more that instrumentalists are often criticised precisely for giving the wrong analysis of trivial truths, and this might seem to weaken the credentials of his support for *Trivial Truths (Epistemic)*. I want now to argue for the claim that even if the arguments in the

[31] For his defence of instrumentalism see Cowie (2014). For other presentations of the view see: Kornblith (1993, 2002); Goldman (2002); Alston (2005); Sharadin (2016); Grimm (2009).

literature concerning instrumentalism are successful in their original context, they do not target institutionalism in the current context of assessing whether EPM is successful against UMET.

What is epistemic instrumentalism? One way of understanding it is this. Instrumentalists offer an answer to the question: why is it that evidence for a proposition constitutes a reason to believe it, and that evidence against a proposition constitutes a reason not to believe it? The question is metanormative: it is taken as common ground between instrumentalists and their opponents that evidence for a proposition does constitute a reason to believe it, and what we want to know is what explains the reason's normativity (Buckley 2020: p. 9297). The answer that instrumentalism offers "is that there is reason to believe in accordance with one's evidence because this an excellent means of fulfilling the goals that one has, or should have" (Cowie 2014: p. 4003). Closely related are teleological views, which claim that whether or not an epistemic agent has a reason to believe *p* on a given occasion "depends on the value of the result of believing that *p*, or on the intrinsic value of believing that *p*, on that occasion" (Steglich-Peterson 2011: p13).[32]

How, then, do critics of instrumentalism claim that it gives the wrong analysis of cases of trivial truths? First consider the following epistemic norm:

> **Evidence Norm**: For any given truth-related consideration that tells in favour of the truth of *p* and any agent X, if X possesses C, then C constitutes a warranting reason for X to believe that *p*.
>
> (Buckley 2020: p. 9299.)

[32] Other presentations of this type of view can be found here: Lynch (2004); Kvanvig (2003); Horwich (2006); Papineau (2013). This is different to instrumentalism: Horwich, for example, explicitly states that "we should acknowledge that true belief has a non-instrumental value—a value for its own sake" (p. 351). However, he says that there is "an important relationship" between truth's intrinsic value and its instrumental value: "It is presumably because most truths are useful in practical inference—and not merely to those individuals who discover those truths, but also to all the rest of us to whom they are communicated—that our society, simplifying for the sake of effectiveness, inculcates a general concern for truth for its own sake" (ibid). Though this explains only the sociological fact that we do find truth valuable rather than the normative fact (if indeed it is one) that truth is valuable, Horwich later suggests that when we say that truth is valuable for its own sake, we have in mind that an intrinsic concern for truth is of moral value. The plausibility of this is not the present concern: the point is that this teleological view shares with instrumentalism the view that the genuine normativity of the fact that *e* is strong evidence for *p* – the way in which is gives me a genuinely normative reason to believe *p* – derives from sources that are not strictly speaking epistemic.

Tenable norms like this cause problems for instrumentalism in the following way. Suppose I possess excellent evidence that the number of specks of dust on my computer screen is *n*. Perhaps I live with an obsessive flatmate who devotes all her free time to tracking the number of specks of dust on things around her, and who has invested in high-tech gadgets that reliably determine how many specks of dust there are on things, and perhaps she has just deployed her latest gadget on my screen and reliably informed me that the number of specks of dust is *n*. Then according to *Evidence Norm*, I have a warranting reason to believe that the number of specks of dust is *n*. But clearly, I have no goals that will be served by my so believing. Thus Cowie is wrong to claim that the reason to believe in accordance with evidence is that doing so furthers our goals.

In response, instrumentalists often make a distinction between genuinely normative epistemic reasons, and epistemic reasons like my reason to believe that the number of specks of dust on my computer screen is *n*, which are not genuinely normative (Leite 2007; Cowie 2014: p. 4014; Steglich-Peterson 2011). Now, whatever the relationship between instrumentalism and institutionalism, clearly the institutionalist needs to buy into this distinction. So arguments from critics of instrumentalism and the teleological view that target the tenability of this distinction appear at first glance equally to target institutionalism.

Paakkunainen (2018) presents arguments of this type, aimed at establishing that "[t]here's no good distinction between genuinely normative and not-genuinely-normative epistemic reasons: there are just plain old epistemic reasons – all of which are, in a perfectly good sense […], normative" (p. 123). In order to support this conclusion, she considers the various treatments of cases of trivial truths provided by instrumentalists, and finds their insistence that the epistemic reasons in these cases are not genuinely normative to be undermotivated. In the dog convention example, Paakkunainen agrees with Leite that the agent has no good reason to form the belief that the next person to come out of the door is likely to be accompanied by a dog. But she thinks that this is best explained not by saying that the epistemic reason possessed by the door-stander is not normative, but instead in terms of the "different, and quite familiar, distinction between epistemic reasons to believe that *p*, and reasons to seek knowledge or beliefs about subject

matter M" (p. 129). It's not that the door-stander's epistemic reason to form the belief that the next person to come out of the door will be accompanied by a dog is not normative – like all epistemic reasons, it is – but that the door-stander lacks a reason to seek knowledge on the subject of whether or not the next person to walk through the door will have a dog. "This doesn't support a distinction between genuinely normative and not-genuinely normative epistemic reasons," Paakkunainen says. "The presence of reasons to seek knowledge or beliefs about M isn't yet the presence of genuinely normative epistemic reasons, for it's not yet the presence of epistemic reasons at all. Hence the absence of reasons to seek knowledge […] isn't yet any evidence of the absence of something further, namely genuinely normative epistemic reasons" (ibid).

Now we might wish at this point to reply on behalf of the instrumentalist that the epistemic reason in the dog convention case cannot be normative, because normative reasons are those things in accordance of which it is irrational not to act. Since the door-stander is not being irrational if they fail to form the belief that the next person to come through the door will have a dog with them, the epistemic reason they have to believe this cannot be normative. In response, Paakkunainen argues that to get instrumentalism from Leite's example in this way, you would need some sort of principle connecting epistemic reasons and rationality. But there is no reason to suppose that the most plausible such connecting principle will underwrite instrumentalism. In particular, Leite would need a principle like:

> **N1**: A (possessed, undefeated) epistemic reason, *q*, to believe *p* is normative only if [one would be irrational, or would display an epistemic fault, if one failed to respond to *q* by coming to believe that *p*].
>
> (p. 131.)

But there is no reason, Paakkunainen believes, to prefer N1 to the following norm, which preserves the normativity of the epistemic reason in Leite's example:

> **N2**: A (possessed, undefeated) epistemic reason, *q*, to believe *p* is normative only if [one would be irrational, or would display an epistemic fault, if one failed to respond to *q* by coming to believe that *p when one considers the question whether p*, and considers *q*].

(ibid.)

So the only motivation for the distinction between epistemic reasons that are genuinely normative and those that are not is undermotivated, Paakkunainen thinks. However, consider now the following passage from early on in her paper:

> Let's call all non-instrumentalist, non-teleological accounts of epistemic reasons *intrinsicalist*. For instance, the following is a version of intrinsicalism: evidence that *p* constitutes an epistemic reason to believe that *p* because it's epistemically rational to apportion one's beliefs to one's evidence, and epistemic rationality just is responsiveness to epistemic reasons [...]. It's in principle open to intrinsicalists to say that on their view, epistemic reasons aren't yet guaranteed to be genuinely normative; and to hold that, to be genuinely normative, epistemic reasons must meet some further (non-instrumentalist, non-teleologist) condition. So the distinction I criticize isn't, as such, an exclusively instrumentalist or teleologist one.
>
> (p. 124.)

It seems that epistemic institutionalism as formulated by Cowie is just such an intrinsicalist position that nonetheless appeals to the same distinction between genuinely normative and not genuinely normative epistemic reasons that Paakkunainen is criticising.[33] For on epistemic institutionalism, epistemic reasons remain categorical: in the dog convention case, the agent has an epistemic reason to believe that the next person to come out of the door will have a dog with them, whether the agent likes it or not. There is no escaping this epistemic reason in the sense of preventing it from *applying*. It is just that the reason is not genuinely normative for the agent.

Thus the institutionalist can easily accept:

> **N2*:** A (possessed, undefeated) epistemic reason, *q*, to believe *p applies* iff [one would display an epistemic fault if one failed to respond to *q* by

[33] Hence the presence of genuine difficulty in ascertaining the relationship between instrumentalism and institutionalism. If institutionalism is on the intrinsicalist side of Paakkunainen's divide, then one might wonder whether the two views are even compatible, let alone whether institutionalism is a type of instrumentalism. And yet, as Cowie's argument for *Trivial Truths (Epistemic)* illustrates, institutionalism is best established by much the same arguments that instrumentalists use to support their view. In the text, I am defending institutionalism only, not necessarily instrumentalism. And institutionalism is on the intrinsicalist side of Paakkunainen's divide. To resolve the taxonomic confusion, the question of the applicability of Paakkunainen's divide to the entire literature would need to be considered. For my purposes, I do not need an answer to this question.

> coming to believe that *p when one considers the question whether p*, and considers *q*],

and it seems to me that they will easily be able to accept similarly institutionalised versions of *any* epistemic norm that Paakkunainen might care to defend.

So – continuing the theme of the presence of taxonomic obscurity in these waters – institutionalism begins to look like it might have more in common with Steglich-Petersen's teleological view than with instrumentalism. On this view, epistemic reasons *per se* are not normative at all, they just *inherit* normativity from other reasons. In responding to Steglich-Petersen, Paakkunainen notes that he appears to be appealing to a norm like:

> **N3**: A (possessed, undefeated) epistemic reason, *q*, to believe that *p* is normative only if [one ought to believe that *p*]. [34]
>
> (p. 134.)

Her complaint is that N3 is doubtable, and that Steglich-Petersen offers no rationale to support it. "It would be nice to have some argument for [N3], and an articulation of the sense of "ought" in play in [N3], which Steglich-Peterson doesn't provide," she says (p. 134). But of course, at this point in our discussion, we do have an argument for N3. The argument is simply that the non-Humean instrumentalist account of practical rationality has taught us that the truth or falsity of any claim of the form "A has a normative reason to φ" must be a function of A's subjective motivational set. But norms like N2* can apply regardless of the contents of A's subjective motivational set. Thus, it would appear that N2* combined with the theory of practical rationality accepted in Chapter 1 leads us to something like N3. This supports the institutionalist's use of the distinction between genuinely normative epistemic reasons and merely institutional epistemic reasons – or, more precisely, between institutional epistemic reasons that successfully inherit normativity from other reasons, and those that receive no such inheritance.

In sum, then, whilst the debate between instrumentalists and proponents of the teleological view on the one hand and those like Paakkunainen on the other who think

[34] This norm is numbered N4 in Paakkunainen's paper.

that all epistemic reasons are normative is unlikely to reach consensus any time soon, it does seem that institutionalism is a fairly reasonable view to bring to the task of assessing whether EPM is successful against UMET.

§2.2.2.4: Comparison: Epistemic Reasons and Moral Reasons

Let us henceforth assume that epistemic reasons are merely institutional, and this is established on the basis of considerations of ordinary explanation and the discussion in §2.2.2.3. Notice that in the discussion in §2.2.2.3, trivial truths have more or less dropped out of the picture. After all, a trivial truth is one that is insignificant and wholly irrelevant to our concerns. But these are not the only sort of truths there might be epistemic reasons to believe, but which it would be practically irrational to believe nonetheless (see §2.2.2.5). Henceforth, then, I will talk about *practically non-rational truths*, rather than trivial truths. A practically non-rational truth is one that an agent would not be practically irrational not to believe. There may be some truths that it would be decidedly practically *ir*rational for an agent to believe. These would be one kind of practically non-rational truths. Trivial truths are another kind.

How do moral reasons compare to epistemic reasons? First, with respect to ordinary explanation, the answer to the questions of why someone should act in accordance with their moral reasons is characterised by neither the positive nor negative feature of ordinary explanation in the case of institutional reasons. Recall that the negative feature was that, in explaining why you should act in accordance with (e.g.) your legal reasons, we would not give further legal reasons, but step outside the institution of the law. Moral reasons are not like this: faced with an agent unmoved by his moral reasons, who asks "but *why* shouldn't I murder my rival in order to secure this job promotion?", we would still provide moral reasons. It's true that we could leave morality and appeal to his prudential reasons ("how would you like to spend twenty years in prison?!"), but from the moral point of view something seems a little off about such a response. The moral reasons are supposed to apply even if the prospective murderer can be certain that he will not be caught and will face no negative consequences from carrying out his crime. And this shows that, likewise, moral reasons are unlike institutional reasons with respect to the positive feature of ordinary explanation. With institutional reasons, we do leave the institution behind,

and appeal to (moral or) prudential reasons. But as just noted, something about this in the moral case seems a little off.

How do moral reasons compare to epistemic reasons with respect to the discussion of §2.2.2.3? Here is Cowie's example:

> Consider, for example, Tyson and Deontay. They are on good terms. One day, however, the opportunity arises for Tyson to enslave Deontay. And, as it happens, Tyson has no non-moral reasons not to do this. Tyson doesn't, for example, have any prudential reasons to refrain from enslaving Deontay. Presumably, though, it is still the case that Tyson *morally* ought not to enslave Deontay [...] and so he has a moral reason not to enslave him. What would we say now? Does Tyson have a genuinely normative reason not to enslave Deontay? [...] The answer is pretty obvious. Of course Tyson's reason is a genuinely normative reason. This is *totally unaffected* by Tyson's current lack of non-moral reason to do otherwise. To think otherwise would be non-standard, to put it mildly.
>
> (Cowie 2019: p. 69.)

Once again, the disanalogy between institutional reasons and moral reasons is obvious: clearly, Tyson still ought not enslave Deontay. The normativity survives Tyson's indifference. (Or, at least, that is what a believer in morality is committed to, according to AfO.) This is quite unlike the case of practically non-rational truths. In those cases, whilst the epistemic reason still *applies*, it is not normative. For moral reasons to be on a par with epistemic reasons, it would have to possible for Tyson to have a moral reason not to enslave Deontay, but for this reason not to be normative because Tyson has no interests that would be served by acting in accordance with his moral reasons. According to AfO, though, this is precisely what the moralist does *not* think.

§2.2.2.5: A Final Consideration from Rowland

The claim against the argument from analogy is that moral reasons to perform morally required actions that the agent would not be practically irrational not to perform are unlike epistemic reasons to form beliefs that the agent would not be practically irrational not to perform, in the sense that the former but not the latter are usually thought to be genuinely normative.

Rowland appears to reject this claim when he considers the case of a character he calls Ella:

> Suppose that an agent, Ella, desires only psychological contentment, and only ever desires this, and that Ella is extraordinarily psychologically fragile. Because of her fragility, in order for Ella to be psychologically content she would have to isolate herself from almost all other agents and only engage in extremely simple tasks. Ella must block off from consideration a vast number of propositions regarding her own psychological state, the state of the world, and the status of her friends and family among other things because she finds considering these matters extremely disturbing. Suppose that Ella is extremely successful at blocking out all these considerations over the course of her life. In this case, if [instrumentalism] is right, there is no reason for Ella to believe many propositions about the world, herself, and her friends and family. For instance, even if she were well aware of R, and R is extremely good evidence that her father has cancer, there is no reason *at all* for Ella to believe that her father has cancer. But this is exactly what we do not think.
>
> (Rowland 2013: p. 5, emphasis in original.)

But the phrase "there is no reason at all for Ella to believe that her father has cancer" is ambiguous. If it means that there is no *epistemic* reason for Ella to believe that her father has cancer, then Rowland is surely right to say that "this is exactly what we do not think". But of course, epistemic institutionalism does not commit us to thinking this. The institutionalist thinks that there is a reason for Ella to think her father has cancer, but this reason is not normative. If on the other hand the claim that "there is no reason at all for Ella to believe that her father has cancer" is intended to mean that Ella has no *normative* reason to believe this, then this claim is actually true, and should not be "exactly what we don't think". For the truth that her father has cancer is, for Ella, a practically non-rational truth.

Summary and Conclusion

§2.3

We have considered the case against ***Parity*** and found it convincing. Consequently, the score we originally came to when marking the Moral Moorean Argument against McPherson's epistemic indicators stands. Thus, we can conclude that when compared to the original anti-SEW version, Moral Moorean Arguments are much less effective.

As I said in §2.0, the discussion in this chapter has been in the service of two aims. The first was to bolster the case for ***Plausibility***. Since I take the preceding discussion to show

at least that moral Moorean Arguments do not provide an easy way for the UMET-critic to simply dismiss UMET out of hand, I take it that ***Plausibility*** has been thus further supported by the discussion in this chapter. In line with ***Consequence***, this gives us further reason to inquire as to the consequences UMET might have for moral thought and talk as we currently know it. From this point on, this thesis will concentrate on this task. The eventual result will be that we arrive at the objection to UMET that I have already advertised – the objection from loss. Having covered the ground that has been covered in this chapter should make it easy, when we finally arrive at that objection, to see how this objection is different from a moral Moorean argument, despite their surface similarities.

Chapter 3:
Abolitionism

§3.0.0: Abolitionism and the Objection from Loss

So far in this thesis most of the discussion has been in the service of defending:

> **Plausibility**: There are arguments that together present a strong *initial* case for UMET. These arguments have sufficient persuasive force to merit our taking UMET seriously.

I have defended *Plausibility* in two ways. First, I have discussed AfO, which is itself simply a combination of Joyce's AfR and the argument from disagreement that Leiter attributes to Nietzsche. Second, I have considered and rejected the claim that moral Moorean arguments undermine *Plausibility*.

In chapter 1, I also offered a brief defence of:

> **Consequence:** If *Plausibility* is true, then it is reasonable to inquire as to the consequences that UMET might have for ethical thought and talk as it currently exists in society.

My main defence of *Consequence* was that UMET-proponents need to inquire as to the consequences that UMET might have for moral thought and talk because if they fail to do so, and so fail to formulate a convincing U-NWP proposal (or a collection of such proposals), they leave themselves vulnerable to the objection from loss. I said that I think that this objection is the most formidable that the UMET-proponent faces, and I shall be pressing this objection in Chapter 5.

In this chapter, I begin doing as *Consequence* says it is reasonable to do, given the case for *Plausibility*. I begin inquiring into the consequences that UMET might have for ethical thought and talk as it currently exists in society. This brings me straight into the territory of the U-NWP literature. In this chapter and the next, I will focus on two of the options represented in that literature. In this chapter, I will discuss abolitionism. In the next, I will discuss substitutionalism. In the course of the discussion of both chapters, a picture will

begin to emerge of ethical discourse as practically indispensable; this practical indispensability will be at the heart of my formulation of the argument from loss.

Abolitionism and Self-Defeat

§3.1.0: More Terminological Precision — Discourse Implicating and non-Discourse Implicating Ethical Judgements

Recall that in this thesis, the terms "ethics" and "morality" have precise and different meanings: ethical evaluations of actions (and characters) are evaluations related in the right way to the serving or frustrating of the ends of others; moral considerations are those governed by concepts (like NICRs) of which an error theory is to be accepted (at least for the sake of argument).

As I have already said, my highly plausible contention that the actions of the Croydon cat killer were cruel looks like a perfect candidate for a non-moral, ethical consideration. Certainly I am assessing the actions and character of this unknown villain in terms of the effects they have had on others. I use the word "cruel" because I am thinking of the pleasure the villain has taken in the suffering and death of the cats, and it is obvious that my assessment here depends on an appreciation of the effects of these actions on others – namely, the cats themselves and the humans who cared for them. But it is not obvious that I am helping myself to any suspect moral concepts in evaluating the cat-killer's actions as cruel: I am not assuming that the killer had some NICR not to kill the cats, or that we can expect convergence upon fundamental philosophical truths about morality that capture or explain precisely how and why the killing of the cats was wrong. So my contention is ethical, but not moral.

For the purposes of this chapter, I need a distinction within the realm of the non-moralised ethical. This is the distinction between what I shall call *discourse implicating* and *non-discourse implicating* ethical judgements. This distinction is most easily understood in light of the condition formulated in §1.1.4 for identifying the point at which any given instance of MET constitutes UMET:

> **UMET Condition:** It is necessary and sufficient for a MET to amount to UMET that the ethical web constructible from the uninfected nodes belonging to the old error-infected web is insufficient to serve *as the basis on which to continue a social practice of ethical discourse*.

Here we need to concentrate on the part of *UMET Condition* that is italicised above. Whether or not a given ethical utterance is discourse-implicating depends on whether, in making the utterance, the speaker commits themselves to the feasibility of ethical *discourse* – i.e. a shared practice of exchanging ethical utterances in such a way that the defensible ethical utterances can be, however contestably, sorted from the indefensible.

This is admittedly vague, so allow me to elucidate. I shall explain *one* clearly identified way in which an ethical judgement can amount to a discourse implicating judgement. This is the not the only way for a judgement to be discourse implicating. Consider the claim of Finlay's that we encountered in §1.4.1:

> *Relational Ethical Application*: the term "ethically wrong" is applied to actions that frustrate certain ends or violate certain standards.

(Here I have substituted the words "ethical" and "ethically" for the words "moral" and "morally" as it appeared in *Relational Moral Application* in §1.4.1. This is for the sake of compliance with my own stipulations on the use of the terms "moral" and "ethical". Note that merely pointing out that a certain action violates some arbitrary end is not in itself discourse implicating.)

Now, a central theme in Finlay's analysis of moral language is the need for shared ends to ground what I am calling the social practice of ethical discourse. Suppose that Alice, an ethical vegan, tells Connor, a meat-eater, that he should not eat meat. Alice applies the term "ethically wrong" to eating meat because it frustrates certain ends relating to the welfare of animals. Now Connor defends himself. He says that eating meat can be ethically acceptable, subject to certain provisos all aimed at reducing the suffering caused to the animals whose flesh we eat. Here, there appears to be enough overlap in the ends they prioritise for them to appear to be exchanging ethical utterances with the aim of sorting the defensible from the indefensible. It's not as though Connor just doesn't care about animal welfare. When Connor says that eating meat can be ethically acceptable, he makes a discourse-implicating ethical judgement.

But if UMET is true, then Connor's judgement, precisely because it is discourse implicating, must be error-infected. Here is an obvious explanation the UMET-proponent could offer for the source of error-infection, given that their commitment to UMET is based on AfO: whilst it is true that Connor and Alice's ends overlap to a certain extent – both, for instance, wish to see an end to the worst abuses of animals commonly seen in today's practices of factory farming – they still diverge sufficiently that in exchanging discourse implicating utterances they are talking at cross-purposes. Alice is not prepared to tolerate any breeding and killing of animals for food; Connor is. This might be due to a genuine difference in normative reasons. Maybe Alice's fully rational counterpart would want Alice to do everything possible to bring about a complete cessation of human's use of animals for food, whilst Connor's fully rational counterpart would advise Connor to do everything possible to retain the institution of animal farming whilst making it less abusive of animals. In that case, their normative reasons diverge and the social practice of ethical discourse that they join – i.e., the practice of exchanging ethical utterances in such a way that the defensible ethical utterances can be sorted from the indefensible – is a confused enterprise.

So discourse implicating ethical judgements are those like Alice's and Connor's which commit agents to the feasibility of the social practice of ethical discourse. Non-discourse implicating ethical judgements are those that do not commit the agent to the feasibility of ethical discourse. But what might such an ethical judgment look like? We may wonder if there could really be any. After all, the main example of a non-moralised ethical judgement I have offered so far – my judgement that the actions of the Croydon Cat Killer were cruel – looks discourse implicating. And what would be the point of an utterance that expressed an ethical judgement that could not contribute to the social practice of ethical discourse?

One form of a non-discourse implicating ethical judgement would be grounded in the prioritisation of extremely idiosyncratic ends that the agent expects nobody else to share. A more interesting form could be generated if we start with Connor's discourse implicating ethical judgement and ask ourselves how he could modify it in order to make

it a non-discourse implicating ethical judgement. Once again let's think about it from the perspective of *Relational Ethical Application*.

One option for Connor would be to embrace a fairly radical individualistic relativism. If he was trying to purge himself of discourse implicating judgements, he could choose to jettison the judgment "eating meat is ethically acceptable subject to certain provisos" in the sense that eating meat does not violate certain ends that are to be assumed for important to both speaker and audience. He could replace this with the following story: *he* has ends *e*, these are ends with which *he* is comfortable when he thinks hard about the institution of animal farming, and animal farming could be reformed so as not to violate ends *e*, even though animal farming in any form whatsoever does violate ends e_0 held by Alice, which are the ends *she* is comfortable with when she thinks about the institution of animal farming. And if another agent, Ned, had ends that included keeping the cost of meat as low as possible and found that dedicating time and effort to resisting political pressure to implement new animal welfare legislation promoted that end, then Connor would have to say that his own non-discourse implicating ethical judgement that new legislation is required *does not apply from Ned's perspective*. He couldn't say that Ned was making an ethical *mistake* in opposing the proposed new legislation; he could only say that Ned's ends were different to his. Of course, Connor's ethical utterances would now be entirely dialectically toothless, but that is the price an agent pays for radical individualistic relativism.

I hope that what I have said so far gives enough of a sense of the distinction between discourse implicating and non-discourse implicating ethical judgements. Of course, I have not provided any necessary conditions on a judgement's being discourse implicating. In fact, I do not believe it would be possible to provide necessary conditions, for reasons that should become clear in Chapter 4. (Preview: the extent to which an ethical judgement is discourse implicating is a continuous variable.)

I can, however, offer a highly defensible sufficiency claim. Many philosophers, e.g. Railton, equate the *moral* point of view – where their use of the word "moral" does not have the connotations of error-infection that that word carries in this thesis – with the *socially rational* point of view:

> Moral evaluation seems to be concerned most centrally with the assessment of conduct or character where the interests of more than one individual are at stake. Further, moral evaluation assesses actions or outcomes in a peculiar way: the interest of the strongest or most prestigious party do not always prevail, purely prudential reasons may be subordinated, and so on. More generally, moral resolutions are thought to be determined by criteria of choice that are *non-indexical* and in some sense *comprehensive*.
>
> (Railton 1986: p. 189.)

Of course, what Railton calls "moral evaluation" should, in the context of this thesis, instead be called "ethical evaluation". Railton is not suggesting that there is anything suspicious about the type of evaluation described in this passage. Moreover, Railton here seems to be talking about the type of ethical evaluation that involves us in the social practice of ethical discourse. And he identifies this practice as the taking up of the socially rational point of view. By the point in 'Moral Realism' at which the quoted passage occurs, Railton has already offered an analysis of *non-moral value* which essentially reduces to the notion of *objectified subjective interest*, with this latter notion highly reminiscent of the discussion of practical rationality in Chapter 1:

> Give to an actual individual agent *A* unqualified and imaginative powers, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history and so on. *A* will have become *A+*, who has complete and vivid knowledge of himself and his environment, and whose instrumental rationality is in no way defective. We now ask *A+* to tell us not what *he* currently wants, but what he would want his non-idealised self A to want – or, more generally to seek – were he to find himself in the actual condition and circumstances of A.
>
> (Railton 1986: p. 173-4.)

In this passage then, what *non-moral value* is for A, is whatever is in A's objectified subjective interest. An agent identifies non-moral value from the point of view of individual practical rationality. What "moral" value is (read: what value is from the perspective of *the social practice of ethical discourse*) is whatever is in the amalgamated objectified subject interest, i.e. the objectified subjective interest of society as a whole.

Now, it would appear that there are various points of entry into the social practice of ethical discourse. In identifying "moral" value – again, recall that Railton is not using this word in the way that I use this word – with that which satisfies the aggregated objectified

subjective interest, Railton is joining a consequentialist tradition with a rich intellectual history going back to Bentham. Someone like Korsgaard might join the social practice of ethical discourse from a quite different perspective, influenced more by Kant, thinking of the demands imposed upon creatures like us by the nature of our rational agency, and not assigning central importance to considerations of aggregated objectified subjective interest. In my view, this does not mean that she joins a different social practice. What Korsgaard contributes to is still recognisable as *the* social practice of ethical discourse. She just has different ideas to Railton about what considerations should be brought to bear on the questions that that practice exists to answer. So it is certainly not necessary to give such pride of place to the notion of social rationality in order to join the social practice of ethical discourse. However, my claim is that it is sufficient:

> **Social Rationality Sufficiency**: An agent who makes ethical judgements by taking up the perspective of social rationality joins the social practice of ethical discourse. The ethical judgements at which the arrive in this manner are thus discourse-implicating ethical judgements.

§3.1.1: Why Abolitionism is Self-Defeating

With the distinction discourse implicating and non-discourse implicating ethical judgements in hand, let us now turn to abolitionism as a U-NWP policy option. It would be usual at this point to begin by explaining in a succinct manner what it is that abolitionists argue. But it is surprisingly difficult to do this.

Olson (2014: p. 179) distinguishes between *partial* and *complete* abolitionists. Partial abolitionists target just some subset of moral discourse. Thus, Olson counts Anscombe's (1958) famous discussion of the place of deontic concepts in our secular age as a defence of abolitionism – of partial abolitionism, that is. Bentham must also surely count as an abolitionist of this sort, because he wants to eliminate talk of natural rights. Pereboom (2001: chapter 5), who wants to dispense with the concept of moral blame, must also count as a partial abolitionist. But these sort of abolitionist proposals seem to be based on LMETs and thus have little to do with UMET. Our focus must be on complete abolitionism.

What would complete abolitionism look like? Arguing (possibly) for complete abolitionism, Garner (2007: p. 500) introduces the policy by telling us that "the advice of the moral abolitionist is to give up moral language, and with it the arrogance and interference that a belief in the objectivity of morality often occasions". But interference with what? It seems that what universal abolitionists have in mind here is interference in the form of disruption to humans pursuing what they desire or what is in their interests. Hinkfuss, for example, argues that in the "moral society" (i.e. a society in which belief in morality is widespread, and people are concerned to acquire true moral beliefs, avoid false ones, and act in accordance with the moral beliefs they have), various negative effects can be expected, and indeed are observed. These negative effects include the psychological sufferings of guilt and moral denigration, widespread economic inequality, and an increased danger of wars and violent revolutions. These concerns are clearly about morality's cost to human's ability to pursue what is in their interests. Indeed, at various points, he explicitly frames the discussion in this way. For example:

> The question here is not whether the moral institution has on some occasions a useful effect. *It is whether it is worth preserving, given the sum of its effects on and within society*.
>
> (Hinkfuss 2019: p. 35, emphasis added)

It appears, then, that Hinkfuss wishes to proceed by arguing that moral discourse as we know it does a poor job of getting us what we want, or satisfying our needs, or promoting our interests. Now if the arguments by which the abolitionist has arrived at UMET do not touch practical rationality – as will be the case here since Hinkfuss' abolitionist views are motivated by similar worries to those discussed in §1.1 concerning inescapable authority – then this might initially appear to be exactly as it should be. For we have a well-defined target (viz., moral discourse as we know it) and a well-defined perspective from which to assess its worth (viz., the standpoint of practical rationality).

But wait. As a UMET-proponent, Hinkfuss cannot legitimately make discourse implicating ethical judgements. He needs to restrict himself to non-discourse implicating judgments. Does he succeed in doing this? I think not: the italicised part of the above quotation constitutes a clear instance of adopting the socially rational point of view. If we

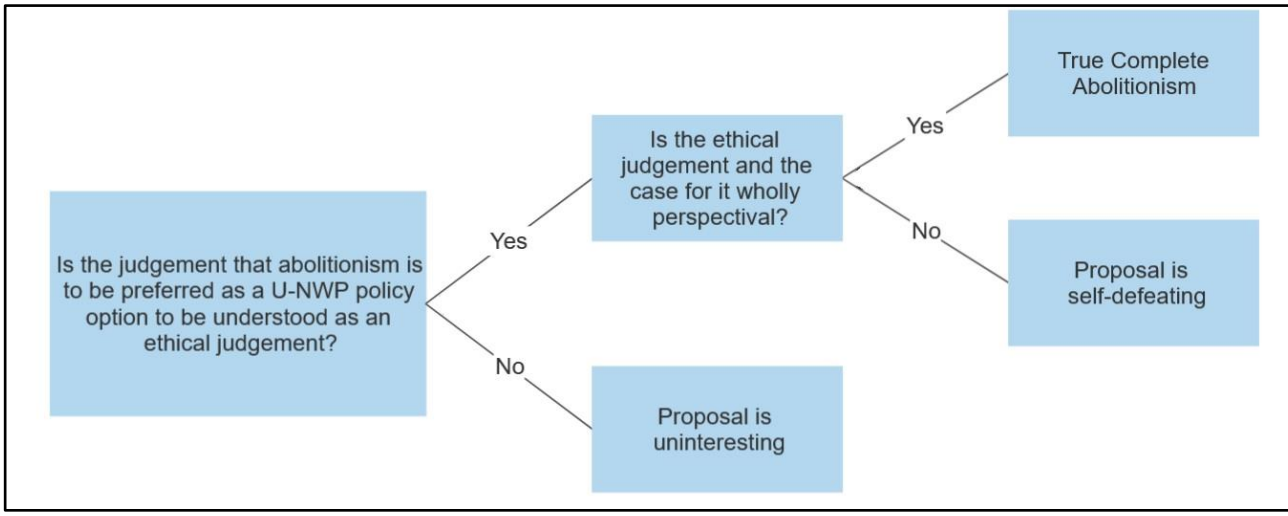
accept *Social Rationality Sufficiency*, then Hinkfuss violates *UMET Condition* to the extent that he thinks that we can take the socially rational perspective without involving ourselves in systematic error.

Here is another way of making the same point. A major part of the case for UMET considered in Chapter 1 was that we should not expect convergence with respect to normative reasons for all agents. As I understand him, Hinkfuss is committed broadly to accepting this view of practical rationality. So, if it is to be argued that we should abolish morality because is costly, that it doesn't serve our interests, then the natural question is: *whose* interests? Who are "we"? Isserow (2019: p133) suggests understanding the abolitionist thus: "by 'we' abolitionists mean to refer to *most* of us, who presumably want to avoid an untimely demise, see to it that others are well, and live in a stable and co-operative society". This is very obviously a sensible way to interpret Hinkfuss: it makes sense of his claims that morality leads to a hierarchical, authoritarian, elitist society in which large-scale violent conflict is possible and people needlessly suffer the ill-effects of moral denigration and guilt. But now we might think that in choosing to address themselves to those whose desires include *that others are well*, the argument to be given is ultimately to be made from the perspective of social rationality. But by *Social Rationality Sufficiency*, the ethical judgement that abolitionism is to be preferred as a U-NWP policy option is itself a discourse implicating ethical judgement, which on UMET is illegitimate. The spectre of self-defeat has been raised.

So the complete abolitionist treads a fine line, which we could illustrate with a decision tree. The first question the complete abolitionist faces is: is your judgement that complete abolitionism is to be preferred as a U-NWP policy option itself an ethical judgement? Certainly, the abolitionist can immediately avoid self-defeat by answering this question in the negative. They might, for example, present their abolitionist proposal as a purely egoistical policy: it pays not to be duped by the error-infected discourses of a moral society. (It might even form part of such a proposal that one should do all one can to keep the moral society going – it might be beneficial to oneself if others continue to make all the mistakes associated with participation in the error-infected practice.) However, for such unimpeachability in terms of coherence the abolitionist would surely pay a price in

terms of relevance. For such an abolitionist simply does not address "most of us, who presumably want to […] see to it that others are well". So, let us say that the abolitionist answers in the affirmative: the judgement that abolitionism is to be preferred is to be understood as an ethical judgement.

There is of course no *immediate* problem with abolitionists presenting their judgement that abolitionism is to be preferred as an ethical judgement. It is perfectly legitimate for them to do so. They simply have to be careful to restrict themselves to non-discourse implicating ethical judgements. This includes when they turn their attention to defending their U-NWP policy option. The advertisement of the advantages of complete abolitionism itself needs to be non-discourse implicating. So, having answered the first question on our decision tree in the affirmative, the abolitionist now faces the question: can you honestly say that you have succeeded in avoiding any discourse-implicating ethical judgements? If the answer here is "no", then the abolitionist's proposal is self-defeating. And we have seen that Hinkfuss' answer must be "no": he clearly defends his brand of abolitionism from the socially rational point of view, which by *Social Rationality Sufficiency* makes his ethical case for abolitionism discourse-implicating rather than wholly perspectival.



So, what would a complete abolitionist proposal that avoided self-defeat (i.e., "True Complete Abolitionism" in the decision tree above) look like? So far in this section, I have discussed Hinkfuss. But earlier I also mentioned Garner's recommendation that we "give up moral language", with all of its "arrogance" and "interference". Neither the

judgement "morality is objectionably arrogant" nor the judgement that "morality generates undue interference" seem to be inherently discourse implicating, so does this proposal belong in the box labelled "True Complete Abolitionism" in the decision tree? It would appear not. Garner's proposal and defence of that proposal is essentially the same as Hinkfuss'. It is true that Garner is more cautious than Hinkfuss: instead of the great polemicising against morality that characterises Hinkfuss' writing, Garner emphasises that "we may never find an indisputable answer to the question of whether the moral overlay helps us more than it harms us" (Garner 2007: p. 511) and suggests, fairly modestly, that we "just cut back on our use of moral language and see how things go" (ibid). But the fact remains that this is the crucial question for Garner – whether "the moral overlay helps us more than it harms us" – and this is a question that seems comprehensible only from the perspective of social rationality. Furthermore, much of Garner's case for abolitionism is the same as Hinkfuss' – indeed, one section of Garner's paper consists of an approving discussion of Hinkfuss' arguments. So it would appear that Garner is to be placed in the same category as Hinkfuss: his proposal is self-defeating, at least when he turns his attention to defending it.

It seems to me that this pattern is going to be entirely general. I do not think that any *interesting* abolitionist proposal stands much chance of ending up in the box labelled "true complete abolitionism". The trouble is that in order to be *interesting*, the proposal is going to have to appeal to "people like us […] who wish to see to it that others are well", as Isserow puts it. But it is difficult to see how a proposal could be interesting in this way if the case for it was not made from the perspective of social rationality. But by ***Social Rationality Sufficiency***, this means that any such proposal will be self-defeating.

Thus, the conclusion we reach is as follows. *Universal abolitionists involve themselves in self-defeat when they attempt to advertise the supposed advantages of their policy proposal by adducing considerations made from the point of view of social rationality*. That is the conclusion of this chapter, and we have reached it rapidly. However, my claim is that the brief argument I have given in the present part of the chapter goes a long way to establishing this conclusion. Certainly all of the defences of universal abolitionism as

a solution to the U-NWP that I have come across have been made from the viewpoint of social rationality.

§3.1.2

Although my simple argument is now complete, some illustration of it is required to make it fully compelling. This I shall do by looking at two abolitionist proposals – one which is not presented as a U-NWP policy option, and so escapes the argument made above, and one of which is so presented, and clearly falls victim it.

Proposal 1: Blackford

§3.2.0: Blackford — Not a UMET-proponent

Although the editors of a recent volume on moral abolitionism (Garner & Joyce 2019) saw fit to put Blackford's contribution in the part entitled 'The Case for Abolitionism', Blackford does not appear to endorse complete abolitionism. My aim in this part of the chapter is to defend this claim. Thus, the conclusion to be established in this part of the chapter is simply:

> *Blackford Non-UMET*: Blackford does not endorse UMET, does not offer a U-NWP policy proposal, and the modifications he wants to see made to social discourse do not amount to complete abolitionism.

Note that defending this claim requires me to engage in taxonomy, not exegesis. There are no obscure passages in Blackford in which it is not clear what his view amounts to. What I am aiming to resolve is what I see as a general confusion in the U-NWP literature, concerning what the problem amounts to. The problem, as my defence of *UMET Condition* in §1.1.4.2 and the discussion in §3.1.0 suggests, concerns what to do if the errors in morality as we know are sufficiently pervasive that uninfected ethical nodes are insufficient to ground a social practice recognisable as ethical discourse. The problem does not concern which opinions *within* the social practice recognisable as ethical discourse are the most rationally defensible. But Blackford's discussion, as we shall see, concerns the latter of these two topics, not the former. In fact, he makes a pretty clear contribution to the social practice recognisable as ethical discourse.

Blackford shares with Joyce a belief that morality's appearance of inescapable authority is misleading, and thinks it would be desirable if widespread belief in NICRs were to diminish. (Blackford does not talk about NICRs precisely; I am here recasting Blackford's discussion in Joyce's terms.) But he doesn't think that the abolition of thought and talk presupposing NICRs will amount to the abolition of everything recognisable as ethical discourse: "my somewhat tentative conclusion," Blackford writes, "is that disbelief in objective moral authority will not comprehensively change the everyday language that we use in our thoughts and conversations" (2019: p. 68). Blackford does not propose that we cease making "ought"-statements, or evaluating things as "good" or "bad", or using thick ethical terms.

§3.2.1: Blackford on The "Horrible Mismatch"

As already remarked in the previous section, Blackford's issue is with inescapable moral authority. His basic reason for thinking that this authority is not real is that morality is (merely) "social technology" (p. 61).[35] Most people, he thinks, nonetheless believe that moral norms are inescapably authoritative. There thus exists what Blackford calls a "horrible mismatch" (p. 60) between the reality of morality as an evolved social technology and the appearance (to many of those utilising this technology) of morality's being something with special authority.

The abolitionism that Blackford supports is abolition of this idea of morality as something more authoritative than it really is. At the societal level, though he seems to think that abandoning belief in inescapable authority would be desirable, he notes that it is ultimately unrealistic:

> [I]t appears to me unrealistic to expect a majority of people – at any time in the foreseeable future – to accept that their various moral systems are human inventions. Therefore, if we're tempted to think that widespread acceptance of this difficult truth will solve any of our social and global problems, it's a temptation that I recommend we set aside.

[35] This is a view Blackford attributes to Mackie. He further states that he does not believe that the thesis that "objective moral authority is part of what our ordinary (affirmative first-order) moral judgements *mean*" is correct (p. 60, emphasis in original), and that it would be disappointing if discussion of Mackie's work focussed exclusively or mainly on "such a doubtful thesis". Blackford thus appears to endorse LMET, and to attribute this view to Mackie. For further doubts that Mackie was a UMET-proponent, see Berker (2019).

(2019: p. 71.)

Hence his proposal is pitched more at the level of a subset of the population who can use this insight to assist their thinking in applied ethical contexts:

> Once we view morality as a form of social technology, the new perspective can liberate our thinking about many cultural and political issues. If morality's original function was to promote intragroup cooperation, perhaps we can reshape it to assist without our modern predicament, where particular societies consist of many tribal groups and no society can flourish in ruthless competition with others. [...]
>
> If we care about ends that morality has not served ... we can attempt to repurpose morality to some extent. If, moreover, some moral norms are no longer socially functional, or perhaps never were, we have a reasonable basis to discard them. [...] It is possible, indeed, that much in our traditional moral norms does more harm than good. If so, perhaps we cannot overturn all our local moral systems at once [...] but it might be justifiable to ignore, or publicly defy, the most oppressive requirements.
>
> (p. 69-70.)

The spirit and substance of Blackford's version of abolitionism, then, can be aptly summarised thus: *if we recognise ethics for what it is, i.e. mere social technology, instead of allowing ourselves to be duped by the myth of what it is, i.e. a collection of inescapably authoritative norms, then we are likely to be able to make better use of the technology*.

§3.2.2: Blackford's Affinities With Pragmatic Naturalism

Blackford and Mackie are not alone in emphasising morality's status as something created by humans. The idea is at the heart of a (non-error theoretical) metaethical view developed by Kitcher (2011) – pragmatic naturalism.

§3.2.2.1: Pragmatic Naturalism

Pragmatic naturalism offers an interesting account of ethical truth which can be used to buttress Blackford's partial abolitionism, explain the mismatch between morality's role and its appearance as authoritative, and motivate a version of LMET.[36] Kitcher bases his view on an "analytic history" of the ethical project which describes how the human species got from a condition in which psychological altruism first evolved and led to the

[36] In keeping with my current policy of not assessing Blackford's proposal, I will also offer no assessment on the relative merits of pragmatic naturalism. As it happens, though, this is the metaethical view to which I now subscribe, having abandoned UMET on the basis of the arguments presented in this thesis.

establishment of a "coalition game", to the condition of immersion in the ethical project as we know it today. His aim here is to emphasise the natural evolution of the ethical project; in many places he does not require his history to be accurate, only possible. The idea is to explain how we *could* have come to be in the position of immersion in the ethical project, without appealing to anything not strictly consistent with naturalism. Most of the details of his proffered natural history need not detain us; the main thing to be emphasised is that for Kitcher, the ethical project was founded, in crude form, in order to solve a particular problem affecting early human societies: conflict arising from failures of altruism. Hence, Kitcher speculates, the earliest incarnations of the ethical project took the form of publicly communicated normative guidance, embodying rules to be followed in the group, backed up by a system of punishment. The ethical project then evolves as modifications are made to the code to enable it to better serve the functions for which it was introduced.

How can this offer an explanation of the mismatch that Blackford discusses? Part of Kitcher's analytic history tells the story (which Kitcher proposes as *a possible explanation* of the evolution of the ethical project, not necessarily as accurate history) of the introduction of the device of the "unseen enforcer".[37] The early rules would have been vulnerable to being broken in circumstances in which the offending behaviour would have gone unwitnessed by other members of the group and punishment would therefore have been unlikely. So the group introduce the idea that there is some other omnipresent agency that observes each member's conduct, even when no other members of the group are around; this agency is supposed to have the power to inflict punishments of its own. The idea is that belief in this unseen enforcer increases compliance with the code, thus allowing the ethical project to discharge its functions more effectively: there will be less rule-breaking, and therefore less conflict generated by failures of altruism.

[37] In his discussion of the notion of the unseen enforcer, Kitcher emphasises that there are drawbacks to its introduction as well as advances. One of the drawbacks is that is paves the way for some people in society to claim special knowledge of the nature or wishes of the unseen enforcer, leading to a sort of tyranny whereby those claiming this special knowledge have undue influence over the contents of the social code to be adopted. This recalls Hinkfuss' concerns about elitism discussed earlier, and thus offers another illustration of how pragmatic naturalism could be used to accommodate the insights of abolitionists without having to buy into complete abolitionism and thus face the objection of self-defeat.

Thus, an explanation for the appearance of the inescapable authority of morality: it is an idea with a human history that could in principle be traced back to the early introduction into the ethical code of a cruder concept like that of the unseen enforcer. We have the impression of morality's inescapable authority because our moral ideas have been handed down to us over many generations of human society in which those ideas have continually evolved in ways that enable the ethical project to discharge its functions more effectively.[38]

Central to pragmatic naturalism is a conception of truth that depends on the idea of progressive change. The challenge is to explain how a change can count as truly progressive or regressive without first helping ourselves to a suspect notion of pre-existing ethical truth. Kitcher solves this problem by putting progress first, and by modelling ethical progress on technological progress rather than on progress in the accumulation of truth. At a very simple level the idea is this: the ethical project was started for a particular reason, viz. that early humans required a method of reducing social conflict arising from altruism failures. That is the fundamental function of the ethical project. Up to a certain point, changes to ethical codes can be classified as progressive or regressive according to whether they improve or hinder the society's ability to discharge that function. Later on in the ethical project, when new functions emerge – for example, when the division of labour in society has led to different members of the society having different roles, with some roles being more desirable than others, such that members of the society now have new desires for the roles thereby created, and thus a new and more enriched conception of the good – there might be balancing between functions to be considered. A change might be regressive when measured against some functions of the ethical project but progressive when measured against some others. This introduces into pragmatic naturalism an element of pluralism.[39]

[38] Of the many oversimplifications I make in this incredibly brisk exposition of Kitcher, one deserves mention here and will be important shortly: the process of refinement and evolution of the ethical concept will not be linear and some changes will in fact be regressive, inhibiting rather than improving the ability of the society who adopts them to discharge the functions of the ethical project.

[39] Kitcher thinks he can avoid pragmatic naturalism's entailing what he calls *rampant* pluralism, which would make his notion of ethical truth vulnerable to arguments from disagreement. He seems to think that if he cannot suppress the risk of rampant pluralism, his notion of ethical truth is sunk. However, if his arguments aimed at neutralising the threat – arguments having to do with minimum conditions on ethical discourse that

Thus, on Kitcher's view, moral progress is not achieved as a result of moral falsity's being replaced by moral truth that exists prior to the progress, waiting to be discovered. It is the other way around: moral falsity is replaced by moral truth when moral progress is achieved, and moral progress is made when alterations to the ethical code are made such that it can discharge its functions more efficiently, more completely or more reliably. To take an example: it is not that society made a progressive step forward in abolishing slavery because it came to apprehend a truth (that slavery is wrong) to which it had hitherto been blind; rather the truth that slavery is wrong obtains, and obtained even when slavery was practiced, because at the time that slavery was practiced, making a change to the ethical code that abolished slavery would have been progressive. It would have helped to remedy a huge class of systematic altruism failures, to wit, those of the slave owners who failed to respond to the wishes and interests of the slaves.

§3.2.2.2: Blackford's Partial Abolitionism as Progressive Change, and Pragmatic Naturalism as LMET

Whilst brisk, I hope that the exposition of Kitcher in the previous subsection makes it clear how Blackford can be seen as offering a version of LMET. The option would be open to Blackford to adopt Kitcher's view of ethical truth, and then frame his abolitionist proposal thus: abandoning the notion of inescapable moral authority would be a progressive change in our ethical code. This is because as things stand, the notion of inescapable authority prevents the ethical project from discharging its proper functions. This would then generate a form of LMET focussed on the idea of inescapable authority: the idea would be that the notion of authority owes its existence in our ethical framework to changes to the ethical code occurring many generations ago. These changes might have been progressive at the time, in the sense that they enabled the societies that adopted them to make headway with the ethical project, reducing violations of the code. However, Blackford could continue, the notion of inescapable authority is now holding us back; it

---

have to be met for the ethical project to continue to fulfil its original function – are found not to work, it is not clear to me that that would be the end of pragmatic naturalism's appeal. Instead pragmatic naturalism could form the basis of an LMET targeting reasonable convergence. This would be in addition to underwriting an LMET targeting inescapable authority. So pragmatic naturalism has the potential to lead to more than one LMET.

is preventing the ethical project from developing in ways that will enable it to discharge its functions (especially recently evolved functions) more effectively. The idea of inescapable authority would be "in important respects like our evolved sweet-tooth: useful in ancestral environments but potentially harmful in our much different modern context" (Fraser 2017: p. 158).[40] If the idea of ethical truth is connected, as it is in Kitcher's view, to changes that would be retained through an indefinite series of progressive changes, then the notion inescapable authority is in this sense a moral falsehood. Thus we arrive at a version of LMET, targeting the notion of inescapable authority. Because it would be a progressive step to excise all judgements governed by the concept of NICRs from the ethical web, these judgements are *moral* judgements, in the special pejorative sense that word has in the context of this thesis.

§3.2.3: Classifying Blackford's Proposal

Where, then, is Blackford to be placed on the decision tree presented in §3.1.1? Answer: nowhere. He does not land in the "proposal is uninteresting" box, because far from being uninteresting, his proposal shares many features with a highly plausible metaethical theory that offers a potential explanation for the existence of the widespread mistaken belief in the inescapable authority of morality. It also offers an appealing recommendation for how we can make better use of the social practice of ethical discourse we have inherited. Neither does he land in the "proposal is self-defeating box", because he does not wish to do away with the social practice of ethical discourse, and therefore can legitimately help himself to any tenable discourse implicating ethical judgements he needs in order to make his case. (Of course, because he needs to restrict himself to *tenable* ethical judgments, he cannot make any that imply the inescapable authority of morality. But he seems to manage to comply with this restriction.) Finally, he does not land in the elusive "true complete abolitionism" box, because as we have seen he wants to "reshape" the practice of ethical discourse, not abolish it.

[40] I am quoting Fraser somewhat out of context here. In this paper he appears to take himself to be discussing *complete abolitionism*. It seems to me that he falls prey to the argument of §3.1 just as much as any complete abolitionist.

So the decision tree diagram does not apply to Blackford, and the reason it does not apply is that the decision tree starts from the question: "Is the judgement that abolitionism *is to be preferred as a U-NWP policy option* to be understood as an ethical judgement?" But Blackford does not subscribe to a metaethical view that satisfies *UMET Condition*, and so he does not subscribe to UMET. And so, whatever his proposal to reshape the practice of ethical discourse amounts to, it does not amount to a U-NWP policy proposal. This shows *Blackford Non-UMET* to be correct, and the argument developed in §3.1 does not apply to his proposal.

Proposal 2: Marks

§3.3.0: Desirism

Marks (2013, 2019) calls his abolitionist proposal "desirism". It is explicitly motivated by UMET, and he makes similar claims to Joyce about the presuppositions of morality. He says that the common view of morality is of something "real and pervasive" (2013: p. 4), emphasises the importance of inescapable authority:

> The essential core of morality, as I understand it, remains its universal, unchanging and absolute authority in matters of human behaviour. If a candidate for morality did not have this feature, then it would be running for the wrong office.
>
> (2013: p. 12.)

His argument for UMET takes the form of a (fairly crude) evolutionary debunking argument relying explicitly on a Harman-style premise to the effect that in order to earn a place in our view of the world, morality must show itself to be necessary to the causal explanation of our moral beliefs (cf. Harman 1977: Chapter 1). This leads him to the conclusion that "there is no such thing as morality" (Marks 2013: p. 16).

Of course, Marks does not want to deny that there is such a thing as morality in the sense of an observable social phenomenon - i.e., in the sense that there exists a fairly entrenched social practice committed to the existence of that which he says doesn't exist. This observable social phenomenon he calls "empirical morality". By empirical morality

Marks means *the belief in metaphysical morality*. ("Metaphysical morality" is his term for morality in the sense of that word as it appears in his conclusion that "there is no such thing as morality".)

Marks also distinguishes between ethics and morality in roughly the way that we have been doing thus far. For Marks, the ethical questions are "what is morality?" and "how shall one live?" His answers, respectively are "a myth" and "without that myth" (p. 3).

To avoid terminological confusion we should compare Marks' terminology with mine. What Marks calls "metaphysical morality" is just what I have been calling "morality" – i.e., it is the target of his MET. What he calls "empirical morality" is very close to what I have been calling "the social practice of ethical discourse". There is a small difference, however, between Marks' use of the term "empirical morality" and my use of the term "the social practice of ethical discourse". This is that the concept of the social practice of ethical discourse is such that we can always ask, for any given case for MET, whether the social practice of ethics necessarily commits the error identified in the case for MET. That is precisely how we distinguish LMETs from UMET. But Marks takes "empirical morality" simply to *be* belief in what he calls metaphysical morality. If this is how "empirical morality" is defined, then it becomes an open question to what extent the social practice of ethical discourse and empirical morality overlap. To get UMET rather than a version of LMET, Marks needs it to be the case that "empirical morality" cannot be reformed so as not to commit itself to "metaphysical morality" without the whole practice collapsing. But Marks steadfastly refuses to consider questions in this territory. This is because he does not want to make controversial empirical claims about how widespread the moral concepts he targets with his EDA are. But this refusal must surely count as a major weakness in his case for UMET. One way in which UMET can be resisted is by rejecting the claim that ethical discourse is in fact committed to inescapable authority (as we saw in §1.4). Sociological facts about the social practice of ethical discourse as we actually find it – from the philosophy journals to the tabloid opinion pieces – must surely be highly relevant to the question.

However, as a UMET-proponent, Marks is of course committed to the claim that the errors in morality are sufficient to render the social practice of ethical discourse

irredeemably error-infected, regardless of the shortcomings in his argument for this claim. So, however weakened his *case* for UMET is by his refusal to engage in sociological analysis of the social practice of ethical discourse, still his *commitment* to UMET makes it legitimate for him to call this practice "empirical morality". So, where it is natural to do so in my discussion of Marks, I shall use his preferred term – "empirical morality". It is to be remembered that empirical morality is ultimately the same thing as the social practice of ethical discourse.

Marks ought to see his task as advertising the benefits of his abolitionist proposal without making the case from the perspective of "empirical morality". That is, he must make the case for it without *joining the social practice of ethical discourse*. To do this, he must avoid making discourse implicating ethical judgements. And *Social Rationality Sufficiency* tells us that in order to do that, he needs to avoid adopting the perspective of social rationality.

Does Marks at least try to avoid the perspective of social rationality? Sometimes it might seem so. On the basis of considerations other than those occupying us right now, Marks proposes a distinction between

> the claim that an amoral regime would leave us (society, humanity) feeling happier or more satisfied than we currently do and the claim that on reflection (e.g. after reading this book) you, the reader, will find empirical amorality more attractive and motivating than empirical morality. The latter claim […] is the one I will defend.
>
> (2013: p. 40.)

Here Marks is using "empirical amorality" as an alternative name for desirism, his abolitionist proposal. So it looks like he might be able to refrain from ever adopting the perspective of social rationality, and that might make his case immune to the complaint of self-defeat. But in fact, he tends to support his desirist policy by appeal to considerations that seem to land him very squarely in the socially rational point of view, and of course if he didn't do this, then his case would be less interesting than it appears to be. And indeed, Marks is happy to be seen doing precisely this:

> I shall certainly feel myself at liberty to put forward considerations suggesting a rosy outcome for a world that embraced amoralism; and that thought would, naturally, form part of the appeal I claim for amorality.
>
> (ibid.)

Why, then, is empirical morality better dropped, according to Marks? He discusses eight alleged disadvantages of empirical morality and seven possible advantages of desirist amoralism. I shall discuss his identified disadvantages of morality in the following section. In the section after that, I will discuss the advertised advantages of abolishing morality.

§3.3.1: Putative Disadvantages of Empirical Morality

With respect to most of the putative disadvantages that Marks discusses, I have the same complaint: it is most implausible that the disadvantage advertised is really characteristic of everything that might be identified as the social practice of ethical discourse. Since Marks is trying to advertise the advantages of abandoning empirical morality, he should be identifying disadvantages of it that pervade more or less the whole of the practice. But he does not do this. The disadvantages he identifies only affect fairly specific phenomena observable in the social practice; moreover, these tend to be things that are themselves suspicious for general ethical reasons. This suggests that the disadvantages that Marks identifies have nothing to do with the putatively near-global presupposition error concerning inescapable authority, and everything to do with certain types of moral discourse being particularly amenable to criticism from within the legitimate ethical web.

Disadvantage #1: Morality is Angry

"Morality," Marks tells us, "is an emotionally fraught phenomenon" (2013: p. 40). And anger, he thinks, is probably the central moral emotion:

> I myself have become so sensitized to the morality-anger connection that I see it now as almost an identity. In other words, it is not only that being moral often, even typically, means being angry, but also that being angry means being moral. For when I consider myself being angry about something, I find that there is an implicit judgement that that about which or about whom I am angry is somehow (morally) wrong or bad.
>
> (ibid.)

This strong claim – that the link between empirical morality and anger is nearly one of identity – must be rejected. Considering the identity in the direction "being angry entails being moral", this seems to me simply a quirk of Marks' personal psychology. (Personally, I have often found myself angry without the slightest sense that my anger was justified from the point of view of the social practice of ethical discourse.) In the other direction, "moral judgement entails moral anger", Marks is on slightly stronger ground. Anger is after all one of the three central moral emotions in the CAD account of moral psychology (Prinz 2007). Still, it is far from clear that if we restrict our attention to those parts of the social practice of ethical discourse that have nothing to do with anger, we do not find enough to ground an ethical practice of adopting the socially rational point of view. But this is what Marks would need for the connection between empirical morality and anger to be relevant as a consideration against retaining the social practice of ethical discourse.

In fact, criticism of unhelpful moral anger occurs not infrequently in ethical discourse as we find it. Most obviously, Christian ethics often shows disapproval of anger. (Examples: Psalm 37:8; Proverbs 14:29; Ecclesiastes 7:9; James 1:19-20.) And in advertising the reduction of anger as a benefit of amorality, Marks himself must surely count as an ethical critic of anger. The self-defeat here is obvious.

Disadvantage #2: Morality is Hypocritical

Here the idea is that morality tends to cloak real motives. "Do you suppose it is only a coincidence that the vast majority of the time that you deem something to be morally wrong, it is something that you don't like on other grounds?" Marks asks (2013: p. 42). The overall complaint is that moral disputants present themselves as responding to moral reasons, when in fact they are responding to no such thing.

But of course any sociologist of ethics will find that in general, a demonstration that a person believes some moral proposition *z* purely because *z* expresses moral disapproval of something that they don't like on other grounds is taken, within the institution of the social practice of ethical discourse, as evidence that this person's belief that *z* is less reliable than it might initially seem. So consider the set of all possible ethical judgements that could possibly be made. Divide these into two groups: in set A we have all those

ethical judgements that are unreliable because biased; in set B we have those ethical judgements which have been made in an unbiased way. Marks' criticism only applies to set A. But, not only is there no reason to suppose that set A constitutes the whole of the social practice of ethics, the suspicious credentials of the judgements in set A is emphasised by the practice itself. (Kramer 2009: Chapter 7 emphasises impartiality as a form of moral objectivity.)

Disadvantage #3: Morality is Arrogant

Here, Marks is keen to rally against moral posturing. There appear to be two parts to this critique. First, he is concerned with the *egotism* of moral discourse. That is, he objects to moral discourse the aim of which is to improve the speaker's own self-image and "public acclaim". But this is just moral grandstanding (Tosi and Warmke 2016), which is the subject of much ethical criticism. (Just consider current arguments over the perceived "virtue signalling" of certain proponents of left-wing identity politics.) So this disadvantage does not apply to a sufficiently significant portion of the practice of ethical discourse to make the rejection of empirical morality attractive.

Second, Marks is concerned that "this sort of posturing can reach orgiastic proportions. One typical place to observe this is at a murder trial – again , the more heinous, the better" (2013: p. 43). This seems to recapitulate themes from his (2011), in which he was writing for a non-philosophical audience and bemoaned what he called "moral pornography". But one doesn't have to travel to the fringes of society to find ethical criticism of this style of moralising. Once again, Marks' target is too narrow to be relevant to a sweeping form of abolitionism that recommends jettisoning the whole practice of ethical discourse.

Disadvantage #4: Morality is Arbitrary

Marks complains that dialectical moves typical of moral discourse, like "because it's wrong!" or "because it's the right thing to do!" "are either empty of content or thoroughly ambiguous" (p. 44). Now, if AfO is sound, then he is of course entirely right about this.

Nonetheless, two questions immediately arise: can these expressions be given determinate, non-ambiguous content, and if so, do we actually find such attempts when we turn our sociological gaze to ethics? For suppose the situation turns out to be as follows. Often, dialectical moves of the sort that Marks questions here are in fact empty

of content, or ambiguous. Nonetheless, there is perfect sense to be made of dialectical moves like this; it is just that they are often made by people who have not bothered to make any sense of them, such that it is entirely obscure what they might mean when they say that something is wrong. Furthermore, often these dialectical moves are made by those who know how to make sense of them in this way. If the situation turned out to be like this, then once again Marks' actual target would turn out to be narrower than he thinks: he would be claiming that these sorts of dialectical moves are *in general* "empty of content or thoroughly ambiguous" when the most the sociological data would support is that these sorts of dialectical moves are *often* empty of content or thoroughly ambiguous.

Now, in Chapter 5 I will attempt to show that there is sense to be made of these dialectical moves in ways that avoid all the problems of inescapable authority, truths concerning foundational propositions about morality, irreducible normativity and the like. If this attempt is successful, then the disadvantage of ethical discourse that Marks identifies here is not as pervasive as it would need to be in order to justify complete abolitionism.

Disadvantage #5: Morality is Imprudent

This putative disadvantage turns out to be an extension of disadvantage #3, and related to disadvantage #6, discussed below. The idea behind disadvantage #5 is that one can be led by (unfounded) moral certainty into acting on principle in ways that have disastrous or unwelcome consequences. Marks gives what he describes as a "trivial example" by way of illustration: you receive bad service at a restaurant. Justice seems to demand leaving a small tip or none at all. However, this will not only displease the waiter but also increase the likelihood that you will once again receive poor service should you end up eating at the same restaurant and being served by the same waiter. A less trivial example with the same essential logical structure might be: popular public (retributive) sentiment demands harsh punishment for crime, but this can lead to undesirable outcomes from the point of view of recidivism rates. But once again, Marks' target is too narrow: many approaches to ethical discourse take the form of explicit appeal to consequentialist reasoning. So the

disadvantage of morality here, whilst real and relevant, does not target the whole observable phenomenon of the social practice of ethical discourse.

Disadvantage #6: Morality is Intransigent

Here the basic complaint is that morality, because of its absolutism, leads to intransigence and thus blocks the way to compromise. But this putative disadvantage is closely connected with disadvantages #3 and #5 above, and similar comments to those made under those headings apply here.

Disadvantage #7: Morality is Useless

Here the complaint is that "it is simply not informative to tell somebody (including oneself) to 'do the right thing' because, as noted earlier, that phrase is empty of content" (p. 46). Obviously, this complaint presupposes that the complaint embodied above under the heading of disadvantage #4. So those comments apply here too.

Disadvantage #8: Morality is Silly

Here Marks illustrates his complaint by asking us to consider his experience of attending a colloquium being given by an academic philosopher who was discussing the ethics of abortion. She presented an argument for the conclusion that abortion was morally permissible. Marks' reaction I interpret as a sudden appreciation of a sort of cognitive chasm that separates us from the contents of our moral beliefs, rendering those beliefs impossible to maintain:

> I remember the moment it happened. I was attending a colloquium being given by an academic philosopher, who was defending a novel theory of why abortion is (sometimes) morally permissible. As a knee-jerk "liberal" I have always supported abortion rights. However on this particular occasion I was suddenly struck by the intricacy of the philosopher's argument and the absurdity of believing (or assuming) that some reasoning of this sort must underlie what, "in fact" makes abortion sometimes permissible. […] I therefore raised my hand and asked my colleague, "Do you really believe there is a moral truth of the matter about abortion?" To my newfound amazement, she answered, "Yes". Which is exactly how I would have replied a few years previously. Now I can only marvel at this attitude – just as how I marvel at how "grown" people in 21st Century America can believe in miracles, even highly intelligent and educated people in Ivy League seminaries.

I can understand the nature of the vivid thought experience that Marks describes as his reaction to the philosopher's answering his question in the affirmative. These sorts of

thought experiences are part and parcel of reflective life, and moral discourse is particularly apt to trigger them. Of all of Marks' eight points, this is perhaps the one to which I am most sympathetic. But we shouldn't allow the thought experience to dictate too much of our view of ethical reality. Most importantly, the affirmative answer to the question "is there a moral truth of the matter about X?" could be fleshed out in a myriad of ways, as the vast metaethical literature attests. Some of these ways are revisionary, in such a way that the notion of moral truth grounded cannot be targeted by Marks' critique, which defines morality at the outset as involving inescapable authority. To the extent that these attempts are identifiably a part of the huge mass of discourse that constitutes the relevant social practice, Marks' putative disadvantage #8, like all the forgoing ones, does not apply to the whole of this practice.

§3.3.2: Marks' Proposed Advantages of Amoralism

Marks advertises seven advantages of his brand of abolitionism. For my purposes, the most important question is not whether these advantages really obtain, or whether they really are advantages. The question is: can Marks argue that they are advantages consistently with his own error theoretical view? That is, can he make his arguments entirely from the safely non-error-infected domain of practical rationality? Or does the defence of amoralism on the grounds of these putative advantages necessarily stray into the realm of social rationality? As in the previous section, I shall go through his advertised advantages in turn.

Advantage #1: Amorality is guilt-free

Marks describes the feeling of guilt as "a ubiquitous pain in the human psyche and society" (p. 48) and argues, somewhat simplistically, that "if there were no such thing as morality (not to mention, sin) or, more to the point, no *belief* in morality, then the perpetual guilt machine would be effectively shut down; amorality would have sabotaged it" (p. 49).

I said above that the question in this section is not whether the advantages that Marks considers really would accrue to amoralists, or whether they really are advantages. But in

this instance it is relevant that the advantageousness of dispensing with guilt altogether is dubious. Consider these words of Blackburn's:

> There is a way of thinking, more common in popular psychology texts than in philosophy or literature, that invites us to think of guilt and shame as bad feelings, like nausea, that we ought just to wish away. They are there to be cured. But that is too simple. Guilt, for instance, typically involves the wish to have done otherwise, and if I really wish to have done otherwise, I won't find that wish just a brute uncomfortable fact about my own consciousness, one that I might in turn wish away. My last word is not "this is a nasty state to be in, so I wish I could get rid of it". My last word is "I wish I had done otherwise".
>
> (Blackburn 1998: p. 20.)

My point here is not to rule in favour of Blackburn and in against Marks. Perhaps guilt might be better jettisoned along with the specific morality system to which it naturally belongs, leaving shame – an emotion that is better able to understand itself (and guilt) than guilt is able to understand itself.[41] My point is just that sceptics like Blackburn seem to be owed an explanation as to why we really would be better off without guilt.

But in order to provide this, Marks would clearly have to enter into the realms of ethical discourse. Would it be legitimate for him to do so? If so, then there is enough of the ethical web left intact after the error-infected parts of it have been excised for UMET to be false. At most LMET could be true. If, on the other hand, it is not legitimate for Marks to enter the necessary realm of the ethical in order to defend his claim that the guiltlessness of amorality really is an advantage, then amorality cannot reasonably be defended by appeal to this putative advantage.

Advantage #2: Amorality is tolerant

Here Marks writes:

[41] This, of course, is a reference to Williams (1993). Even if the characterisation of Williams' view is correct, which is open to dispute, still Williams was obviously only an LMET proponent and brought many complex and carefully examined ethical resources to the project of illuminating the emotions of guilt and shame and their connection to practical necessity of various sorts. This is not characteristic of Marks' writings, and I am fairly sure that Marks would want to eliminate shame as well. If he did so wish, then the "advantage" of amoralism here would be highly dubious indeed. If he did not so wish, then what other reason could there be for resisting the elimination of the emotion of shame than its social utility in promoting cooperation and cohesion? But making a case like that would put Marks firmly in the territory of social rationality, contra his UMET commitment.

> In this country [i.e., the USA, but I assume he would make similar claims about the UK and Australia] one finds examples everywhere of laws or attempted legislation to enforce certain conceptions of right and wrong, whether about drug use, sexual behaviour, or how to die. I have been involved in movements of resistance to the imposition of uniform standards in some of these areas, but previously I saw myself as a member of the moral opposition to religious subversion of the secular polity. Now I would say that the enemy has all along been as much morality as religion, for what I dispute is the quintessentially moralist conviction that one's preferences are intuitions of a universalising imperative reality.
>
> (2013: p. 49.)

But it is not so easy to oppose interference by moralist busy-bodies without joining the practice of ethical discourse. *Why* is it an advantage of amoralism that it is tolerant? Presumably, that it cuts down on conflict. But then why is *that* a good thing? Assuming the conflict is needless, it seems obvious that less of it is better – but this is because the needless instigation of conflict is so obviously irrational from the social point of view. (From some *individual* points of view, conflict is obviously advantageous. To generate a general conclusion of the sort that Marks wants here, you *have* to adopt the socially rational point of view.)

Advantage #3: Amorality is interesting

The discussion that Marks provides us with here is an extremely odd advert for amoralism, in that it seems to have little to do with the topic at hand. He says that he is "learning to appreciate how interesting the world can be when moral conviction is replaced by boundless curiosity" (p. 51) and gives the example of an experience he had when a couple of teenage girls stepped off the curb into the path of his car as he was driving down a street one day. Before his conversion to amoralism, he tells us, he would have been seething with rage that they would inconvenience him like this. Instead, this time, he was merely curious as to what they could have been thinking.

I have no application of the pattern of argument I am applying to most of Marks' other touted advantages of amorality with respect to this one. Instead I will offer the observation that discourse implicating ethical conviction is no barrier to boundless curiosity. For example, a school leaver might be attracted to the study of psychology at university because a career as a forensic psychologist appeals to them. This might be because of

boundless curiosity concerning the minds of those who commit certain crimes. This can obviously co-exist with the discourse implicating ethical conviction that what these individuals do is wrong; indeed, part of the appeal to the school leaver of a career in forensic psychology might be that they believe it will be ethically worthwhile – that is, highly defensible from the perspective of social rationality – to spend their career doing something that can help to reduce crime.

In sum, what Marks reports here appears to be more a quirk of his own psychology (i.e. that he was previously prone to road-rage and has since gained stoic-like control over his judgements) than a substantial advantage of amoralism over moralism.

Advantage #4: Amoralism is explanatory

Here the claim is that "amorality has the potential to resolve various anomalies that beset morality". The sole example he gives is moral luck: why should a drunken driver who kills a child be judged more severely than one who does not, even though both behaved in the same way, and took the same risks? Marks claims that amoralism offers the explanation that "it is natural for desire to focus more on what has actually happened than on motives or other causal factors".

But the same question arises as usual: why is this an advantage? The question is one of justification: *ought* we care more about the drunk driver who is unfortunate enough to kill someone than the one who luckily arrives home without incident? Marks can only offer the fact that amoralism explains why we do this as an advantage if we ought to do this. But this brings us immediately into the realm of ethical discourse.

Advantage #4: Amoralism is simple

Here the idea is that amoralism strips away notions of: moral obligation, moral responsibility, moral goodness, moral rights, conscience, and many others besides. Marks touts this as an advantage. But whether or not this is an advantage depends on whether these things that have been stripped away were desirable or not. Marks does not adduce any considerations to show that these things are undesirable, and adjudicating on this matter is likely to involve joining the social practice of ethical discourse.

Advantage #5: Amorality is compassionate

Here Marks writes of how when "moral judgement is cast aside, my heart aches to see the silent video of the prematurely aged Osama bin Laden davening and stroking his beard in front of his little TV set in his shabby room in his last hideout in Abbottabad." The individual judgement that this represents an improvement in his sensibility may well be an excellent candidate for being a non-discourse implicating ethical judgement. Nonetheless, accounts of the value of compassion do have a rich history in the social practice of ethical discourse, and it does seem to me that Marks is tapping into this history here. Unlike with some of the other putative advantages of desirism, the self-defeat here is not automatic. But there is certainly a risk of it.

§3.3.3: Marks' Reply to the Charge of Self-Defeat

I have been arguing that Marks involves himself in self-defeat. He attempts to defend a thoroughgoing and wide-ranging abolitionism, but in motivating it consistently needs to venture into territory which on UMET must be error-infected. Two parts of the discussion in *Ethics Without Morals* need attention in order to see what Marks' response to this type of objection is, and why it doesn't adequately defuse the objection as I have formulated it here.

In Chapter 5, Marks considers the possibility that "the very notion of amorality I have put forward could count as morality on a different rendering" (p. 56). He says that there are two ways in which this objection might be formed. The first is that "as a matter of fact, 'morality' has all along meant something closer to what I have been calling 'amorality' than to what I have been calling 'morality'" (ibid). This is the closest to the way I have presented the objection. Unfortunately, Marks *does not answer this version of the objection*. This is for the reason noted in §3.2.0, that Marks does not want to enter into sociological analysis of how widespread is the commitment to the inescapable authority of morality. Joyce went to considerable lengths to stress how important to morality is the commitment to NICRs. Marks does no such thing. This obviously represents a serious failing in his case for UMET. (Of course, he could simply appeal to Joyce's arguments.)

The second way in which Marks says the possibility that the very notion of amorality he has been defending could count as morality is that "it would make more sense to *reform* the definition of morality along the lines of what I have been calling 'amorality' than to retain the old meaning and thence discard morality" (ibid). This seems to be a reference to the policy proposal of substitutionalism, which will occupy us in Chapter 5.

The other part of *Ethics Without Morals* in which Marks presents material that could be used to respond to the objection to his amoralism that I have presented in this part of the present chapter occurs in Chapter 7, where he tries to explain what ethics is. (The chapter is entitled 'What is Ethics?') What he tries to do is explain the distinction between ethics and morality by giving three ways in which they differ. In this way he hopes to show that his case for desirism has been harmlessly ethical, not self-confoundingly moral. But the fact that he sees the relevant distinction as that between ethics and morality is the *source* of the problem of self-defeat, not part of its solution. For the first difference that Marks offers is that ethics is hypothetical, not categorical. But the important distinction is not between the hypothetical-ethical and categorical-moral, but between the discourse implicating ethical and non-discourse implicating ethical. This is because Marks is a UMET-proponent arguing in favour of complete abolitionism, not an LMET-proponent whose diagnosis of error in the ethical web leaves enough of it still standing to ground a social practice of ethical discourse. Marks should say not just that legitimate ethics is hypothetical, but furthermore that it is non-discourse implicating. But this would be a bad fit with the advertisement of all the benefits of desirism that he has provided.

The second difference Marks offers is that ethics is practical. Here the claim is that "amorality, unlike morality, holds out the real hope of reaching conclusions about ethical issues". The example he gives here is, in a way, quite extraordinary:

> When I became a step-father I suddenly faced mind-boggling issues which, at the time, I conceived as moral ones. For instance, as we all climbed into my wife's car one day, I naturally headed for the passenger seat, but so did my tween stepson. Both of us had been used to occupying that seat, since theretofore we had never all been in my wife's/his mother's car at the same time. Now what? Well, I automatically conceived the question as a moral one; and it should come as no surprise, as my case against morality has made plain, that my moral conclusion coincided with my personal

> preference: *A spouse/parent should have pride of place in any family car*. I was not about to be treated like a child who sits in the back seat.
>
> (p. 87.)

The point Marks aims to make with this example is that with morality, clashes like this are irresolvable, because both parties feel themselves to have the justificatory clout of morality behind them, whereas when morality is discarded, compromise is possible. That is what he tells us happened in the car seat case: when he abandoned morality he was able to reach a sensible compromise with his step-son.

It seems unlikely either that the benefits of dispensing with ethical discourse that Marks gained in the car seat kerfuffle will automatically follow in all situations of conflict or that Marks could not have gained these benefits by seeing the issue as one to be settled by the method of finding the correct discourse implicating ethical judgement that applied to his circumstances. In disagreements where more is at stake, between parties who have fewer reasons to foster warm and friendly relations than step-parents and step-children do, matters are unlikely to be magically resolved by the simple expedient of giving up morality. And in the precise disagreement Marks describes, it is plausible to suppose that the social practice of ethical discourse has the resources to support the discourse implicating ethical judgement that Marks ethically *ought* to compromise with his step-son.

Once again, this characteristic of ethics as opposed to morality does nothing to distinguish the putatively error-infected social practice of ethical discourse from a more minimal evaluative approach, because both are alike in respect of the proffered distinction. Both sometimes "hold out real hope of reaching conclusions about ethical issues" and sometimes do not, depending on the nature of the conflict.

The third and final characteristic is that ethics is motivating. With respect to morality, Marks says, "it is a commonplace that a person can conclude or believe that s/he ought to do something and still does not do it, or that s/he ought not to do it but does it anyway". Ethics, understood in the desirist way, is not like this. Here, Marks suggests that although hypothetical imperatives are not error-infected as (inescapably normative) categorical imperatives are, we would do well to eliminate even hypothetical imperatives from our

normative repertoire. Instead of saying "if you wish to master the cello, you should practice every day", we could say "if you wished to master the cello, you would practice every day". Marks makes this suggestion because of the "moral baggage" that the words "should" and "ought" and the like bring with them: even when appearing in hypothetical imperatives, these terms are "easily misunderstood, given our long history of using these terms absolutistically" (p. 90). And now there is no room for desirist ethics to fail to motivate: if Anna desired to master the cello she would practice every day. Does Anna not practice every day? Then it would appear that she doesn't really desire to master the cello. As Marks puts it:

> We could say: in lieu of prescription, prediction. This is the sum total of ethics. This is desirism. This is amorality.
>
> (p. 91.)

Unlike with the first characteristic that Marks proposed for distinguishing ethics from morality, my trouble with this is not that the characteristic does not do anything to distinguish the two. If the idea is that desirism, having as it does desires as its only currency, with nothing of even hypothetically prescriptive force, is to be distinguished from empirical morality on that count, then I am happy to agree that desirism manages to stand apart from the social practice of ethical discourse. The trouble is that this way of delineating the moral is a terrible fit with the advertised advantages of ethics over morality discussed in §3.3.2. There, we saw Marks consistently and repeatedly *recommend* jettisoning morality and adopting instead a desirist perspective. But apparently, we are to learn at the conclusion of this book that desirism has no means of recommending anything: it is not in the business of recommending, it is in the business of predicting. Quite apart from the fact that it is hard to recognise mere predictions as any sort of ethical judgements, even non-discourse implicating ones, this involves Marks in the most basic inconsistency: he appears to be recommending that we adopt a perspective that dispenses entirely with recommendations.

§3.5: Might There Be Stronger Abolitionist Proposals To Be Found?

Now, I don't think it should be too controversial to say that the abolitionist proposal which I used in §3.3 to illustrate the argument of §3.1 is especially weak. This might invite an objection: I have not considered the strongest possible complete abolitionist proposal, and so it is too early to conclude that any complete abolitionist proposal will necessarily conform to the pattern suggested in §3.1. Thus an abolitionist could escape my dilemma by formulating a better complete abolitionist proposal.

This is an objection that I think it is worth taking seriously, but it is difficult to respond to it adequately in the absence of a detailed abolitionist proposal that improves upon Marks'. So the first thing to say in response to this objection is simply to concede that there is a limitation in my analysis here: a future abolitionist might be able to come up with a better proposal that escapes my critique. I suspect, however, that the prospects for abolitionists here are dim, for the reason discussed in §3.1.1: in order to be *interesting*, a case for abolitionism will likely have to consist of discourse implicating ethical judgements, and thus the argument will be self-defeating. Viewed in this way, Marks' discussion merely offers an especially clear example of a pattern we should not be too hopeful to see avoided in any future abolitionist proposals.

The second thing to say in response to this objection is that the suspicion I articulated in the previous paragraph seems to be strengthened rather than weakened by abolitionists' track record. I have already mentioned Hinkfuss and Garner, and shown that the considerations they adduce in favour of their abolitionist proposals are made from the perspective of social rationality. So, they too must involve themselves in self-defeat if they are to be interpreted as complete abolitionists. There are other abolitionist proposals in the literature that do not even present themselves as policy options with respect to U-NWP, and for this reason are likely to be of fairly limited help to the proponent of abolitionism as a solution to U-NWP. This is the case with Moeller (2009, 2019), whose abolitionist proposal is derived in part from Daoist wisdom about emptying the heart-mind. It seems to me that a UMET-proponent who adopted this philosophy would soon

cease to be a UMET-proponent: the discussion in Chapter 1 will not be of interest to someone who has emptied their heart-mind.

And anyway, Moeller too seems to involve himself in self-defeat. One main reason for this is that he advertises law as one of two "antidotes" to morality, and this comes with deeper problems. Moeller claims that law has reached the point where it is able to function amorally. This claim is based on viewing law as an autopoietic social system as in the social systems theory of Luhmann (2008). Regardless of what is to be said for and against social systems theory, it seems to me to be very difficult to find any system of law currently operating on Earth that does not claim moral authority for itself in some way. Of course, to make good on this claim would take me far into the waters of philosophy of law, so I shall leave this matter here due to considerations of space.

The third and final thing to say in response to this objection is simply that even if an abolitionist proposal can be formulated which escapes the argument of the present chapter, still the proponent of this proposal, as a proponent of UMET, will face the objection of the final chapter: the objection from loss.

Chapter 4:
Substitutionalism

§4.0

On the basis of the discussion in the previous chapter, we can conclude that abolitionism is probably not an available U-NWP option. The problem with it is that it any interesting attempt to motivate it will inevitably be incompatible with the abolitionist's own error-theoretical motivation for supporting the policy. In this chapter I want to make a similar argument about substitutionalism. Substitutionalism is a form of revisionism; as I will explain in more detail presently, it is the recommendation that we reorient moral thought and talk away from the error-infected concepts currently associated with it, and towards new hygienic concepts that can play the same or similar roles.

Substitutionalism takes inspiration from versions of moral realism that are revisionary in some way. Herein, the substitutionalist might hope, lies a great strength of substitutionalism: the past careful work of moral realists of a revisionist bent can be repurposed to provide a neat solution to U-NWP *even though these analyses fail to vindicate moral realism as their authors intended*. However, I think instead that the very opposite is the case, and that herein lies a great weakness of substitutionalism: the more the substitutionalist does to convince us that the policy proposal could be made to work, the less credible error theory becomes. All the substitutionalist's toil is for the revisionary moral realist in the end. Or so I shall argue.

Substitutionalism – The Policy

§4.1.0: Salvaged Concepts and Revised Concepts

Here is how substitutionalism works. Consider an error-infected moral concept like **MORALLY WRONG**. On UMET, this is a defective concept. But suppose that we could identify some alternative concept, **β**, that is *close to* **MORALLY WRONG** in the sense that it can be put to many of the same practical uses, and suppose further that **β** is not

error infected. The proposal is that we could cease using the term "morally wrong" to pick out the concept **MORALLY WRONG** and to begin using it instead to pick out **β** instead.[42]

The concept **β** does not need to be a *perfect* replacement for the error-infected concept it replaces. It just needs to be able to be put to *many* of the same uses in *many* contexts. For example, one of the things that a substitutionalist will want from the β-concepts will presumably be that said concepts can function so as to secure for the error theorist the benefits of social co-ordination that moral discourse can secure. If an error theorist is part of a group that considers itself to have some moral duty to some other group, for example, and there is debate within the group to which the error theorist belongs about precisely what action this duty demands of them, the error theorist will find it convenient to have some β-concept that replaces the error-infected **MORAL DUTY**, and can be used to avoid distracting and pointless diversions into metaethical territory when contributing to this debate. However, it might be that where dispute is more fundamental, the difference between the original error-infected moral concepts and the β-concepts makes some sort of normative difference. In that case, the error theorist might have to bring metaethics into the argument. Still, so long as these types of situations are rare enough, perhaps substitutionalism can remain a viable solution to U-NWP. That, at least, is the promise.

These β-concepts, Lutz (2014) calls ***salvaged concepts***. We should think for a moment about the process of identifying them. The substitutionalist will need to pick such a concept out by means of a ***reforming definition***. And reforming definitions of moral concepts have of course been proposed by those who do not endorse UMET. For example, Lewis gives the following definition of the term 'value': "Something of the appropriate

[42] How essential is it to the substance of the substitutionalist proposal that we use the same term? Suppose a UMET-proponent impressed by the feasibility of the substitutionalist proposal but foreseeing all sorts of linguistic confusion decides to introduce the term "exchortally wrong" to pick out the salvaged concept. Will the arguments that I will develop over the course of this chapter target such a policy? This depends on what the substitutionalist wants to say about exchortality in general. If they want to say that exchortality is close enough to morality that an exchortalist and a moralist can profitably discuss many or most ethical matters without the differences between morality and exchortality making themselves felt too much, then this substitutionalist faces the arguments of the present chapter. If the claim is that exchortality is sufficiently different to morality that an exchortalist cannot engage in distinctly moral arguments, then the exchortalist escapes the arguments of this chapter, but now appears to be a partial abolitionist. But we saw in the previous chapter that partial abolitionists are best seen as LMET-proponents, not UMET-proponents. So in either case, replacing moral talk with exchortal talk appears not to be a move available to UMET-proponents.

category is a value if and only if we would be disposed, under ideal circumstances, to value it" (1989: p. 113). Although this seems to him an adequate analysis of the term "value", he explicitly acknowledges that it does not quite capture the *entirety* of our pre-theoretical ideas of what an analysis of "value" should capture. Indeed, Lewis thought that the question of whether his analysis showed that value was something different to our pre-theoretical idea of it, or that there was no such thing as value, was essentially "a matter of temperament":

> What to make of the situation is mainly a matter of temperament. You can bang the drum about how philosophy has uncovered a terrible secret: there are no values! (Shock horror: no such thing as simultaneity! Nobody ever whistled whilst he worked!) You can shout it from the housetops – browbeating is oppression, the truth shall make you free. Or you can think it better for public safety to keep quiet and hope people will go on as before. Or you can declare that there are no values, but that nevertheless it is legitimate – and not just expedient – for us to carry on with value-talk, since we can make it all go smoothly if we just give the name of value to claimants that don't quite deserve it. […] When it comes to deserving a name, there's better and worse but who's to say how good is good enough?
>
> (Lewis 1989: p. 137.)

Who indeed? Well, substitutionalists, apparently. They think they can have solid reasons for believing that no β-concepts are going to "deserve the name", even though it *is* expedient to attach those names to those concepts. The coherence of their policy proposal depends on its being the case that it *is* expedient but not quite legitimate to use β-concepts for moral concepts. In the following subsection, we shall see that Lutz has a principled method for sorting concepts that 'deserve the name' from those that don't.

Let's call the concept picked out by Lewis' reforming definition **LEWIS-VALUE**. When he says that it is a matter of mere temperament whether **LEWIS-VALUE** deserves the name "value", I understand him to be saying that it is *indeterminate*. Meanwhile, it will be important to substitutionalism as I understand it here that is determinate that this concept doesn't deserve the name "value" (or, at least, that it doesn't deserve the name "moral value"). Fairly obviously, we can imagine a third view here: that it is determinate that this concept *does* deserve the name. Indeed, reforming definitions have been offered by those who appear to hold that the concepts picked out by their reforming definitions

determinately do deserve the name. Consider Railton's (1986) identification of moral rightness with that which is rational from the social point of view, where rationality is understood in the instrumental sense. This 'misses out' categoricity, and yet Railton does not say that whether or not the concept **THAT WHICH IS RATIONAL FROM THE SOCIAL POINT OF VIEW** deserves the name 'moral rightness' is indeterminate: he is trying to offer a brand of moral realism. I shall call those who offer reforming definitions of moral terms in order to present a metaethical theory that competes with UMET for acceptance – as does Railton's moral realism – *revisionists*.[43] I shall say that whilst substitutionalists use reforming definitions to pick out salvaged concepts, revisionists use their reforming definitions to pick out *revised concepts*. There is a difference between substitutionalists and revisionists concerning their use of reforming definitions: substitutionalists' interest in reforming definitions is revolutionary whereas revisionists' interests are more hermeneutic. I shall therefore distinguish between the use of reforming definitions to pick out salvaged concepts and revised concepts by distinguishing between *revolutionary definitions* and *revisionary definitions*. (See the table at the end of this subsection summarising the stipulative use of terms throughout this chapter.)

The point I wish all this stipulative terminology to render salient is that for any reforming definition, a *pair* of concepts is always available: a salvaged concept and an analogous revised concept. Accepting the reformed concept as a revised concept is to reject UMET; it is to buy into revisionism instead. Accepting the reformed concept as a salvaged concept is to first reject it as a revised concept. Because this point is crucial for my purposes, I am going to labour it somewhat with three quick examples. I think that this will prove useful for what is to come.

Example 1: Railton

This example has already been mentioned. Suppose we take as a reforming definition of the term "moral rightness" "that which is rational from the social point of view". Railton

[43] It is slightly unfortunate that 'revisionism' is also used to name the type of U-NWP policy proposal of which substitutionalism is a version. But for the remainder of this chapter, I will not use the term 'revisionism' to refer to any U-NWP policy proposal; I will use it only to refer to non-error theoretical metaethical views that make use of re-examining definitions. This should avoid confusion. Furthermore, I have already argued in the previous chapter that U-NWP policies that recommend revising moral discourse in some particular way – by simply cutting out talk of obligation, say – are not best seen as U-NWP policies anyway. They are versions of LMET.

may have been offering this reforming definition as part of the case for a metaethical view that directly competes with UMET, but there is nothing to stop the substitutionalist rejecting Railton's naturalistic moral realism and taking the concept **THAT WHICH IS RATIONAL FROM THE SOCIAL POINT OF VIEW** as a salvaged concept, to be used where appropriate as a *replacement* for the error-infected concept **MORALLY RIGHT**. So here we have a pair of β-concepts, one a candidate salvaged concept, the other a candidate reformed concept, both picked out by the same definition.

Example 2: Expressivism

Some substitutionalists are revolutionary expressivists – they think that error theorists' best bet for navigating their way through those parts of life that require normative thought and talk is to substitute in expressivist moral concepts for error-infected moral concepts. (Kohler & Ridge's revolutionary expressivism will be briefly discussed in §4.1.3.) But we can easily imagine an expressivist revisionist.[44] Such a revisionist might claim, for example, that expressivism cannot account for moral objectivity and authority, and might consider their own expressivism to be revisionary on this count. Once again, we have a pair of β-concepts, one a candidate salvaged concept, the other a candidate reformed concept, both picked out by the same definition.

Example 3: Finlay

As already discussed in §1.4, Finlay' analysis of "morally wrong" is relational: to say that *x* is morally wrong is to say that *x* violates certain standards or frustrates certain ends held by the speaker to be of overriding importance (Finlay 2014). This dispenses with inescapable authority – the main presupposition of morality identified by AfR in Chapter 1. Thus, it is a candidate revisionist account.[45] But once again, we can construct a pair by considering an analogous reforming definition used not for Finlay's own hermeneutic purposes but for a suitably minded substitutionalist's revolutionary purpose. Such a

[44] True, most modern expressivists tend to argue that their position is not revisionist. Blackburn, for instance, is fond of advertising the ability of his expressivist quasi-realism to accommodate all our pretheoretical intuitions about morality. Still, one could buy into an expressivist theory like his whilst disagreeing with him about the extent to which it is revisionary.

[45] It would be revisionist if we were convinced, with the error theorist, that most standard users of moral discourse do in fact suppose that moral norms are inescapably authoritative. Finlay does not think that this is the case, and so for him, his own theory is not revisionist.

character would insist that because Finlay's analysis does not accommodate inescapable authority, it is defective as an interpretation of what speakers actually mean when they utter sentences that ascribe moral wrongness to actions. However, this character would continue, the concept that can be extracted from Finlay's analysis is useful for the error theorist who wants a way of participating in moral discourse without always needing to embark on lengthy metaethical digressions. Yet again, we have a pair of β-concepts, one a candidate salvaged concept, the other a candidate reformed concept, both picked out by the same definition.

In general

These examples seem entirely generalisable. Consider any candidate salvaged concept a substitutionalist could possibly formulate. Then we can use the definition the substitutionalist provides to pick out instead a candidate revised concept. There is always the possibility that this candidate revised concept will provide a plausible form of revisionism. Working in the opposite direction: consider any revised moral concept. The possibility cannot be dismissed out of hand that an error theorist will be able to provide a convincing argument to the effect that this revised concept does not deserve the moral name the revisionist gives it. Such an error theorist, if they find substitutionalism attractive, might nonetheless wish to purloin the revisionist's analysis as the basis on which to formulate a candidate salvaged concept.

Summary of Stipulative Terminology

	Option A	**Option B**
Name of View	UMET, with substitutionalism the favoured U-NWP policy option	Revisionism
Type of Reforming Definitions Offered	Revolutionary definitions	Revisionary definitions
Type of concepts picked out by reforming definitions	Salvaged concepts	Revised concepts.

§4.1.1: Negotiable and non-Negotiable Platitudes

So far, I have provided only a basic sketch of the substitutionalist proposal. I shall now try to bring a more detailed picture into view. I'll start with Lutz, who is a useful point because his concerns are so general: he does not propose a particular revolutionary definition of "morally wrong", instead he is content simply to make plausible the idea that some such suitable substitution of hygienic concepts for old error-infected ones could be made to work.

So, let's start by following Lutz (2014: p362ff) in returning to thinking about why, according to UMET, the concept **MORALLY WRONG** is error infected in the first place. The problem derives, Lutz reminds us, from the fact that the meaning of **MORALLY WRONG** is governed by a set of platitudes. Here Lutz reminds us of the distinction between negotiable and non-negotiable platitudes that was discussed in Chapter 1. On the one hand we have non-negotiable platitudes. These concern *what it takes to be* a moral concept. (Recall from Chapter 1 Joyce's non-negotiable platitude concerning NICRs.) Then, there are negotiable platitudes: these are less strict. It might be a negotiable platitude concerning the moral concepts that pain is a morally relevant feature of any scenario; that a possible action I could take would help to alleviate the pain someone is in can provide me with a moral reason to perform that action.

With the distinction between negotiable and non-negotiable commitments at the forefront of our minds, Lutz makes his proposal:

> Let us assume for a moment that there is nothing that satisfies all of our commitments about morality, but there is something that satisfies all but one of our commitments about morality. (Call the unsatisfied commitment the *defective commitment*. Call the remaining cluster of satisfied commitments the *salvaged concept*.) Suppose we make the following recommendation: While there is nothing that satisfies all of our commitments about morality, the salvaged concept is close enough; let's use our moral language to talk about the salvaged concept, instead. Would following this recommendation be compatible with moral realism? That depends on the nature of the defective commitment. If the defective commitment is negotiable, then the salvaged concept will still capture all of the non-negotiable commitments about morality, and so the thing that satisfies the salvaged concept will still count properly as morality. But if, on the other hand, the defective commitment is non-negotiable, then the salvaged concept will not capture all of the non-negotiable commitments

> about morality, and so to talk about the salvaged concept would no longer constitute talk about morality.
>
> (Lutz 2014: p. 364.)

What most interests me about this quotation is that Lutz can be seen as providing an answer to Lewis' rhetorical question – he is giving us a principled way of deciding what is and is not "close enough" to deserve the name of moral value. The method is this: look at the defective commitments and ask whether or not they are negotiable. If one or more of the defective commitments is non-negotiable, then the concept does not 'deserve the name'. (It is for this reason that we have to endorse UMET.) What we have is a candidate salvaged concept. On the other hand, if only negotiable commitments are unmet – i.e., if the concept satisfies all the non-negotiable commitments – then what we have is a genuine moral concept. It's just that some of our less important commitments about this concept were mistaken.

It is easy to see why this method of sorting reformed concepts into neat categories of salvaged and revised should appeal to many UMET proponents. For one thing, many of them will have arrived at UMET on the basis of an argument that follows the 'recipe' discussed in Chapter 1 – first identify a presupposition of morality and then argue that this presupposition is false. The presupposition will embody a non-negotiable commitment of morality; any reformed concept that fails to meet this non-negotiable commitment will not 'deserve the name' of moral concepts. However, we should have doubts that this method will be successful. We might have arrived at UMET on the basis of an argument (or collection of arguments) like AfO, which is distinctive in not identifying any particular commitments of morality as inherently non-negotiable. And even for error theorists who have come to UMET by a more traditional argument like AfR, we might think that the substitutionalist proposal itself, if successful in the sense of identifying reformed concepts that really do play the roles in our lives we would need them to, starts eroding our confidence in the 'presupposition' premise of said traditional argument. That is what I shall eventually go on to argue in this chapter. First, there is some more exposition to be done.

§4.1.2: Functional Continuity

One specific substitutionalist proposal comes from Kohler & Ridge (2013).[46] Their brand of substitutionalism is expressivist: "if the normative error theory is true, we should become expressivists" (2013: p428). As the reference to *normative* error theory rather than moral error theory makes clear, Kohler & Ridge's revolutionary expressivism is pitched as a solution to G-NWP rather than U-NWP. This presents them with something of a problem that does not afflict those offering revolutionary expressivism as response to U-NWP more narrowly. Whereas a UMET proponent who rejects GNET can offer the advice "if UMET is true, then we should become expressivists about moral terms" unproblematically, construing the 'should' in the sentence they utter as the uninfected 'should' of practical reason, Kohler & Ridge face a problem of self-defeat. After all, what is the significance of the word 'should' in the revolutionary expressivist advice when the problem is G-NWP rather than U-NWP? If GNET is true, then there is nothing that we 'should' do, and so the advice is self-defeating.[47]

Related to the problem of self-defeat is the problem of circularity. Kohler & Ridge would face this, they point out, if they solved the problem of self-defeat by stipulating that the 'should' in their advice concerning what to do if GNET is true is to be read in their revolutionary expressivist way. Then the problem of self-defeat is avoided, but only at the cost of making any case for their proposal circular. Kohler & Ridge get around both the problem of self-defeat and the problem of circularity in one go by appealing to the idea of functional continuity. The idea is to identify the functional role the error-infected concepts play in our lives, and look for reformed concepts that play the same functional role. This avoids the sort of "brute shift" in language that would be represented by advising that we should reorientate our normative vocabulary around concepts of what is, say, "green with yellow trimmings" (in Stevenson's famous example). This can then be

[46] Kohler & Ridge do not *call* themselves substitutionalists: the term is Lutz's. Nonetheless, they clearly are substitutionalists: normative terms are to have their meaning changed by substituting in a new, non-error-infected, meaning to replace the old meaning.

[47] Svoboda (2015) offers revolutionary expressivism as a solution to U-NWP and is keen to point out to his reader that in concerning himself with U-NWP he sidesteps the problem of self-defeat that might appear to afflict Kohler & Ridge's proposal. Another way that Kohler & Ridge could solve their problem of self-defeat is to pitch their revolutionary proposal as a solution to the narrower U-NWP, by saying that GNET *implies* UMET. I will not discuss this here, but I believe that this can be persuasively argued using paraconsistent logic.

used to answer the problems of circularity and self-defeat: because the reformed concept is functionally continuous with the error-infected concept, and because normative thought and discourse is indispensable ("not really optional for humans", is how Kohler & Ridge put it – p435), the rationale for revolutionism is only unobjectionably circular.

There are questions we might have at this point. Can Kohler & Ridge help themselves to the claim that normative thought and discourse is "not optional", or is there assumed normativity here to which they have no right as GNET proponents? Can they distinguish between objectionable and unobjectionable forms of circularity? I shall not pause to consider such questions. I am not concerned with G-NWP anyway. Instead I want to make the obvious and reasonable point that the appeal to functional continuity is both natural and helpful for any would-be substitutionalist. The more functionally continuous with the old moral concepts some candidate reformed concept **β** is, the more suitable it is going to be for the role the substitutionalist has in mind. Conversely, there will be no interest for anyone whatsoever in a reformed concept like "that which is green with yellow trimmings" which is entirely functionally discontinuous with current normative concepts. More generally: if we have two candidate reformed concepts $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, both of which are uninfected, but where $\boldsymbol{\beta}_2$ is more functionally continuous with the old moral concepts than $\boldsymbol{\beta}_1$, then, *ceteris paribus*, $\boldsymbol{\beta}_2$ will be the one to take as our reformed concept. This much, I take it, is uncontroversial.

Throughout the previous paragraph, I spoke in ways that assume that functional continuity is a graded phenomenon rather than a binary one. I spoke, for instance, of one candidate salvaged concept being *more* functionally continuous with old moral concepts than others. This seems natural. Functional continuity strikes me at least as a graded phenomenon rather than a binary one.

Indeed, an obvious method of assigning a numerical values to concepts' functional continuity suggests itself. Consider some moral concept **α**, thought to be defective because some pretheoretical commitment(s) or the other are unsatisfiable. Suppose we have *n* commitments: C_1, C_2, … C_n. Some of these commitments will be more important than others in terms of the contribution they make to allowing **α** to play the regulative role in our lives. So, let's give each of C_1, C_2, … C_n a functional importance coefficient

f_n. Each functional importance coefficient will be greater than 0 and less than or equal to 1. The more functional importance accorded to a particular commitment, the closer to 1 will that commitment's functional coefficient be.[48]

Armed with the complete list of commitments and their functional importance coefficients, we then compare with **α** some candidate salvaged concept **β**, using the formula:

$$F = (f_1C_1 + f_2C_2 + \ldots + f_nC_n)/(f_1 + f_2 + \ldots + f_n)$$

with each C being given a value of between 0 and 1 depending on how well the reformed concept under consideration fulfils the commitment. For the sake of simplicity, I shall suppose that when it comes to deciding this, it is a binary matter whether the commitment is satisfied or not – i.e., each C has a value of either 0 or 1. However, if we thought that this too was a graded matter, we could easily assign each C the appropriate value in the range [0,1]. Either way, functional continuity emerges as a continuous variable rather than a discrete one. Functional continuity is a thoroughly graded matter.

I want to emphasise, though, that the *f*-values are a measure *only* of the contribution the commitment to which they are attached makes to allowing the unrevised concept to be put to the uses to which it is in fact put. In general, I am going to follow a practice of assigning to any commitment thought by the substitutionalist to be non-negotiable an *f*-value value of 1. But this is not because the non-negotiability *per se* is directly relevant to a revised concept's F-value. The F-value is a measure only of how functionally continuous with the unrevised concept is the revised concept. The reason that putatively non-negotiable commitments will be assigned an *f*-value of 1 is that it seems plausible that a putatively non-negotiable commitment appears non-negotiable *precisely because* that commitment is especially important to the functioning of the unrevised concept. (In support of this point, recall Joyce's argument in favour of considering the commitment to

[48] Surely the idea of assigning weights in this manner must be allowed? Personally, I find the idea no less strange than the familiar idea of a graded notion of belief or intensity of preference. Any scepticism on this score, I hope, will vanish when a toy example is considered in a few paragraph's time.

NICRs non-negotiable: it was that the *whole point* of morality was to refer to such reasons, that without referring to such reasons we *miss out* the inescapable authority to which we appeal when morally condemning acts.) Thus, the F-value is first and foremost a measure of how functionally continuous it is with the unrevised concept, and not a measure of how "close" the revised concept comes to *deserving the name* of moral concepts. Let's call this latter dimension of evaluation *conceptual continuity*.

By conceptual continuity, all I mean is this: two concepts are conceptually continuous to the extent that they are in fact the same concept. Thus the concepts **SQUARE ROOT OF SIXTEEN** and **TWO SQUARED** are conceptually continuous: both are identical to the concept **FOUR**. Whilst functional continuity makes sense as a graded phenomenon, I think there is something to be said for treating conceptual continuity as binary and saying that two concepts α and β are either conceptually continuous or they are not, even though there are likely to be some close calls in which the matter is probably indeterminate (see fn. 51). However, the argument I press in the next part of the present chapter will not depend on taking on a view on this matter.

Now, two people can clearly possess the same concept, even where one of them has false beliefs about that concept that the other does not. We saw an example of this in §1.1.1.1 when considering the concept **MOTION**. Consider the concept **α** possessed by a 17th century mariner, which he labels "motion", and the concept β possessed by a 21st-century physicist, which she labels "motion". The mariner has some false beliefs about his concept that the physicist does not have: he thinks that his concept is of something that is absolute. In fact, his concept is of something that is relative. Nonetheless, the mariner and the physicist seem to possess the same concept. So here, α and β are conceptually continuous: they are the same concept. Both are identical to the concept **MOTION**. Meanwhile, the concepts **EPILEPSY** and **DEMONIC POSSESSION** are clearly separate concepts, i.e. conceptually discontinuous. This is so even if the class of events that the mariner will take to be correctly described as "fits of demonic possession" is exactly coextensive with the class of events that a contemporary person will take to be correctly described as "epileptic fits". We have already seen one proposed method of determining how conceptually continuous with an α-concept is any β-concept that it is

proposed could stand in for it. This is Lutz's method from §4.1.2, where **β** is conceptually continuous with **α** iff all of the non-negotiable commitments of **α** are met by **β**.

Let's now turn to the toy example promised earlier (in fn. 48) to illustrate the notion of a numerical measure of functional continuity. Suppose we have a total of four commitments concerning the concept **MORALLY WRONG**. Our commitments are:

1. If it would be morally wrong for A to φ in C, then A has an NICR not to φ in C. (f = 1.0)
2. If it would be morally wrong for A to φ in C, then A's φ-ing in C would be detrimental to the interests of at least one other person. (f = 0.8)
3. If it would be morally wrong for A to φ in C, then A has the ability not to φ in C (i.e., 'ought' implies 'can'). (f = 0.9)
4. If it would be morally wrong for A to φ in C, then it would not be extremely difficult for A to refrain from φ-ing in C, unless C are highly unusual circumstances (f = 0.2)

Suppose that the moral concept **MORALLY WRONG** is defective because the non-negotiable commitment to NICRs is defective. The substitutionalist identifies a concept **β** that fulfils the three remaining commitments. How functionally continuous with the error-infected moral concept is **β**? Applying the formula, we get: F = (1*0 + 0.8*1 + 0.9*1 + 0.2*1)/2.9 = 0.655 (to 2 d.p.). The conclusion seems to be that **β** will do 65.5% as good a job as the old concept **MORALLY WRONG** when pressed into service as a replacement.

Is 65.5% a pass mark? I don't propose to give an answer. As will become apparent, the argument that I give against substitutionalism will be unaffected by this detail. So here I am happy to let any substitutionalist fill in the details as they see fit. Let them list all the commitments – defective and otherwise – that they think are important; let them also specify the functional importance coefficients precisely as they see fit, and let them be the judge of whether the score that my formula assigns to their proposed salvaged concept is sufficiently high to count as pass-mark. My argument will target substitutionalist proposals *however* these details are filled in.

§4.1.3: Summary of The Substitutionalist Proposal

To sum up so far, substitutionalists propose reorienting moral thought and talk around salvaged concepts. Salvaged concepts would be conceptually discontinuous with defective moral concepts, in virtue of dropping any defective non-negotiable commitments that govern moral concepts, but they would be sufficiently functionally continuous with moral concepts to be put to similar uses. Functional continuity is a graded phenomenon. Lutz conceives what I am calling conceptual continuity as a binary phenomenon. I am happy to do the same (see fn. 51).

I shall now proceed to formulate what I see as the main problem for the substitutionalist.

The Problem of Shrinking Space

§4.2.0

In this part of the chapter, I am going to present my objection to substitutionalism. Before getting into the intricacies of the argument, I will try to summarise it as briefly and simply as possible. This should make what is to come easier to follow.

As we have seen, when a substitutionalist formulates their proposal, they will do so by offering some candidate salvaged concepts, around which we are to reorientate ethical thought and talk. For any such set of salvaged concepts, we will be interested in the question of how functionally continuous they are with the old error-infected concepts. In §4.1.2 I suggested a way of answering this question numerically. Substitutionalists now seem open to two risks. From one direction, the risk is that the salvaged concepts are insufficiently functionally continuous with error-infected concepts for the proposal to be practicable. In this case, substitutionalism will have to be rejected as a solution to U-NWP. From the other direction, the risk is that the salvaged concepts will be so functionally continuous with the error-infected concepts that UMET is rendered implausible. Here the idea is that what the salvaged concepts secure is the feasibility of something eminently recognisable as ethical discourse. But as I formulated it in §1.1.4, that is precisely what is distinctive about UMET: it is the theory that moral concepts are

so error infected that no error-free practice that is truly recognisable as ethical discourse is possible.

I call this problem for substitutionalism the problem of shrinking space. I shall now set about formulating this problem more precisely.

§4.2.1: Two Conceptions of Conceptual Continuity of Reformed Moral Concepts

In this section, I want to compare two conceptions of the conceptual continuity of reformed moral concepts with their unreformed (and defective) predecessors. We have already seen one conception of this, namely Lutz's, which I shall call the Substitutionalist Conception.

> **SUBSTITUTIONALIST CONCEPTION OF THE CONCEPTUAL CONTINUITY OF REFORMED MORAL CONCEPTS WITH THEIR UNREFORMED PREDECESSORS:** A reformed concept β is conceptually continuous with an unreformed moral concept α iff β satisfies all the non-negotiable commitments of α.

For short, I shall call this *Conception S*. It should be compared with:

> **HERMENEUTIC CONCEPTION OF THE CONCEPTUAL CONTINUITY OF REFORMED MORAL CONCEPTS WITH THEIR UNREFORMED PREDECESSORS:** A reformed moral concept β is conceptually continuous with an unreformed moral concept α to the extent that users of moral language can be charitably interpreted as using concept β on the assumption that they believe themselves to be using concept α.

For short, I shall call this *Conception H*.

Four things about these alternatives are worth emphasising at this point. First, these two possibilities do not exhaust logical space. We can easily imagine conceptions of the continuity of reformed moral concepts with their unreformed predecessors other than Conception S or Conception H. In particular, we could imagine a conception based on an externalist semantics of moral terms of the sort upon which Cornell Realists base their metaethical view. On a conception like this, a reformed moral concept would be conceptually continuous with an unreformed predecessor to the extent that the reformed

concept had its extension fixed in the same way.[49] However, a conception like this would be obviously irrelevant in the present context: it is agreed by all parties to discussion of U-NWP that the semantics of moral terms is not strongly externalist in this way. (If the semantics of moral terms are strongly externalist in this way, then it is a simple mistake even to enter into discussion of U-NWP.)

But an internalist semantics of moral terms does not necessarily mean that the meaning of our moral terms are pellucidly clear to us at all times or even most of the time. Opacity of introspection can interfere with our access to the meanings of our own terms.[50] So the second thing to emphasise is that on conception H, we need to start distinguishing between different conceptions of identical concepts. Consider some moral concept **M** (the concept **MORAL OBLIGATION**, say). Suppose that Sally believes that if A is morally obliged to φ then A has an NICR to φ, but suppose also that Joyce is right that there is no sense to be made of such reasons. Now suppose that a candidate reformed concept, **β**, descended from the defective concept **M**, presents itself. And finally suppose that someone who views Conception H as the correct conception of conceptual continuity finds that on Conception H, **β** is highly conceptually continuous with **M**. (We can interpret Sally as using **β** when she makes judgements about which agents are subject to which moral obligations.) Now we risk committing ourselves to the view that **β** *is* **M**, even though (ex hypothesi) **M** is defective and **β** is not. To avoid saddling the proponent of Conception H with contradicting commitments, the natural things to say are that **M** is non-defective after all, that Sally possesses the concept **M**, that the correct conception of **M** is **β**, but that Sally's own conception of **M** is **α**, and it is **α** that is defective. Sally thus possesses a non-defective concept, but is mistaken about her own concept, in the sense that she has an incorrect conception of her own concept.

Third, there is an fairly close connection between what I have just said (and indeed Conception H generally) and Finlay's criticisms of the presupposition premise of AfR, discussed in §1.4. Recall that Finlay argues that the mistake the UMET proponent makes is to assume that mistaken suppositions concerning NICRs infect the meanings of moral

[49] This sort of view is defended by Boyd (1988), Brink (1989), and Sturgeon (1988).
[50] This is well emphasised in Finlay (2008), which I discussed in §1.4.

terms, whereas in fact (according to Finlay), mistakenly buying into the notion of NICRs does not deprive statements made in terms of moral obligation of truth. Finlay's criticisms of UMET take on a new significance in the present context; I discuss this in §4.4.

The final thing to emphasis is that the UMET-proponent is not committed to Conception S just in virtue of their commitment to UMET. Indeed, it is not at all clear that AfR is not itself based on conception H. When Joyce claims that the *whole point* of moral claims is to invoke NICRs, this seems to be an interpretative claim concerning the speaker-meaning of moral terms.

§4.2.2: The Impact of Conception H on Substitutionalism

Suppose we think that Conception H is the best conception of the conceptual continuity of reformed moral concepts with their unreformed predecessors. Then conceptual continuity becomes a function of functional continuity. The function in question is *at least* one of proportionality and looks suspiciously like it might be one of identity.[51] Allow for now that on Conception H, conceptual continuity is identical to functional continuity. (I will consider in §4.3.4 whether this can be challenged.) In that case, if the substitutionalist is successful in showing that a particular concept **β** enjoys a high degree of functional continuity with the defective moral concepts hitherto used in moral contexts, then as a corollary the substitutionalist has shown that **β** also enjoys a high degree of conceptual continuity with the defective moral concepts. But then this means that standard users of moral language are interpretable as using concept **β** after all. Or rather, this means that standard users of moral language are interpretable as using a concept **M** the best conception of which is **β**. The upshot is that standard users of moral concepts are not using defective concepts after all. They are using non-defective concepts, but hold mistaken views about their own concepts: their conceptions of their concepts are off. Whilst this

[51] It does seem to me that there is much to be said for treating conceptual continuity as binary. But if conceptual continuity is a function of functional continuity, and functional continuity is graded rather than binary, how can conceptual continuity be binary? The natural view to take here is an analogue of the Lockean Thesis concerning the relationship between graded belief and binary belief: we should say that two concepts are identical when the surpass a certain threshold of functional continuity. This would make the best sense of close calls. It would also make the best sense of failures of transitivity, where agent A appears to possess concept α, B appears to possess β, and C appears to possess γ, and it seems reasonable that α and β could be different conceptions of the same concept, that β and γ could be different conceptions of the same concept, but unlikely that α and γ could be different conceptions of the same concept.

amounts to the diagnosis of widespread error on the part of moral believers, it does not amount to a diagnosis of *the* error on the part of moral believers that gives rise to error theory.

I call this the *problem of shrinking space*, because one way to think about it is in terms of the following squeeze on logical space open to the substitutionalist: consider some candidate reformed concept **β**. One of the first things an error-theorist will be interested to know about β, when assessing it as a possible salvaged concept, will be its F-value. Of course, β cannot have F=1, for then it would be *the exact same concept* as the unrevised and error-infected moral concept it is intended to replace. But it needs to have a value above a certain threshold, *t*, otherwise it will be rejected as unfit for the substitutionalist's purposes. But now recall one of the very first points I emphasised at the start of this chapter: salvaged concepts and revised concepts always come in pairs. So now consider the exact same reformed concept, this time as a candidate revised concept rather than as a candidate salvaged concept. Won't the prospective revisionary realist be as interested in the F-value of **β** as the prospective substitutionalist? After all, the would-be revisionist has a hermeneutical interest in the reforming definition: the idea is supposed to be that a large portion of the moral judgements we have hitherto made, all the while under some misapprehension about our conception of moral concepts, can be captured by the revised concept. If the F-value is too low, then the revised concept will be too discontinuous with actual use of moral concepts (by users of those concepts whose conceptions of the concepts are supposed to be mistaken) to be plausible as a revised concept – standard users will not be interpretable as using this concept, even with a mistaken conception of it. As the F-value increases, the plausibility of revisionism increases. Just as in the substitutionalist setting, though, the F-value of the reformed concept cannot be 1. In this case, the revised concept and the unrevised concept are one and the same, and so there is no need for revisionism. (Also, we could never have diagnosed the erroneous conception in the first place: there must have been some judgements in which the erroneous conception showed up by making a normative difference, e.g. by continuing to insist that the transgressor of a certain moral standard *really had a reason* to refrain from so doing

when all there is to say really is that the transgressor, by not subscribing to our moral standards, has become our *enemy*.)

So, just as the substitutionalist needs **β** to exhibit an F-value of more than *t* but less than 1, the revisionist needs **β** to exhibit an F-value of more than, say, *T* but less than one. So now we have two thresholds in the range [0,1]. One of these thresholds, *t*, represents the threshold at which the functional continuity of **β** with the defective concept becomes useful enough to make a good salvaged concept. The other, *T*, represents the point at which, on Conception H, **β** achieves such conceptual continuity with the defective concept that UMET no longer looks attractive. That is, the substitutionalist needs to identify a concept **β** with an F-value equal to *x* such that $t \leq x < T$.

Thus, the substitutionalist needs to worry about where in the range [0,1] *t* and *T* will be located. There is no particular reason for optimism that the range $t \leq x < T$ will be especially wide. After all, consider the set $\mathbf{S_\beta}$ of all circumstances in which **β** can be used for the same purposes as the defective concept. Then in all circumstances belonging to $\mathbf{S_\beta}$, standard users of moral discourse are to be most charitably interpreted as having all along been using a concept M the best conception of which is **β**. And once this is admitted, then the risk that *t* actually *equals* *T* looms sharply into view. This would be the end of substitutionalism: the logical space the substitutionalist can call home has shrunk to zero. Hence the name 'the problem of shrinking space'.

As suggested in §4.2.0, this can be put in the form of a dilemma for the substitutionalist: pick any analogous pair of salvaged and revised concepts you care to identify. Ask yourself: are these reformed concepts sufficiently functionally continuous with the putatively error-infected moral concepts that they can play much the same normative role in our lives as a social species as those concepts? If the answer is no, then this reformed concept is going to be neither plausible as a revised concept nor practicable as a salvaged concept. But if the answer is yes, then how much credence can we invest in UMET? For if the reformed concept is practicable as a salvaged concept, then it is also plausible as a revised concept. There needs to be some rationale for using the reformed concept as a salvaged concept rather than as a revised concept. The only rationale that has been offered is that the reformed concept fails to secure one or more non-negotiable platitudes

governing the use of moral concepts. But the fact that β achieves a sufficiently high F-value to be used for the substitutionalist's purposes itself casts doubt on the identification of the defective commitment as non-negotiable – at least on Conception H.

How big a problem is this dilemma for the substitutionalist? There seem to me to be two main ways they could attempt to neutralise the threat. The first is to argue that conception S should be preferred to conception H. The second is to emphasise the functional shortcomings of the salvaged concept. In parts four and five of the present chapter, I will assess the prospects of each of these strategies. I will suggest that neither are likely to succeed.

Before that, it would help to illustrate the problem of shrinking space with a fully worked out example. For whilst I think that the statement of the problem of shrinking space that I have given in this part of the chapter is fairly clear, it is nonetheless pitched at a fairly general and abstract level. What I shall do in the next part of the present chapter, then, is to dramatize the problem by considering, at a more practical level, how someone who favours substitutionalism as a solution to U-NWP will need to put the theory into practice.

The Problem of Shrinking Space – A Dramatization

§4.3.0: One Reforming Definition and Four Characters

For the purposes of my dramatization, I will need four characters and one reforming definition. In order to keep the dramatization general, I will not specify a particular reforming definition, and simply refer to "reforming definition R", which picks out an undefined reformed concept **β** that is supposed to be reasonably functionally continuous with the error-infected **MORALLY WRONG**. For the purposes of imagining the dialectic situation that obtains between my four characters more clearly, I suggest the reader simply selects their favourite of the three examples I provided in §4.1.0 (or their own example, if they have one they prefer), and suppose **β** to be the corresponding reformed concept.

The first character, I shall call Subeara (*Sub*eara the *sub*stitutionalist – named for easy identification). Subeara is a UMET-proponent. Because the concept picked out by definition R exhibits a high degree of functional continuity with the error-infected moral concepts with which she wishes to dispense, Subeara proposes R as a revolutionary definition, and the concept it picks out as a salvaged concept.

The second character, I shall call Fumihiro.[52] Fumihiro is a standard user of moral terms; he is innocent of the supposed mistakes in morality. Although he is in the habit of carefully reflecting on his moral beliefs, and attempting to correct any mistakes in his first-order normative commitments, he has never engaged in distinctly metaethical reflection. He straightforwardly believes that there are such things as NICRs, irreducibly normative favouring relations, truths about foundational propositions about morality, and the like. He believes this because this is the natural assumption that his enculturation has left him with.

The third character, I shall call Peter. Peter is a revisionist: for him R picks out a reformed concept. Peter thus rejects UMET. To return to Lewis' question of whether the concept picked out by R "deserves the name" "morally wrong", Peter answers that it does. (For ease of identification, this character is named after Railton, an actual moral revisionist who rejects UMET.)

Finally, I will eventually need to bring a character I shall call Katrina into play. Like Fumihiro, Katrina straightforwardly believes in NICRs, irreducibly normative favouring relations, truths about foundational propositions about morality, and the like. Unlike Fumihiro, though, Katrina *has* engaged in much metaethical reflection. It is just that her belief in the inescapable authority of moral reasons has survived this reflection. Her metaethical opinion is essentially that Christine Korsgaard is right about everything.[53] (Thus: *Ka*trina the neo-*Ka*ntian.)

[52] Fumihiro – possibly meaning *wide sentence* in Japanese. The idea here is that Fumihiro's use of moral terms captures the full range of the things we pretheoretically take ourselves to mean when we use them.
[53] See Korsgaard (1996, 2009).

§4.3.1: Substitutionalism in Practice

We are to suppose that Subeara says something like "because the concept **MORALLY WRONG** is error-infected, I shall henceforth use the term 'morally wrong' to pick out, not the concept **MORALLY WRONG** but the reformed concept **β**. Because **β** is sufficiently functionally continuous with **MORALLY WRONG**, it can be put to many of the same uses, and so by adopting this policy I gain the advantage of being able to function properly in society, which my realisation that UMET is true would otherwise jeopardise." She then goes out into the world and, for the most part, continues as does everyone else who uses the error-infected concepts. She reads in the newspaper that the government is planning to cut its foreign aid budget. She wants the aid budget to remain at its current level because he thinks that the objective interests of citizens of the affected countries will be harmed, and she is motivated to act in ways that she believes (or hopes) will increase the chances that government will rethink their policy. She writes to her MP, and in her letter she describes the plans as "a dereliction of our moral duty to citizens of those countries whose aid it is proposed to reduce". If when later pressed on how something can be "a dereliction of moral duty" given that moral concepts are radically error-infected, she will not retract her claim. She will say that cutting the aid budget *is* a dereliction of moral duty, in the sense of "dereliction of moral duty" that *she* means it.

Now Subeara, as a UMET-proponent, is of course a cognitivist. She thinks that that the utterance "the proposed reduction of the aid budget is a dereliction of our moral duty" is an assertion. Nonetheless, she must also be alive to the fact that utterances like this are used to *do* things. (If not, then why is she so keen on substitutionalism? Substitutionalism is motivated entirely by the observation that the ability to make moral assertions is *useful*.) Now suppose that Fumihiro writes a similar letter to his MP in which he also calls the proposed plans "a dereliction of our moral duty". Is there a difference between Fumihiro and Subeara concerning what they aim to *do*? Perhaps it cannot be assumed that their aims are completely identical: Fumihiro might well have epistemic aims that Subeara does not have. Nonetheless, the similarities between what they aim to do are surely great: both wish to *articulate their disapproval* of the proposed reduction in aid to their MP; both aim, however remotely, to *effect a change in public policy* by utilising the tools of

representative democracy; if they both know that their MPs approve of the plans, then both aim, however unlikely they realise success in the endeavour to be, *to persuade the MP* that the proposal should not be adopted. Even if Fumihiro has epistemic aims that Subeara does not have – to identify the moral truth with respect to the proposed plan – it still seems fair to say that the similarities in their aims in writing the sentence "this plan is a dereliction of our moral duty to the citizens of those countries whose aid it is proposed to cut" far outweigh the differences.

On conception H of conceptual continuity, we are thus well on the way to attributing the *same* concept **DERELICTION OF MORAL DUTY** to both Fumihiro and Subeara. But on the substitutionalist story, there was supposed to be an identifiable difference between the two. Is Fumihiro's possessing epistemic aims that Subeara does not enough to ground this difference? Not on conception H, it would appear. The substitutionalist will have to insist on conception S to make this stick. I'll discuss the prospects for that strategy in the fourth part of the present chapter.

Now Lutz would respond that there *will* be identifiable differences between Subeara and Fumihiro – differences, that is, in their aims in uttering or writing moral sentences – it's just that they will only show up in certain contexts. For suppose that in conversation with Subeara, Fumihiro made some moral claim that could not be supported by arguments using only concept **β**, but instead required the full-force concept **MORALLY WRONG**. In that case, Subeara would have to point the error out to Fumihiro; he would have to show Fumihiro that there is nothing in the world answering to the concept **MORALLY WRONG**, and that all their previous apparent agreement on other moral topics was an artifact of the functional continuity of **β** with the error-infected concepts with which Fumihiro is working:

> Of course, since the salvaged concept is not identical to the normal concept, there will be conversations in which the Substitutionalist's interlocutors will be interested primarily in aspects of the moral concept that the Substitutionalist, as an error theorist, will reject. For instance, in a philosophical context [Fumihiro] might well be using moral language in order to talk specifically about the categorical nature of moral claims. In such a context, [Subeara] could not felicitously use his reformed [i.e. revolutionary] moral language to communicate with [Fumihiro]. But why would he? The reason he cannot use his reformed language with

> [Fumihiro] here is because he disagrees with [Fumihiro] on the very point at issue! Rather than using a reformed moral language to demonstrate the extent to which his commitments align with [Fumihiro's], [Subeara] should be trying to convince [Fumihiro] to become an error theorist.
>
> Lutz 2014: p367

How convincing this response is will occupy us later. For now, I am interested in how the problem of shrinking space shows up in ethically engaged contexts, not whether it can be solved. Here, Lutz explains a feature of substitutionalism that below I call 'abandonment', and in this phenomenon the problem of shrinking space has begun to show. Lutz's point is that the salvaged concept is sufficiently functionally continuous with the moral concept it replaces to be useful in various concepts (i.e., has functional continuity exceeding threshold *t*) but is conceptually discontinuous with the moral concept in such a way that its conceptual discontinuity will show up in various contexts. If conceptual continuity is simply a function of functional continuity (as it is if Conception H is the correct one), then its functional continuity must also not exceed threshold *T*. That is, Lutz here seems to be trying to demonstrate that there is room in the interval [*t*, T] for a defensible **β** to carve out a respectable living, so to speak. I'll discuss the prospects for this strategy in the fifth part of the present chapter.

So far in this part of the chapter, I have compared an instance of Fumihiro's utterance of a moral sentence with an instance of Subeara's utterance of a moral sentence. To further illustrate the occurrence of the problem of shrinking space for the substitutionalist, I think it will be helpful to compare a co-operative exchange of moral views between Fumihiro and Subeara with a similar exchange between Fumihiro and Peter. Recall that like Fumihiro, Peter thinks that some things really are morally wrong and morally right. Unlike Fumihiro, he does not think that A's being under a moral obligation to φ necessarily means that there is a NICR for A to φ.

In both the conversation between Fumihiro and Peter and the conversation between Fumihiro and Subeara, we are to assume that both parties to the conversation are careful and cooperative. They are not out to score dialectical victory in order to "win" the "argument"; their conversation is a joint venture aimed at forming the most tenable opinion concerning the matter under discussion. Both parties in each conversation always

do their best to imagine the situation from the points of view of all those affected, to be impartial, to strive to support policies which treat everybody fairly, and to take steps to combat any unconscious implicit bias they might have against certain parties.

Suppose the topic is capital punishment. Is all capital punishment morally wrong, or might there be circumstances in which it is permissible? Imagine two neighbouring possible worlds, one in which it is Fumihiro and Subeara who are jointly trying to identify the best view to take of this matter, and one in which it is Fumihiro and Peter. Let's say that at the end of the conversation, Fumihiro and his partner in ethical exploration in each possible world reach the following highly nuanced conclusion. Capital punishment is morally permissible in response to certain extremely serious crimes, but the overwhelming majority of executions actually performed in the USA are gravely morally wrong. Furthermore, because the circumstances in which capital punishment is permissible are so rare, and because even in these circumstances it is not obligatory, and also because it is too easy to be led on the basis of unreliable reactive attitudes to mistake circumstances in which capital punishment would be gravely wrong for circumstances in which it is permissible, the only morally acceptable law is one that outright prohibits capital punishment in all circumstances. Importantly, let's also stipulate that in each possible world, the conversation he has with Subeara or Peter is identical. Subeara's contributions to the conversation are word-for-word identical to Peter's.

We can now notice two phenomena that are likely to be pervasive in the lives of substitutionalists as they go about their daily life, uttering moral sentences in conversations with morally serious individuals who are not error theorists. These are Divergence and Abandonment.

Divergence

At the end of their conversation in the possible world in which Fumihiro discusses the topic of capital punishment with Subeara, Fumihiro could be forgiven for thinking that they precisely agree. After all, their verdicts at the end were identical: the only morally acceptable law is one prohibiting capital punishment, for a variety of not-so-simple reasons including that the circumstances under which capital punishment is permissible are rare but not non-existent.

But now consider that Subeara appears perfectly able to say to himself: "Fumihiro and I both agreed that laws permitting capital punishment are morally wrong – but he was wrong and I was right". This is obviously odd-sounding. There is even something reminiscent of the oddness of "p, but I don't believe that p" in this – although, there is not quite the same appearance of flat-out contradiction. Still, that Subeara is able to say "Fumihiro and I agreed that p, but I was right and he was wrong" is information that if available to Fumihiro seems likely to alter his perception of the situation. Perhaps Fumihiro could even be forgiven for feeling somewhat aggrieved – isn't there something *arrogant* about Subeara's attitude? It appears that Fumihiro and Subeara disagree about something that was never touched upon in their conversation: the existence of irreducibly normative favouring relations, or NICRs, or some such.[54] But, given that this topic had no normative relevance to their discussion, why should the fact that Fumihiro holds some *extra* mistaken view mean that he is wrong when he says, with Subeara, that a law making provisions for capital punishment would be morally wrong? If the conjunction "a law making provision for capital punishment would be wrong, and there are no NICRs" is true, it follows that it is true that a law making provision for capital punishment would be wrong. Fumihiro believes this. Further, he believes the conjunction "a law making provision for capital punishment would be wrong and there are NICRs". This conjunction is false. But why should the falsity of the second conjunct in Fumihiro's view suddenly rob the first conjunct of veracity? This is what Subeara seems committed to when he thinks "Fumihiro and I both agreed that laws permitting capital punishment are morally wrong – but he was wrong and I was right".

Of course, the substitutionalist will say that there is a simple answer to this question, namely that the first conjunct in Subeara's conjunction is *not* the same as the first conjunct in Fumihiro's. That's because Subeara and Fumihiro are using the term "morally wrong" differently. Subeara's use of the term is tied to the salvaged concept; Fumihiro's is tied to the defective concept. Once again, in order for this distinction to be real, the

[54] We have already stipulated that this topic was not broached in the discussion of capital punishment. After all, if it had been broached, then Subeara and Peter would have had to respond in different ways. But the conversations were word-for-word identical.

substitutionalist will need either to cast doubt on conception H, or cast doubt on my claim that it looks likely that the threshold of functional continuity *t* at which reformed concepts become serviceable as salvaged concepts is going to be equal to the threshold T at which reformed concepts are best seen as revised concepts.

Abandonment

This is the tendency of substitutionalism to break down wherever the functional continuity of the salvaged concept is insufficient for smooth communication in an ethically engaged context, as explained by Lutz in the most recent quotation.

This phenomenon is going to occur in situations in which the substitutionalist converses with moralists who believe in such things as NICRs and identifiable truths about foundational moral propositions not because they haven't thought much about them, but because they have thought about them, and think that there are such things and that this metaethical view has important normative implications.

Suppose now that Subeara discusses the ethics of capital punishment with Katrina. (Recall that Katrina is a neo-Kantian whose metaethical view is highly influenced by the work of Korsgaard.) Katrina starts with the following first-order moral view, which is not a million miles away from the view that Subeara and Fumihiro came to on the basis of their discussion of the issues: capital punishment is morally wrong. It is morally wrong because it offends against the fundamental dignity of human beings. This is the primary reason that it is wrong. There are also secondary reasons why it is wrong, such as that it causes additional emotional trauma to relatives of those sentenced to death, that there is always the substantial risk of executing the wrongly convicted, and that once capital punishment is introduced for the most heinous crimes, the risk is that policy will creep into allowing executions for lesser crimes. But even in cases where none of these factors are present – there is certainty of guilt, the crime is of the worst kind imaginable, nobody's interests will be harmed by the execution, even the offender's – still the mere fact that the execution would offend against the fundamental dignity of human beings counts sufficiently heavily against it that it would be morally wrong to go ahead with it. Katrina here seems to be in the grip of a belief in an inescapably authoritative categorical reason.

Subeara, qua revolutionist, needs now to say to Katrina: your mistake is to think that there is something answering to the concept **MORALLY WRONG**, to which we can refer with the term "morally wrong". In fact, the concept **MORALLY WRONG** is inherently confused; in consequence *nothing* is morally wrong, and we shouldn't be using moral terms to pick out unrevised moral concepts at all.

Katrina, at this point, is likely going to be puzzled. After all, Subeara has until this point apparently been quite happy to use moral terms! To remove her confusion, Subeara will now explain: all this time I have been using the term "morally wrong" to apply to a salvaged concept. How satisfied is Katrina likely to be with this response?

I think she will prefer Peter's analogous move at this point. Peter will say to Katrina: your mistake is that you are confused about what we mean when we say that something is morally wrong. You are thinking in terms of inescapably authoritative categorical reasons, and I admit that pretheoretically this seems entirely natural. But on closer inspection, the only analysis we can give of something's being morally wrong is that it falls under the extension of the concept **β**. And the hypothetical death sentence we are considering here does not fall under that concept.

Isn't this an advantage for Peter's way of going about things compared to Subeara's *relative to the Subeara's own criteria for assessing U-NWP policy proposals*?! For one thing, Peter stays in "the moral game", so to speak, after the point at which Subeara has to abandon it. It seems to me that Peter and Katrina are more likely to understand one another than Katrina and Subeara are. More generally, Peter manages to converse with Katrina about important matters of ethics with much less friction. That was always supposed to be one of the main pay-offs of substitutionalism in the first place.[55]

[55] Here, substitutionalism appears to be in danger of collapsing into conservationism: by the substitutionalists own lights, we should become revisionist moral realists.

§4.4

My argument so far has been that the problem of shrinking space puts the substitutionalist at great risk of either no longer being able to defend UMET or having to admit that the policy proposal is not practicable. With respect to the first risk, conception H has been doing a lot of work. The substitutionalist, of course, is likely to reject conception H in favour of conception S. Most arguments for UMET, after all, follow the "recipe" discussed in Chapter 1, the first ingredient of which is a presupposition premise; most standard arguments for UMET thus *depend on* conception S, because they identify a non-negotiable commitment of morality.

But the substitutionalist is in a precarious position here. Recall the argument discussed in §1.4. We saw that Finlay thinks that the UMET-proponent who diagnoses a commitment to NICRs as a non-negotiable commitment governing the use of the term "moral obligation" is guilty of interpreting standard users of moral discourse in an objectionably uncharitable way. This was based largely on Finlay's claim that revisionists *make broadly the same judgements as everybody else*.[56] I argued in §1.4 that, when Finlay's argument is used to target Joyce, the best conclusion to draw was the one I labelled *Stalemate*. However, when Finlay's arguments are redeployed in the current context to target the substitutionalist more specifically, nothing like *Stalemate* seems to be the correct response. Instead the substitutionalist seems quite clearly guilty of the charge of uncharitable interpretation that Finlay levels against the UMET-proponent more generally. Hence, I wish here to argue for the conclusion:

> *Application of Finlay*: Finlay's arguments against UMET, when targeted more narrowly against the substitutionalist, show that if the substitutionalist holds firm to conception S, they are guilty of interpreting standard users of moral discourse in an objectionably uncharitable way.

[56] Finlay's relational account of the meaning of **MORALLY WRONG** is revisionist in the sense of the term relevant to this chapter: Finlay's definition must count as a reforming definition, and Finlay rejects UMET. So the concept **MORALLY WRONG** picked out by Finlay's reforming definition is a revised concept, in my sense.

The argument for *Application of Finlay* is simple. Recall the essential point of dialectical tension between UMET and Finlay's revisionist view. The question was: on which side of the witch/phlogiston vs. water/motion divide does morality belong? To answer, we turned to the essential application conditions. Finlay supported *Relational Moral Application*, and this returns the answer that morality belongs with water and motion. Joyce supports *Absolute Moral Application*, and this returns the answer that morality belongs with witches and phlogiston. But the support for *Absolute Moral Application* was that only a commitment to NICRs allows morality to perform the role it is thought to play in our lives. And it is doubtful that there is any other good reason to favour *Absolute Moral Application* over *Relational Moral Application*. Or, to put it another way: morality ends up in the same category as witches and phlogiston because, and only because, a practice guided by *Relational Moral Application* is not sufficiently functionally continuous with the existing practice of moral discourse to vindicate that practice.

But the substitutionalist, if they wish to use the orthodox response, treads a fine line here – a line so fine, in fact, that it disappears completely. For, on the one hand, the substitutionalist is an error theorist. That is, they think that a practice guided by *Relational Moral Application* is not sufficiently functionally continuous with existing moral practice to vindicate it. On the other hand, for the substitutionalist proposal to work, *Relational Moral Application* *does* need to be sufficiently functionally continuous with existing moral practice to be put to the same (or similar) uses.

To me this seems decisive against the tenability of the substitutionalist's holding on to conception S. Consider the situation in which this puts Subeara in the dramatization of the problem of shrinking space in the third part of the present chapter. She has committed herself to all the following:

1. The only **morally-acceptable-$_{R}$** law is one that prohibits capital punishment.
2. Fumihiro believes that the only **morally-acceptable-$_{A}$** law is one that prohibits capital punishment.
3. A practice governed by *Relational Moral Application* is insufficiently functionally continuous with the practice on the basis of which Fumihiro forms

his moral opinions for Fumihiro's practice to end up on the water/motion side of the divide; instead it ends up on the witch/phlogiston side.

4. A practice governed by *Relational Moral Application* is sufficiently functionally continuous with the practice on the basis of which Fumihiro forms his moral opinions for the relational practice to be put to most of the same uses as Fumihiro's practice.[57]

But now recall that in §1.4, the question of interpretation arose at the choice point between the relational and absolute account of the essential application conditions of moral terms. There needs to be a reason for interpreting speakers to be applying absolute rather than relational essential application conditions. The evidence to ground such an interpretation is going to come from noticing other things that they say and trying to make sense of their commitments on the assumption that they are rational. But Subeara's opposing commitments (3) and (4) in the list above pull in opposing directions here. (3) together with a principle of charity should lead Subeara to reject (4). (4) together with a principle of charity should lead Subeara to reject (3). So, in this case, the accusation that Subeara is objectionably uncharitable in his interpretation of Fumihiro's moral utterances can be made to stick. The substitutionalist cannot hold firm to conception S.

Can The Substitutionalist Find Room in The Interval [*t*, T]?

§4.5

Substitutionalism's compatibility with its own error-theoretical motivations would now see to depend entirely on the prospects of the substitutionalist finding room in the interval [*t*, T] for the salvaged concepts to sit. The prospects for this strategy do not seem promising to me.

[57] Here, "**morally-acceptable-$_R$**" means "morally acceptable, where the concept 'morally acceptable' is governed by *Relational Moral Application*"; "**morally-acceptable-$_A$**" means "morally acceptable, where the concept 'morally acceptable' is governed by *Absolute Moral Application*"

We need some way of distinguishing between UMET and LMETs. I suggested one in §1.1.4. UMET is reached when the uninfected nodes of the ethical web are insufficient to ground the social practice we recognise as the practice of ethical discourse. But it is obvious that the salvaged concepts that the substitutionalist aims to identify are intended to be the basis not of an entirely new social practice that it not recognisable from the current perspective as ethical discourse, but instead of a new social practice that is broadly continuous with the old. Indeed, a large part of the motivation of Lutz's proposal was that UMET is a minority position and that the UMET-proponent will want to find a way of communicating effectively with those whose participation in social discourse puts them squarely in the camp of the old practice, not the new.

Essentially, if the substitutionalist attempts to escape the problem of shrinking space by arguing that there is room in the interval [*t*, T] for their salvaged concepts to carve out a niche, they seem to be offering a proposal that fits squarely into the category of those set aside in the first part of Chapter 3. On this understanding of substitutionalism, the proposal is essentially this: once we free ourselves of our mistaken conceptions of our moral concepts, and understand what they really amount to, we are likely to use them slightly differently, and the consequences of these differences are likely to be positive. Thus, the proposal fits squarely into the same category as Blackford's. Like Blackford's proposal, substitutionalism on this understanding is not a policy option with respect to U-NWP. It is an instance of LMET with a normative agenda. As an instance of LMET with a normative agenda, it may well be defensible. But that does not mean that any would-be substitutionalist's view of themselves as a UMET-proponent who is inventing a whole new style of social discourse to rescue the projects served by an old, unsustainable, radically error-infected style of social discourse is accurate. From this perspective, the character of Subeara I invented to dramatize the problem of shrinking space almost looks a little conceited: it's as though she thinks that just because she is aware of an error infecting much of current social discourse, she can magically set herself above the whole practice, whilst still reaping the benefits of engaging in it. She can't.

Summary and Preview

§4.6

The Substitutionalist ends up in roughly the same position as Marks in the previous chapter: the proposal is incompatible with its own error-theoretical motivations. I think that the fact that this result has occurred twice should make us suspicious about the whole enterprise of developing policy proposals with respect to U-NWP. The unifying theme of the present chapter and the last seems to be this: *something broadly recognisable as ethical discourse seems impossible to get away from*. Maybe this fact alone should cast doubt on UMET. In the next chapter, I assess the prospects of objections to UMET based on this thought. Objections of this sort form a family of objections I shall call objections from loss. I think that there is a version of the objection from loss that is actually persuasive.

Chapter Five:
The Objection From Loss

§5.0

The current state of play: our multi-stranded argument for UMET, the argument from obscurity, is yet to be answered. No particular fault has been found with either of the two main strands that form the focus of this thesis – no particular grounds for optimism about the possibility for rational consensus concerning foundational moral claims has been identified, and inescapable authority has not been accommodated. Nonetheless, the previous discussion has suggested that UMET cannot possibly be correct. It has done this by illustrating the practical indispensability of ethics.

The practical indispensability of ethics has been most obviously demonstrated by the problems abolitionism has with self-defeat. Recall that the basic problem here turned out to be that the only interesting or compelling way to make a case for abolitionism would be to argue that abolitionism recommended itself when assessed from the socially rational point of view. That is, only an ethical case for abolitionism could be compelling. But the whole point of abolitionism is that it proposes doing away with ethics and making do with something less – desire, in Marks' case, and something it is more difficult to identify in other abolitionists' cases, probably because what they wish to make do with just is a form of ethics.

The practical indispensability of ethics and its relevance to an assessment of the chances that UMET might be true has also been illustrated by the discussion of substitutionalism. Recall that the success of substitutionalism depends on a salvaged concept being identified with sufficient functional continuity that the salvaged concept *could do some of the work that current, putatively error-infected concepts do*. It was precisely this that sunk substitutionalism: I argued that once the relevant threshold of functional continuity is passed, there is every reason to suppose that UMET is false. The salvaged concepts just are well-functioning, uninfected ethical concepts. We can see this as just another

illustration of the practical indispensability of ethics: it is because (most) human beings are incapable of abandoning the ethical point of view that the need to identify new concepts with enough functional continuity with the old that we can carry on much as before arises.

This leads us to the objection to UMET that gives this thesis its title: the objection from loss. It is the natural objection to which these considerations point. I provided a simple version of this objection in §0. It went:

> **Objection from Loss – Simplest Version**
> [1] If UMET is true, then there is some sense in which it is illegitimate to adopt an ethical point of view.
> [2] It is not the case that there is some sense in which it is illegitimate to adopt an ethical point of view.
> [3] Therefore: UMET is not true.

I warned at the time that this simple version concealed complexities. These complexities have now come to light. Premise 2 looks exceedingly strong: it does not say just that there is *a* sense in which it *is* legitimate to take an ethical point of view. It says that there is *no* sense in which it is *not* legitimate to do so. But of course much with which we are acquainted in ethical discourse might well turn out to be illegitimate – appeals to inescapable authority and assumptions that there exist truths concerning foundational moral propositions being the two main sources of illegitimacy with which this thesis has been concerned. To be acceptable, premise 2 must be understood as referring to something quite minimal when it mentions adopting an ethical point of view. We adopt an ethical point of view when we make *any* discourse-implicating ethical judgement. Premise 2 says that there is nothing illegitimate about doing this. But it does not say that when we do this, we do not risk involving ourselves in error. If we have been socialised into a culture that makes heavy use of appeal to NICRs and the like, then perhaps we will fall into the trap of believing in inescapable authority when we make discourse-implicating ethical judgements. But premise 2 claims that these errors, if we make them at all, are only incidental: we don't commit these errors *merely* by making discourse-implicating ethical judgements. So there is *nothing* illegitimate about adopting an ethical point of view *per se*.

The objection from loss is not new. It has been made by others before me, and can be made in a variety of ways. It appears in Blackburn (1993: chapter 8) in the following guise. Blackburn opens 'Errors and the Phenomenology of Value' with the observation that Mackie, whilst professing to be a moral sceptic and calling his metaethical view an error theory, nonetheless went on in *Ethics* to offer much in the way of ethical discussion. (Blackburn describes Mackie as "happy to go on to express a large number of straightforward moral views" (p. 149) in the second half of the book. Of course, on the stipulated use of "moral" and "ethical" in force in this thesis, we would have to replace the word "moral" with "ethical" here.) This, Blackburn notes, is distinctly odd. Surely, the natural suggestion is that "[i]f a vocabulary embodies an error, then it would be better if it were replaced by one that avoids the error", or, "[s]lightly more accurately, if a vocabulary embodies an error *in some use*, it would be better if either it, or a replacement vocabulary, were used differently" (p. 150). But Mackie does not do this, Blackburn observes. Instead he simply goes on to talk in more or less the ordinary way about ethics. (Of course, I have already assessed at some length, in Chapter 4, the prospects of one proposal for using ethical vocabulary or a replacement for it differently, and I argued that to whatever extent that this proposal appears attractive, UMET must be unattractive to precisely that extent. Blackburn's comments in this regard seem to me quite friendly towards that analysis.)

Now at the start of Chapter 5 of *Ethics*, in a section entitled 'The Consequences of Moral Scepticism', Mackie asks: "if there is no objective moral truth to be discovered, is there nothing left to do but to describe our sense of justice?" (Mackie 1977: p. 106.) (By "describing our sense of justice", Mackie means using the Rawlsian method of reflective equilibrium to iron out inconsistencies in our commitments, making sure that we do not confuse this "legitimate kind of inquiry" (p. 105) with a search for real objective truth.) He answers his question in the negative: it is not the case that all that is left to do is to describe our sense of justice. Instead, he says, we can "look at the matter in another way. Morality is not to be discovered but to be made: we have to decide what moral views to adopt, what moral stands to take. […] [T]he object [of this exercise] is to decide what to do, what to support and what to condemn, what principles of conduct to accept and foster

as guiding or controlling our own choices…" (p. 106). It is to the suggestion that "we have to decide what moral views to adopt" that Blackburn is responding when we writes:

> [F]rom the standpoint of an error theory, it is quite extraordinary that we should have to do any such thing. Why should we have to choose to fall into error? Surely it would be better if we avoided *moral* (erroneous) views altogether and contented ourselves with some lesser, purged commitments that can be held without making metaphysical mistakes. […] The puzzle is why, in light of the error theory, Mackie did not at least indicate how [to do something like this]. […] And in my view this is enough of a puzzle to cast doubt back on to the original diagnosis of error. In other words, it would obviously have been a silly thing to do, to try to substitute some allegedly hygienic set of concepts for moral ones; but that in itself suggests that no error can be incorporated in mere use of those concepts.
>
> (Blackburn 1993: p. 150)

This is obviously is a version of the objection from loss. Boiled down to its simplest form, it seems to run:

> **Objection from Loss – Blackburn Version**
> [1] If UMET is true, then it would be better if we avoided *moral* (erroneous) views altogether.
> [2] It would not be better if we avoided moral views altogether (as evidenced by Mackie's declining to indicate how to do avoid them).
> [3] Therefore: UMET is not true.

Translating *Objection from Loss – Blackburn Version* into the stipulative terminology of this thesis, we get:

> **Objection from Loss – Amended Blackburn Version**
> [1] If UMET is true, then there is some sense in which it would be better if we avoided discourse-implicating ethical judgements.
> [2] There is no sense in which it would be better if we avoided discourse-implicating ethical judgements.
> [3] Therefore: UMET is not true.

And this is essentially the same argument as *Objection from Loss – Simplest Version*.

It seems to me that the basic thought articulated by this argument is ripe for development. In 'Errors and the Phenomenology of Value', Blackburn's development consists in trying to show that there is no difference in practice between, on the one hand, the sort of activity that we can imagine being founded by someone who takes seriously the need to replace the old error-infected concepts with hygienic ones, and on the other

hand, moral discourse as we already know it. In this way his discussion marks a contribution to the overall project he has advanced in several works – that of showing that his quasi-realism can account for and capture everything we recognise as morality. This provides two reasons to part company with his exposition of the objection from loss at this point. First, in considering the prospects of a project of replacement of the old concepts with hygienic ones, his concerns are closer to those of the previous chapter to this one. Second, Blackburn's project of accommodating everything we recognise as morality pulls in precisely the opposite direction to one component of the present project: that of accommodating the insights of the argument from obscurity, which will probably require *revisions* to the practice of moral discourse.

§5.1: Preview

The version of the objection from loss that I think best shows UMET to be wrong is as follows:

> **The Objection from Loss – Practical Indispensability Version**
> [1] If UMET is true then it is not the case that for most of us engagement with the ethical project is both possible and of overwhelming importance.
> [2] For most of us, engagement with the ethical project *is* both possible and of overwhelming importance.
> [3] Therefore: UMET is false.

I make this argument in sections §§5.4 - 5.9. But first I want to turn to an existing presentation of the objection from loss: that offered by Wiggins (2006: chapter 11). The reason for this is that I want to extract from Wiggins a distinction that will come in useful later. This is the distinction between what I will call "loosely deontic concepts" and "heavily deontic concepts". Formulating this distinction will be the work of §5.2. In §5.3, I will say a little about how the argument to be formulated in the present part of this chapter and the next is different from another well-known indispensability argument.

Thereafter, I will set about defending *The Objection from Loss – Practical Indispensability Version*. In §§5.4 - 5.8, I will offer an argument towards:

> **Indispensability Premise 1**: if UMET is true, then it is not the case that for most of us, engagement with the ethical project is both possible and of overwhelming importance.

In §5.9, I will offer an argument for the second premise, namely:

> **Indispensability Premise 2**: engagement with the ethical project *is* both possible and of overwhelming importance for most of us.

If the arguments for both premises are successful, then I will have shown UMET to be false. In §§5.10 - 5.12, I will tie up some loose ends: revisiting the comparison between my argument and the other well-known indispensability argument, responding to an anticipated objection that because I am concerned with ethical practice rather than ethical truth my argument does not touch the error theory, and then differences and similarities between the objection from loss and conservationist and fictionalist U-NWP solutions.

§5.2: Two Types of Deontic Concept

Wiggins is responding to Mackie's concerns about queerness when he suggests that we begin by looking "first where Mackie himself looked when he sought to compensate for that which his Chapter 1 seemed to sweep away" (Wiggins 2006: p. 326). Wiggins reminds us that "Mackie proposed a form of constructivism" (ibid) on which morality was to be 'a device for counteracting limited sympathies'. He writes:

> By what method then does the construction proceed? Mackie's real method is this. As he moves from topic to topic and question to question, he draws constantly upon the reservoir of implicit knowledge that we all have, but make explicit only piecemeal and in given contexts, of what matters in this or that sphere of activity. At need, Mackie draws freely upon this reservoir. From hence, of course, and from his power to interpret the point of what he finds, his rare judgement and good sense.
>
> Mackie might find this a disappointing description. But I should reply that he ought not to complain. For this is the *right* method.
>
> (p. 328.)

Wiggins' calling this method of construction the *right* method recalls Blackburn's point that what Mackie does after convicting morality of radical error infection looks a lot like what we nonetheless recognise as ethics. And the lesson Wiggins takes from this emphasises the practical indispensability of this practice:

> [S]omething else can be gained by reflection upon this reservoir which Mackie draws upon [...] The thing we gain is a vivid reminder of the possibility of inward ends or purposes or concerns which, in the business of their life at a given place and time, participants in a first-order ethic will

> steadfastly adhere to as if by second nature, distinguishing readily, however essentially contestably, between these concerns and others that *can* be abandoned and may have to be. Is it not here, in the sphere of the unforsakeable, that we find the true source of the deontological ideas and categorical requirements that Mackie insists upon our retaining in first-order morality?
>
> (p. 329.)

Compare this latest Wiggins quotation to *The Objection From Loss – Simplest Version*. What Wiggins offers here is a particularly elegant defence of a weaker version premise 2, to the effect that it *is* legitimate to adopt an ethical point of view. In fact, it is more than merely *legitimate*, it is positively necessary. There are certain ends, Wiggins tells us, which are of such overriding importance that they make a certain demand on us. What sort of demand? Not some mysterious demand of inescapable authority, but simply the sort of demand we can quite easily make sense of from the perspective of Joyce's own "non-Humean instrumentalist" conception of practical rationality: the demands of a concern that is of such overriding importance that it is, as Wiggins' says, "unforsakeable".

How best should we go about making sense of the "sphere of the unforsakeable"? I suggest that we should turn to Williams' brief discussion in the final chapter of *Ethics and the Limits of Philosophy* of the notion of importance and its complicated connections with the notion of deliberative priority. Williams distinguishes between two notions of importance: a relative notion and (what we could call, but Williams doesn't call) an absolute notion. To say that *x* is important in the first, relative, sense, is to say that someone *finds* that thing important. To say that *x* is important in the second, absolute sense, is to say that *x* is important, period. The distinction here is intimately connected to the distinction in Wiggins (1987: Essay 1) between the relative and absolute conceptions of a need. After all, what is it for an agent's need (in the absolute sense) for *x* to be especially bad but for it to be of overriding importance (in the absolute sense) that the agent gets *x*? An important part of Wiggins' discussion concerns the close connection between the absolute notion of a need and essentially contestable matters concerning, for example, what departures from wellbeing should be considered harms. Williams' discussion of importance in *Ethics and the Limits of Philosophy* is likewise connected with essential incontestability. He writes: "I doubt there can be an incontestable account

of this idea [the idea of absolute importance, that is]; the explanations people give of it are necessarily affected by what they find important" (Williams 1985: p. 202).

For Williams, three things are required of this absolute notion of importance, poor though our understanding of it might be: (i) that there is such a notion; (ii) that there is in general no implication from "*x* is important (in the relative sense)" to "*x* is important (in the absolute sense)"; (iii) that questions concerning what is important (in the absolute sense) are to be distinguished and kept apart from questions concerning *deliberative priority*. Nonetheless, there are important – but not simple – connections between questions of importance and questions of deliberative priority.

Deliberative priority is defined thus: "[a] consideration has high deliberative priority for us if we give it heavy weighting against other considerations in our deliberation" (Williams 1985: p203). Complications in the relationship of dependence between deliberative priority and absolute importance come from a number of sources. Something, *x*, might be absolutely important, but be such that nobody can do very much about *x*, or at least that *I* cannot do much about *x*. This can render *x*'s importance largely irrelevant to the contents of my own rational deliberation. Likewise, *x* might be none of my business; there exists, Williams says "a deliberative division of labour" (ibid).

The main thing that Williams extracts from his analysis of the notions of absolute importance and deliberative priority, and the connection between them, is a notion of obligation. This is a notion that we should be careful to call *ethical* obligation, and not *moral* obligation. Williams is, at this point in the text, in the process of arguing that "the morality system" is mistaken in that it attempts to reduce every ethical concern to a notion of obligation, to find general obligations to ground particular ones, and that it is committed to the idea that the only thing strong enough to defeat an obligation could be another obligation. So the notion of *moral* obligation is, at this point in William's text, looking quite suspicious. The notion of obligation at which we arrive by considering the relationship between the absolute notion of importance and the notion of deliberative priority is not supposed to be suspicious in this way. Indeed, this is supposed to be the notion of obligation we end up with when obligations are "rightly seen as merely one

ethical consideration among others" (1985: p. 202). In contrast to the notion of moral obligation, ethical obligation is

> …grounded in the basic issue of what people should be able to rely on. People must rely as far as possible on not being killed or used as a resource, and on having some space and objects and relations with other people they can count as their own. It also serves their interests if, to some extent at least, they can rely on not being lied to. One way in which these ends can be served, and perhaps the only way, is by some kind of ethical life; and certainly, if there is to be ethical life, these ends have to be served by it and within it. *One* way in which ethical life serves them is by encouraging certain motivations, and *one* form of this is to instil a disposition to give the relevant considerations a high deliberative priority – in the most serious of these matters, a virtually absolute priority, so that certain courses of action must come first, while others are ruled out from the beginning.
>
> (p. 206.)

Now compare all that is said in the Williams passage just quoted with Wiggins' claim that it is "in the sphere of the unforsakeable, that that we find the true source of the deontological ideas and categorical requirements that Mackie insists upon our retaining in first-order morality". I hope that two things are clear about the concept of obligation that Williams derives from the twin considerations that some things in life are of absolute importance and that there exist some non-simple connections between absolute importance and deliberative priority. The first thing is that the concept of obligation thus reached probably *just is* the most general conception available of "the deontological ideas and categorical requirements" to which Wiggins refers. The second thing to notice is that NICRs are not required here. That people need to be able to rely to some extent at least on not being lied to generates pressure within ethical codes transmitted in any society to be such as to give the demands of honesty high deliberative priority in many contexts of information exchange. It is no part of the concept of obligation that Williams has arrived at here that people's need to be able to rely on not being lied to provides *any agent in any context* with an NICR not to lie.

Because NICRs are irrelevant to the concept of obligation with which we are dealing here, I shall call the deontological concepts this concept of obligation can be used to ground (i.e. those having to do only with giving high or overriding deliberative priority

to some consideration or other) *loosely deontic concepts.* By contrast, *strongly deontic concepts* will be those to which NICRs are essential.

§5.3: Indispensability, Realism and UMET

Recall that the version of the objection from loss that I will defend in this chapter is:

> **The Objection from Loss – Practical Indispensability Version**
> [1] If UMET is true then it is not the case that for most of us engagement with the ethical project is both possible and of overwhelming importance.
> [2] For most of us, engagement with the ethical project *is* both possible and of overwhelming importance.
> [3] Therefore: UMET is false.

The overall idea suggested by the discussion in the previous section is this: use Wiggins' pointing to the "sphere of the unforsakeable" as the source of deontic concepts, as a way of supporting premise (2). The crucial idea here is that we *have* to be able to secure for ourselves some form of ethical life, and if we have to be able to secure for ourselves some form of ethical life, then ethics as a whole cannot be irredeemably error-infected.

Now at this point some readers might think: *hang on a moment, this sounds a bit like David Enoch's indispensability argument for metanormative robust realism* (Enoch 2011: Chapter 3). And given that it is common ground between the UMET-proponent and me that the strands of argumentation discussed in Chapter 1 can correctly be described as containing *insights*, and that my project is one of finding a way to accommodate these insights on a non-UMET view, this might seem to open me up to an objection right from the outset, which we could call "the objection from unwitting realism". This objection goes: if the argument to be developed in present part of this chapter succeeds, then I will have established a form of robust realism. In this case, I will be unable to accommodate the insights of Chapter 1, for these cannot be captured on a robustly realist metaethical view. If, on the other hand, the argument to be developed here does not succeed, then the threat of UMET has not been neutralised. In either case, the present project cannot succeed.

I have two lines of response to this potential objection, and I want to say something about both lines of response now because doing so will, I think, make the details of the

argument to come easier to understand. First, I'll give a brief exposition of Enoch's indispensability argument.

Enoch's argument from indispensability runs:

> [1] If something is instrumentally indispensable to an intrinsically indispensable project, then we are (epistemically) justified (for that very reason) in believing that that thing exists.
> [2] The deliberative project is intrinsically indispensable.
> [3] Irreducibly normative truths are instrumentally indispensable to the deliberative project.
> [4] Therefore: we are epistemically justified in believing that there are irreducibly normative truths.
>
> (Enoch 2011: p. 83)

This formulation invites the obvious questions: what precisely is an intrinsically indispensable project, and what is it for something to be instrumentally indispensable to some project? By his own admission (on p. 83), Enoch says precious little about this, but I don't think this is a major problem. In my opinion, what little he does say provides sufficient explanation. A project is intrinsically indispensable, he says, if it is "non-optional for us – it is one we are rationally required to engage in" (Enoch 2011: p. 60). With respect to the notion of instrumental indispensability, he gives us the following:

> Something is instrumentally indispensable for a project, I suggest, just in case it cannot be eliminated without undermining (or at least sufficiently diminishing) whatever reason we had to engage in that project in the first place; without, in other words, thereby defeating whatever reason we had to find that project attractive.
>
> (Enoch 2011: p. 69.)

Now I don't propose at this juncture to submit all the premises of Enoch's argument to scrutiny with a view to finding the case for one or more of them wanting. More important is getting clear on why Enoch thinks that this argument, when conjoined with another argument that by this point in his book he has already made, shows UMET to be false. This itself comes in two stages.

First, he thinks that, if the success of the indispensability argument in demonstrating meta*normative* robust realism to be correct is granted, then any meta*ethical* view other than robust realism will look very strange and entirely undermotivated. This includes the

conjunction of robust metanormative robust realism and metaethical error theory. For, Enoch asks, what sort of special objection could there be to irreducible moral normative truth, that sits easily with the acceptance of irreducible normative truth more generally? He runs through a quick survey of the usual suspicions. Could queerness be the culprit? No: "if you think that things like the purported fact that we ought not humiliate others is metaphysically queer, you should be equally metaphysically suspicious of things like the purported fact that two-boxing is the uniquely rational policy in Newcomb's Problem" (p. 89). What about concerns relating to epistemological access? No: "the epistemological challenge […] seems to apply just as forcefully to other, non-moral normative truths and facts" (ibid). And so on.

Of course, in light of AfR, the UMET-proponent does have some pretty beefy grounds of suspicion that are relevantly restricted to the moral sphere. These concern inescapable authority. So Enoch needs a further reply to UMET-proponents whose scepticism about morality is based on suspicion of NICRs. This constitutes the second stage of Enoch's rationale for taking the indispensability argument to target UMET. As we would expect from a robust realist whose project is one of vindicating the whole of morality – vindicating morality in all its *majesty*, we might say – the response at this stage is not to reject the claim that NICRs are necessary for morality. Instead he appeals to his argument concerning the normative significance of objectivity to suggest that we should believe that there are NICRs. And once you believe in NICRs (on the basis of this other argument), and are committed to robust metanormative realism (on the basis of the indispensability argument), then the only way to reject metaethical robust realism would be to adopt a hybrid view on which there are irreducibly normative truths, and there are NICRs, but the normative truths to which NICRs give rise are not irreducibly normative. That would clearly be an idiosyncratic view, to say the least. It is hard to see what any such view would have going for it.

Thus, my first line of response to the objection from unwitting realism is this. Even if in pressing the objection to UMET I am about to press over the course of the next few sections I really did thereby commit myself to buying into some form of metanormative robust realism, this would not necessarily prevent me from making good on my aim to

accommodate the insights of the arguments for UMET in some non-UMET view. For Enoch cannot successfully target the UMET-proponent without leaning more heavily on the argument about the normative significance of objectivity than he does on the indispensability argument. But it is the conclusion of the argument about the normative significance of objectivity that is incompatible with the conclusion of AfR, not the conclusion of the indispensability argument. But if AfR is good enough to generate an error theory of NICRs, then it is good enough to defeat Enoch's argument about the normative significance of objectivity.

My second line of response goes further than this. It is that Enoch makes stronger indispensability claims than I do, and thus arrives at a view with much heavier ontological commitments than the view at which I arrive. For all that I will claim is indispensable is a collection of *loosely* deontic concepts, as that notion was elucidated in the previous section; I will be quite careful to avoid appealing to *strongly* deontic concepts. Enoch, on the other hand, is out to demonstrate the indispensability of irreducibly normative truths, which presumably count as strongly deontic. Thus, Enoch's argument and his resulting view will inevitably be vulnerable to objections that mine will easily escape. Indeed, one way to look at my argument is as delivering *in one sense* roughly the same goods as Enoch's arguments, but doing so at a lower cost in terms of ontological commitment. For the thing that Enoch and I have in common is that we think that UMET is less plausible than some other metaethical view because it is not really plausible that when we ask ourselves what we ethically should do in some difficult circumstance we are asking ourselves a pseudo-question. But Enoch thinks that in order for it to be the case that we are not asking ourselves a pseudo-question when ask ourselves an ethical question, we need some heavily deontic concepts. I do not think we need any such thing. I said in the antepenultimate sentence that what Enoch and I have in common is that we both think that UMET is less plausible than some other metaethical view. This is true of Enoch in the sense that he thinks of some specific metaethical view, namely robust realism, that it is more plausible than UMET. It is true of me in the sense that I think that since UMET is not very plausible, it is quite likely that there is *some* metaethical view that is more

plausible than UMET.[58] So far as indispensability concerns go, I am not out to prove any more than that we should reject UMET.

§5.4: A Disjunction — Intrinsic Indispensability or Mere Exigency

I now begin working towards my argument for *Indispensability Premise 1*, which I will present in §5.8. Throughout my discussion here, I shall refer to *the ethical project*. I shall be using this term in the following way:

> **The Ethical Project (definition):** the general practice of tending to think of the ethical point of view as being often of very high deliberative importance when making decisions whose consequences will affect others.

One claim I wish to make about the ethical project is this:

> **Psychological Inescapability (of the ethical project):** the vast majority of us are psychologically incapable of disengaging from the ethical project to any great extent.

Psychological Inescapability strikes me as very obviously true, and I shall not make strenuous efforts to defend it here. Perhaps it could receive support from a biological story about the innateness of morality (of the sort offered in Joyce 2006), the idea being that since the propensity to think in moral terms is innate, we are very unlikely to be able to do much to resist doing so. However, this argument seems to me to be less plausible than the bare assertion itself: mere *experience* will attest for most of us that we cannot disengage from the project.[59] So I will assume that the reader will grant me *Psychological*

[58] It also happens to be true of me that I think of some specific metaethical view that it is more plausible than UMET, but in my case the view in question is not robust realism but a version of pragmatic naturalism (Kitcher 2011). However, arguing for pragmatic naturalism would take me beyond the present project, and so I remain officially uncommitted in the text.

Of course, the claim that UMET is not very plausible does not *logically entail* that there is some other metaethical view that is more plausible. Perhaps no metaethical view is plausible. Suppose, for instance, that our rational credence in UMET should be (say) 0.16, with reductive naturalism, Cornell Realism, hermeneutic fictionalism, Blackburn's quasi-realism, Cambridge pragmatism, and Kramer/Dworkin-style realism all commanding rational credence of 0.14. Then no metaethical theory is more plausible than UMET, but UMET is implausible nonetheless. So the phrasing in the text expresses optimism on my part that might be undue. Still, notice that in the present example, on none of the views in which we should invest credence of 0.14 are ethical questions pseudo-questions. So, in the context of arguments focussing on the indispensability of ethical life, it is their disjunction that should be contrasted with UMET, not each one individually. In the present example, we ought to invest confidence of 0.84 in this disjunction.

[59] This might not to amoralists, egoists and psychopaths. I am setting such characters aside here, as I did in Chapter 3.

Inescapability. (However, it will not be strictly necessary to my overall argument, so not much changes if this is not granted.)

A different and stronger claim about the ethical project is this:

> **Intrinsic Indispensability (of the Ethical Project)**: for most of us, the ethical project is intrinsically indispensable.

Here, the notion of intrinsic indispensability is to be understood exactly as it is in Enoch's argument – a project is intrinsically indispensable if it is rationally non-optional – that is, if it is rationally required. So, inherent in *Intrinsic Indispensability*, but entirely missing from *Psychological Inescapability*, is rational endorsement of the ethical project. According to the former, but not to the latter, engagement with the ethical project is rationally required (for most of us).

Now I am not (yet) defending *Intrinsic Indispensability*; I am simply introducing it as a claim that will be discussed presently. The important thing to emphasise at this juncture is that it is to be read in accordance with the account of practical rationality discussed in Chapter 1. So, *Intrinsic Indispensability* does not claim that each of us has (or, indeed, any of us have) an NICR to commit to the ethical project rather than disengage from it. For my purposes, the normativity of the 'rational non-optionality' of the project should be understood to derive from the agents subjective motivational set (as that notion was discussed in Chapter 1).

Compare this with what was said about epistemic normativity in Chapter 2. There is no categorically normative reason for you to commit to the project of forming your beliefs about the causal order of the universe on the basis of scientific discovery rather than wishful thinking or astrology; it is just that, given the sort of creatures we are, and given our interests in having largely true beliefs rather than largely false ones, the project of forming your beliefs about the causal order on the basis of scientific discovery rather than wishful thinking is rationally non-optional for you. Likewise, what *Intrinsic Indispensability* claims is that given the sort of creatures we are – emphasising here our nature as social creatures – we have *overwhelming reason* to commit ourselves to the ethical project.

But there is more to intrinsic indispensability than merely having an overwhelming reason. For sometimes we might find ourselves in tragic circumstances, in which we have overwhelming reason to do something that is in fact impossible. Example: in a nearby possible world at some point in recent history, an asteroid of roughly the size that caused the extinction of the dinosaurs is heading for earth. If the technology exists to change the asteroid's course so that it will harmlessly pass Earth by, then deploying that technology is presumably rationally required. (Again, given the subjective motivational set of all the agents concerned.) But if the technology does not exist and cannot be developed in time to prevent disaster, then it would be strange to describe the project of developing the technology and deploying it before the strike as rationally required. This is an impossible project, and we cannot be rationally required to do that which is impossible.[60] Nonetheless, it is clearly deeply unfortunate that the technology does not exist and cannot be developed. Certainly if the technology could be developed, the inhabitants of this world would have overwhelming reason to develop it.

Projects like this I will call *merely exigent projects*. A merely exigent project is one that would be rationally non-optional were it possible to engage in the project successfully, but which in fact cannot be successfully engaged in. Some merely exigent projects can be dropped, with much sadness: the asteroid case is like that. Other merely exigent projects might be psychologically inescapable, such that their mere exigence dooms those for whom the project is exigent to futile engagement in a project with no chance of success. Imagine, for instance, a mother who psychologically cannot accept her son's guilt in a proven case of murder and so spends the rest of her days engaging in the doomed project of proving her son's innocence.

Now suppose that UMET is true, that *Psychological Inescapability* is true, but that no U-NWP solution succeeds in allowing us to engage in the ethical project. Then we would appear to have:

[60] Of the "ought implies can" principle, Williams says that it is untrue, but notes that "[i]f my deliberation issues in something I cannot do, then I must deliberate again" (Williams 1985: p. 195). So, it can't be that from the perspective of practical rationality, what the inhabitants of the unfortunate world described in the text should do is to develop and deploy the technology. If that is their conclusion, they must deliberate again. So the project of developing the technology is not intrinsically indispensable.

> **Mere Exigency (of the ethical project):** for most of us, the ethical project is merely exigent.

The consequence here would be that we are in the same position with respect to the ethical project that the mother of the murderer is in with respect to her project of proving his innocence: doomed inescapably to futile engagement.

Now, whilst *Mere Exigency* and *Intrinsic Indispensability* do not together exhaust logical space – for the ethical project could be rationally optional (or even rationally forbidden!) – I take it that they do exhaust logical space once *Psychological Inescapability* is granted. That is, I think the following inference is a sound one:

> If Psychological Inescapability holds, then either Intrinsic Indispensability or Mere Exigency holds.

And, of course, given that I have asserted that *Psychological Inescapability* holds, then I also accept the disjunctive consequent of this inference. Because this is one of two disjunctive propositions I need in the discussion to come, it will be convenient to give it a name:

> **Intrinsic-or-Exigency:** either *Intrinsic Indispensability* or *Mere Exigency*.

Important here is that *Intrinsic Indispensability* and *Mere Exigency* are mutually exclusive, and so the "or" should be read as an exclusive "or".

§5.5: Another Disjunction — Loose or Heavy

Whether or not the ethical project is intrinsically indispensable, what might be instrumentally indispensable to that project? Instrumental indispensability is to be understood, much like intrinsic indispensability, in the way that Enoch presents it: something is instrumental to a given project just in case without that thing, the whole point of engaging in the project is undermined.

Here are two candidate claims about instrumental indispensability to the ethical project:

> **Instrumental Indispensability of Some Loosely Deontic Concepts** [henceforth abbreviated to: **Loose**]: there is some set S of ethical concepts such that:
>
> i. Some elements in S are loosely deontic ethical concepts
> ii. S as a whole is instrumentally indispensable to the ethical project

iii. S contains no heavily deontic concepts, and no deontic concepts other than those in S are instrumentally indispensable to the ethical project.

Instrumental Indispensability of Some Heavily Deontic Concepts [henceforth abbreviated to: **Heavy**]: there is some set S* of ethical concepts such that:

i. Some elements in S* are heavily deontic ethical concepts
ii. S* as a whole is instrumentally indispensable to the ethical project.

Now *Loose* and *Heavy* together do not exhaust logical space: it might be that one can successfully engage in the ethical project without using any deontic concepts, heavily deontic or otherwise. For the time being, I shall ignore this complication and proceed as if *Loose* and *Heavy* jointly exhausted logical space. It seems unlikely that ethical life can proceed without *any* notion of obligation, even a loose notion that is to be seen "as merely one ethical consideration amongst others" (in Williams' words). And anyway, the UMET-proponent whose suspicions about morality are based in large part on doubts about the inescapable authority of moral obligation is hardly in a position to deny this. Hence:

Heavy-or-Loose: either *Heavy* or *Loose*.

As with *Intrinsic-or-Exigency*, the disjuncts here are mutually exclusive. (This is owing to (iii) in *Loose*.) So, once again, the "or" here is to be read exclusively.

§5.6: An Undisputed Premise

The UMET-proponent is of course committed to:

Heavy-Error: Heavily deontic concepts are legitimate targets of an error theory.

This is common ground between me and the UMET-proponent, and my version of the objection from loss needs to be consistent with it.

§5.7: A Licencing Principle

In Enoch's indispensability argument, it is not enough for him just to argue that the deliberative project is intrinsically indispensable and that irreducible normative truths are instrumentally indispensable to that project. He also needs some sort of licencing principle. His is: "[i]f something is instrumentally indispensable to an intrinsically

indispensable project, then we are (epistemically) justified (for that very reason) in believing that that thing exists" (Enoch 2011: p. 83). To make my version of the objection from loss work, I need a licencing principle a little bit like this. But since Enoch is attempting to do much more than I am, mine can afford to be more modest: the justification does not need to be epistemic and the commitment does not need to be ontological.

Thus, I suggest the following principle:

> **Licence**: If P is an intrinsically indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.

Does **Licence** generate commitments (normative or otherwise) from nothing, avoiding error theories on the basis of merely wishful thinking? I do not think so. We need to remember the distinction between intrinsically indispensable projects and merely exigent projects. Consider again the tragic case of the mother who is psychologically incapable of disengaging from the project of proving her son's innocence. Instrumentally indispensable to any project of proving that *p* is that *p* is indeed true. So if the mother's project counted as intrinsically indispensable, then **Licence** would tell us that we could not be error theorists about the son's innocence. However, in the case as I described it, the son's guilt has been established beyond any reasonable doubt: an error theory about the son's innocence is exactly what we should go in for. But it is of course for precisely this reason that the mother's project is merely exigent rather than truly intrinsically indispensable. The following argument, then, would be unsound:

> [1] The mother is engaged in an intrinsically indispensable project of proving her son's innocence.
>
> [2] It is instrumentally indispensable to the project of proving her son's innocence that her son is indeed innocent.
>
> [3] **Licence**: If P is an indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.
>
> [4] Therefore: her son's innocence cannot successfully be targeted by an error theory; the son is indeed innocent.

Rather, the argument should go precisely the opposite way:

> [1] It is instrumentally indispensable to the mother's project of proving her son's innocence is that her son is indeed innocent.
>
> [2] **Licence**: If P is an indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.
>
> [3] The son's innocence is the legitimate target of an error theory
>
> [4] Therefore: the mother's project is not intrinsically indispensable, even if she lacks the ability to disengage from it.

When this is borne in mind, I think that **Licence** is acceptable. It should certainly be significantly less controversial than Enoch's licencing principle, quoted in the first paragraph of the present section.

§5.8: My Argument for *Indispensability Premise 1*

Let us review the materials I have assembled so far. I have defended the following premises:

> **Intrinsic-or-Exigency**: either for most of us, the ethical project is intrinsically indispensable or for most us, the ethical project is merely exigent.
>
> **Heavy-or-Loose**: either heavily deontic ethical concepts are instrumentally indispensable to the ethical project, or only loosely deontic ethical concepts are instrumentally indispensable to the ethical project.
>
> **Heavy-Error**: Heavily deontic concepts can successfully be targeted by an error theory.
>
> **Licence**: If P is an indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.

My version of the objection from loss essentially boils down to the question of which disjunct of **Intrinsic-or-Exigency** we should think is true. Is the ethical project intrinsically indispensable, or is it merely exigent? Because I think it is deeply implausible that we should be in the position with respect to the ethical project that the mother of the murderer finds herself in with respect to the project of proving his innocence, I think we should accept **Intrinsic Indispensability** and not **Mere Exigency**. The UMET-

proponent, meanwhile, must accept **Mere Exigency** and reject **Intrinsic Indispensability**. I find this a price in terms of plausibility too high to pay. So, I think that we can be reasonably confident that UMET is false.

Thus, my claims are that one of the following two arguments must be sound, and it is the second rather than the first that is sound:

Argument 1: The Ethical Project is Merely Exigent

[1] **Licence**: If P is an indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.

[2] **Heavy**: There is some set of S* of ethical concepts such that:

i. Some elements in S* are heavily deontic ethical concepts
ii. S* as a whole is instrumentally indispensable to the ethical project.

[3] **Heavy-Error**: Heavily deontic concepts are legitimate targets of an error theory.

[4] **Intrinsic-or-Exigency**: either [for most of us, the ethical project is intrinsically indispensable] or [for most of us, the ethical project is merely exigent].

[5] Therefore: For most of us, the ethical project is merely exigent.

Argument 2: The Ethical Project is Intrinsically Indispensable

[1] **Licence**: If P is an indispensable project and C is instrumentally indispensable to P, then C cannot successfully be targeted by an error theory.

[2] **Heavy-or-Loose**: either:

Heavy: There is some set of S* of ethical concepts such that:

i. Some elements in S* are heavily deontic ethical concepts
ii. S* as a whole is instrumentally indispensable to the ethical project.

or

Loose: There is some set of S of ethical concepts such that:

i. Some elements in S are loosely deontic ethical concepts
ii. S as a whole is instrumentally indispensable to the ethical project
iii. S contains no heavily deontic concepts, and no deontic concepts other than those in S are instrumentally indispensable to the ethical project.

[3] **Heavy-Error**: Heavily deontic concepts are legitimate targets of an error theory.

[4] **Intrinsic Indispensability**: for most of us, the ethical project is intrinsically indispensable.

[5] Therefore: there is some set S of ethical concepts such that:

i. Some elements in S are loosely deontic ethical concepts
ii. S as a whole is instrumentally indispensable to the ethical project
iii. S contains no heavily deontic concepts, and no deontic concepts other than those in S are instrumentally indispensable to the ethical project.

Each argument accepts all of the four claims I have made so far: *Licence*, *Heavy-or-Loose*, *Heavy-Error*, and *Intrinsic-or-Exigency*. Because the two disjunctions here are exclusive, which disjunct you choose to commit yourself to in one disjunction, in combination with the other two claims, commits you to a specific disjunct in the other disjunction. The choice is between the conjunctions *Intrinsic-and-Loose* and *Exigency-and-Heavy*.

The fact that we have to choose between *Intrinsic-and-Loose* and *Exigency-and-Heavy* entails that *Indispensability Premise 1* is true. For UMET is committed to *Heavy*, meaning that the UMET proponent has to opt for *Exigency*. On *Exigency*, successful engagement in the ethical project is impossible, even if we are doomed by our psychology to continued failed efforts to so engage. Thus, the UMET-proponent cannot say that successful engagement with the ethical project is possible. Hence:

Indispensability Premise 1: if UMET is true, then it is not the case that for most of us, engagement with the ethical project is both possible and of overwhelming importance.

Functional Continuity, Again

There is a link between the argument given for *Indispensability Premise 1* and the argument against substitutionalism in the previous chapter. This is that in committing themselves to *Exigency-and-Heavy*, the UMET-proponent claims that *in order for ethical life to serve the ends it needs to serve, strongly deontic concepts are strictly necessary*.

If an alternative strategy is attempted, one on which it is admitted that ethical life can, without availing itself of strongly deontic concepts, achieve much of what we recognise as ethics, but on which this conception of ethical life somehow falls short of being "the

real thing", then the UMET-proponent will run headlong into the problems identified for the substitutionalist in the previous chapter. They will have to deny that the high degree of functional continuity between (error-infected) strongly deontic concepts and (error-free) loosely deontic concepts has the result that these are roughly the same concepts. This means rejecting my proposal that the conceptual continuity of any two ethical concepts *a* and *b* is equal to their functional continuity, and this is likely going to mean retreating to the non-negotiable commitment account of conceptual continuity. But in that case the UMET-proponent is vulnerable to Finlay's complaint that error-theorists interpret standard users of moral discourse in a gratuitously uncharitable way. We saw in the previous chapter that in this context, this complaint has more force than in the context in which Finlay originally made the point.

§5.9: How to Engage in the Ethical Project Successfully

Because they are committed to *Exigency-and-Heavy*, UMET-proponents are committed to:

> **Loose Insufficiency:** loosely deontic concepts are not sufficient for successful engagement in the ethical project; strongly deontic concepts are instrumentally indispensable.

I, on the other hand, think that *Loose Insufficiency* is deeply implausible. Instead I claim:

> **Loose Ethical Sufficiency**: loosely deontic concepts are sufficient for successful engagement in the ethical project; strongly deontic concepts are not instrumentally indispensable.

And if *Loose Ethical Sufficiency* is correct, then we can successfully engage in the ethical project. This gives us:

> **Indispensability Premise 2**: engagement with the ethical project *is* both possible and of overwhelming importance for most of us.

In this section, I am going to support *Loose Ethical Sufficiency* and thus *Indispensability Premise 2* by direct demonstration. If someone believes that it is impossible to reach the top of Lord Hereford's Knob on a sunny July day without a pair of sturdy hiking boots, then you can prove that they are wrong by doing so in skate shoes. Likewise, if someone believes that strongly deontic concepts are necessary for successfully engaging in the

ethical project, then you can prove them wrong by doing so using only weakly deontic concepts. What I am going to do, then, is successfully engage in the ethical project by adjudicating correctly an ethical debate, without appealing to any strongly deontic concepts. My success in doing so will show *Loose Ethical Sufficiency* to be correct, and thus show that Argument 1 of §5.8 is to be rejected in favour of Argument 2. UMET will then have been shown not to be true.

The debate concerns the question of whether the death penalty in the United States ought to be retained. Pojman (2013) argues in favour of its retention; Nathanson (2013) argues that it should be abolished. I shall be defending the following claims:

> [1] Pojman's argument can be understood in such a way that Pojman does not commit himself to the availability of heavily deontic concepts;
>
> [2] Nathanson's argument can be understood in such a way that Nathanson does not commit himself to the availability of heavily deontic concepts;
>
> [3] It can be clearly seen that the considerations that Nathanson adduces are decisive against retention of the death penalty.

Because I will be rejecting Pojman's argument, claim (1) is quite dispensable to my case. It is no part of my case against UMET to claim that the elimination from our ethical conceptual toolkit of strongly deontic concepts will make no difference to the normative conclusions to which we come. Indeed, my interest in LMET is in no small part due to my belief that it will make normative differences, that the error of taking strongly deontic concepts to be in good working order leads to concrete normative errors downstream. So, one method of objection to Pojman that is in principle available to me, and perhaps plausible, would be to point to various points in his argument that could be supported only by appeal to strongly deontic concepts, and proclaim: "error!" However, claim (1) is only that Pojman's argument *can be understood* in a way from which commitment to the availability of strongly deontic concepts is entirely absent. I see this as helping to establish *Loose Ethical Sufficiency* in the following way. Even if you are not convinced that Nathanson's arguments against the death penalty are decisive, and even if you think that this is because it is Pojman and not Nathanson who is correct in this debate, still you

don't need heavily deontic concepts. There is a way to make your case in terms only of loosely deontic concepts. Indeed, it is the best way available to you – for any theoretical debt you incur to heavily deontic concepts will bankrupt your argument. Of course, I maintain that no such case will in the end be convincing, for the simple reason that it is not true that capital punishment is ethically acceptable.

Here, then, is how I shall proceed. First I shall briefly summarise Pojman's case for retention. Then I shall emphasise how it can be understood entirely in terms of loosely deontic concepts. Then I shall draw replies and criticisms to Pojman from Nathanson's essay, emphasising once again the dispensability of strongly deontic concepts in Nathanson's argument. These replies will be seen, I hope, to be successful, such that we can establish *that capital punishment is unethical*. Establishing this conclusion obviously constitutes successful engagement in the ethical project. And so *Loose Ethical Sufficiency* will have been proven.

§5.9.1: Pojman – For the Death Penalty

Pojman defends the death penalty on grounds both of desert and deterrence. With regards to desert, he writes:

> Human beings have dignity as self-conscious rational agents who are able to act morally. One could maintain that it is precisely their moral goodness or innocence that bestows dignity and a right to life on them. Intentionally taking the life of an innocent human being is so evil that the perpetrator forfeits his own right to life. He or she deserves to die.
>
> (Polman 2013: p. 160.)

There are clear Kantian overtones to Polman's language here, and so it might be supposed that he makes tacit appeal to strongly deontic concepts of precisely the sort that AfR targets. If that is the correct interpretation of Pojman's claims, then they ought to be rejected on the basis of an LMET targeting inescapable authority. But even if this is Pojman's intended meaning, is it the best interpretation? Does reading this as an appeal to the rational agency of a free moral agent give these considerations their best chance of being persuasive?

Perhaps not. Dworkin (2011: Chapter 9) thinks of human dignity as comprising two parts: self-respect and authenticity. To have self-respect is to take one's own life seriously; a person with self-respects accepts "that it is a matter of importance that his life be a successful performance rather than a wasted opportunity" (Dworkin 2011: p. 203). The principle of authenticity amounts to the principle that every individual "has a special, personal responsibility for identifying what counts as success in his own life" (p. 204). Now Dworkin is of course keen to stress the objectivity of these twin principles. Against proposals that we just "happen to want to live in a particular way" (p. 207) and that this can account for the role of "overall plans and projects" in life – and by extension, ultimately all value in life – he complains that "we aim to meet a standard, not just pick at random from a menu". But suppose that Pojman elected to construct something like Dworkin's concept of dignity only out of the materials available to those committed to *Heavy Error*. On this (somewhat existentialist) view, the twin principles of authenticity and self-respect would be fundamental, and there would be a sense in which they really were just "picked at random from a menu" – like Sartre's concept of a fundamental project. This would not prevent Pojman from developing, from these fundamental values, chosen in the face of the absurd as it were, an account whose normative consequences were broadly similar to those of Dworkin's. Consider Dworkin's question: "[d]o you value your life as objectively important in virtue of something special about your life, so that it would be perfectly consistent for you not to treat other human lives as having the same kind of importance? Or do you value your life in that way because you think all human life is objectively important?" (p. 256). On the present suggestion, Pojman could answer – again, in an existentialist vein, and so deleting the word "objectively" – that, according to *his* values, all human life is important, not just his own. This could then be the sense in which the claim that "human beings have dignity as self-conscious rational agents who are able to act morally" should be understood. And then, he would have got going a relatively well-defined deliberative project from which he could make the claims in the quotation cited earlier, but which made no use of heavily deontic concepts.

Of course, the notion of desert to which Pojman appeals might be targeted by another of the arguments mentioned in Chapter 1 but not discussed in detail: the one based on

scepticism about moral responsibility. If this argument is taken to establish an LMET, then Pojman's normative claims could be rejected here on account of that LMET. However, since the question of scepticism about moral responsibility has not been much of a concern in this thesis, I shall pass over this point here. Clearly anyone who argues against the death penalty on such grounds (as does Pereboom 2001: chapter 8) does so from the ethical point of view. They must take the ethical point of view to be legitimate.

The main normative consideration that Pojman adduces to support the claim quoted above that murderers "deserve to die" is a general principle that *some harm equivalent to that which the murderer inflicted on his victim should come to the murderer himself*. This is an obvious appeal to the "eye for an eye" principle, to which Pojman refers approvingly in the course of defending his desert-based argument from the objection that it represents the pursuit of revenge rather than justice:

> [R]etribution stipulates that the offender be punished in proportion to the gravity of offense. In this sense, the *lex talionis* that we find in the Old Testament is actually a progressive rule, where retribution replaces revenge as the mode of punishment. Revenge demands a life for an eye or a tooth, but Moses provides a rule that exacts a penalty equal to the harm done by the offender.
>
> (Pojman 2013: p. 168.)

Once again, if we imagine a version of Pojman who has constructed an existentialist analogue of Dworkin's idea of dignity and uses that to ground his desert-based argument, the above quotation can be seen as being made from that point of view. We have still not encountered anything that commits Pojman to the use of heavily deontic concepts.

The dispensability of heavily deontic concepts to Pojman's arguments is even clearer in the case of the other part of his justification of retention – that the death penalty is an effective deterrent. Now it is commonly pointed out that it is not proven that the death penalty is in fact an effective deterrent. Pojman accepts that this is the case, but claims equally that it is not proven that the death penalty is *not* an effective deterrent, which he thinks is the claim that death penalty abolitionists need, and to which they frequently help themselves. He then offers an argument based on "common sense" for supposing that the death penalty probably is an effective deterrent. I do not need to go into the details of this

argument to make the obvious point: the debate at this point concerns a simple matter of empirical fact. No need for heavily deontic concepts here.

In conclusion, Pojman's argument can be made in a way such as to avoid commitment to the non-error-infected status of heavily deontic concepts. Thus I have shown the first of my three claims to be correct. Consequently, to the extent that Pojman's arguments constitute successful engagement with the ethical project, then such engagement does not require heavily deontic concepts, and *Loose Insufficiency* is disproved. But of course, *I* do not think that this establishes *Loose Ethical Sufficiency*, and neither should anyone else, because Pojman's arguments are to be rejected. They do not in fact constitute entirely *successful* engagement in the ethical project. To show how entirely successful engagement in the ethical project is possible without strongly deontic concepts, I shall now turn to Nathanson.

§5.9.2: Nathanson – Against the Death Penalty

Nathanson (2013: p. 175) attempts to show that "punishing people by death is an unjust and immoral practice". (Of course, I want to say that what he shows is that it is unjust and *unethical*.) His essay opens with optimism about the prospects of *showing* this, rather than merely formulating an argument for a conclusion that will ultimately be subject to irresolvable disagreement, as often happens in other debates:

> Some controversies are hard to resolve because people on opposing sides differ in their fundamental values. In such cases, it is hard to find values that can serve as a basis for reaching agreement. The death penalty debate is not like this. Both death penalty supporters and opponents generally appeal to the same fundamental values: the pursuit of justice and respect for human life.
>
> (pp. 175-176.)

In the present context, this recalls Finlay's (2014) end-relational analysis of normative language, according to which predicative uses of 'good' (as in 'the death penalty is good') are elliptical for 'it is good for *e* if *p*'. What Nathanson is drawing our attention to here is that both retentionists and abolitionists about capital punishment are both pursuing the same ends – namely, the end of respecting human life and the end of pursuing justice, which we can here conceptualise as people *getting what they deserve*. (As above, it is a

slight but manageable complication that the notion of desert itself may well be successfully targeted by an error theory.)

Nathanson's discussion is in two parts. In the first, he considers capital punishment in theory, and argues for the conclusion that "the principled bases for the death penalty are extremely weak" (p. 176). In the second part, he considers the death penalty in practice, and concludes that even if his arguments about the death penalty in theory are to be rejected, still the death penalty ought to be abolished.

In the first part of his discussion, he considers the same sort of deterrence argument that we saw Pojman consider above. Nathanson aims to challenge the principle that "if a punishment deters more murders and thus saves innocent lives, then it is justifiable". To do so, he points out that "we can imagine punishments that have greater deterrent value than either the death penalty or imprisonment and yet would be wrong to inflict" (p.177). For instance, we can imagine a penal code that demands that not just the murderer but also his family and/or closest friends are to be executed too. This, we might suppose, would have an even greater deterrent effect on potential murderers than the death penalty in its current form. And yet it would clearly be unjust to impose such a punishment. The conclusion that Nathanson draws from this is that "the deterrence argument is not enough to justify [the death penalty], even in theory".

Next, he considers the case for the death penalty from desert, again much as we saw Pojman argue above. He rejects the "eye for an eye" principle because: (i) it would require unjust or barbaric punishments in response to unjust or barbaric crimes (if, for example, we describe a certain crime as the killing of the family of one's enemy, then the "eye for an eye" principle seems to demand the killing of the murderers family); (ii) the "eye for an eye" principle fails to take into account all sorts of factors that we typically take to be relevant to desert other than harm to a victim. Indeed, a victim can suffer the same harm (death) as a result of crimes of varying degrees of seriousness. Nathanson considers the following spectrum:

- A hired killer lies in wait and shoots the intended victim
- An argument degenerates into a fight in which one person strikes the other and kills him

- A person sets fire to a building, thinking that it is empty; several people in the building die in the fire
- A drunken driver kills a pedestrian.

(p. 179.)

All these crimes are serious, but Nathanson's point is that they are not equally serious: the first is more serious than the second; the fourth, whilst still a serious crime, is the least serious of all. And nobody thinks that the death penalty is appropriate for any but the first crime. But in each case the harm to the victim is the same: death. So the "eye for an eye" principle cannot be correct: it is incapable of reflecting the complexity of judgements about desert.

Now I said above that Nathanson's arguments can be seen to be successful. However, I should stress that I think that it is in the second part of his discussion, in which he considers the death penalty in practice, that his arguments prove really conclusive. I do not want to commit myself to the claim (or indeed to its denial) that this first part of his discussion is a clear example of successful engagement in the ethical project. Nathanson is still here highly wedded the notion of desert, and moreover desert in its connection to the deliberate imposition of harms (rather than the distribution of economic resources), and it is entirely possible that such a notion could be successfully targeted by an LMET. So, at least with respect to Nathanson's discussion of capital punishment *in theory*, I will rest content with the same claim as I made with respect to the argument of Pojman's, which I reject: *to the extent that this constitutes engagement in the ethical project*, it is engagement sense of which can be made strictly in terms of loosely deontic concepts. Still heavily deontic concepts have not been needed, as evidenced by our ability to interpret everything that Nathanson says here in terms of Finlay's end-relational theory of the good. Nathanson, in effect, has been arguing that the death penalty is bad, where this is understood as elliptical for the claims that: (i) relative to the end of acting in such a way as to respect human life, the death penalty is bad, and (ii) relative to the end of making sure people get what they deserve and do not get what they do not deserve, the death penalty is bad.

So we turn to the second part of Nathanson's discussion, in which he considers capital punishment in practice. Here he needs to be understood as accepting *arguendo* that his discussion of capital punishment in theory has failed to establish that capital punishment is unjust and unethical. Let's suppose, then, there are certain criminals who really do deserve to receive the death penalty, and that perfect justice would be served by executing them. "It would still not follow," Nathanson says (p. 180), "that the death penalty should be adopted. Why? Because we need to consider how this punishment works out in practice". In actual fact, he argues, the real institution of capital punishment, both as it is now known and in all its feasible future incarnations, violates the values noted earlier to be common between opponents and supporters of the death penalty: respect for human life, and justice.

The death penalty is inconsistent with the value of justice, he argues, because whether or not a death sentence is imposed in any given case typically depends not on actual moral desert, but on a host of clearly irrelevant factors. The three main factors he considers are: (i) race – here he notes that "a large body of research has shown that sentencing in capital cases is very much influenced by both the race of the offender and the race of the victim" (p. 181); (ii) socio-economic status – you are far more likely to be sentenced to death if you are of low socio-economic status than if you are of high socio-economic status; (iii) quality of legal representation – the fate of the offender depends in large measure on the competence of the legal team representing them, with the consequence that those who have to rely on court-appointed lawyers are 2.6 times more likely to receive a death sentence.

The death penalty is inconsistent with respect for human life, Nathanson argues, for the following reasons. First, there are all the factors considered in the previous paragraph. In tolerating a state of affairs in which irrelevant factors play a substantial role in determining who gets executed, the system "expresses indifference toward the value of these defendants' lives" (p. 183). Second, there is not just the possibility but the routine occurrence of errors. Two types of errors are possible here, Nathanson thinks: a genuinely guilty party may receive a death sentence when only imprisonment is truly deserved, or a defendant entirely innocent of the crime with which they are charged might be

erroneously found guilty and receive a death sentence. Appealing to research concerning the frequency of the latter kind of mistake, Nathanson argues that these errors are worryingly frequent. Again, this clearly expresses lack of concern for the value of human life: supporting the death penalty in light of these figures means supporting a policy that one knows will have the consequence that innocent people will be killed. Third, there is the lack of interest the system takes in rectifying mistakes. As illustration of this, Nathanson considers the existence of time limits on the submission of new evidence to exonerate sentenced individuals:

> In *Herra* v. *Collins* (1993) the Court ruled that new evidence in support of a claim of innocence could be disregarded because it had been submitted too late to meet the Texas 60-day deadline. In other words, the Court ruled that a person could be executed even though there was now evidence that he or she is innocent. Why? Because the evidence came in too late.
>
> (p. 185.)

I am sure that all will agree that the disregard for the value of human life here is *blatant*. It should be equally obvious that no appeal to NICRs is necessary for a thoroughgoing and robust opposition to practices such as this.

Now, of course, a supporter of the death penalty could push back on all these points with an insistent optimism that the system could and should be improved: procedures could and should be put in place so that irrelevant factors no longer determine whether a given criminal receives a death sentence, investigative procedures could and should be tightened so that mistakes are not made; time limits could be lengthened or abolished. It seems to me that there are solid grounds for pessimism on all these points, so these replies are not going to work.

§5.9.3: The Lesson of the Debate Just Considered

My discussion of Pojman and Nathanson has been brief and swift, but enough has been said for us to get back to the chase: have my claims (2) and (3) from §5.9.0 been established, with the consequence that *Loose Ethical Sufficiency* has been demonstrated?

Claim (2) was that Nathanson does not appeal to heavily deontic concepts. That this is the case should be clear from the exposition of the previous subsection, especially the

second part of the argument concerning the real-life institution of capital punishment. Nathanson's materials are purely these: (i) a commitment to the end of respecting the value of human life; (ii) a commitment to the end of respecting justice; (iii) some pertinent social empirical facts concerning the actual operation of the real life institution of capital punishment and the extent of the possibility of its reform. Where in these materials is the commitment to heavily deontic concepts? Certainly not in the empirical social facts. And it cannot be plausibly maintained that (i) embodies a commitment to heavily deontic concepts either: clearly most of us *do* care about the end of respecting the value of human life, and so the relevant deontic concepts can all be relativised to this end. (Anybody who does not care about this end is free and welcome to reject the arguments for this reason, but then we are unlikely to suppose that such a person is really party to the ethical project as it relates to the question of capital punishment.) So, a desperate UMET-proponent at this point will have to cling on to the idea that the commitment to the end of respecting justice somehow involves Nathanson in the mess of strongly deontic concepts. The idea here must be that the credentials of justice-talk depends on there being sense to be made of NICRs.

But whatever is to be said with regards to most uses of the word "justice" in ethical discourse as we currently know it, it is clear that in the limited context of the second part of Nathanson's argument against the death penalty, no such accusation of error has anything going for it. Nathanson's point is that the supporter of capital punishment is committed to the possibility of a just institution of capital punishment. Perhaps (but not certainly), in appealing to justice here the supporter of the death penalty involves themselves in the errors of heavily deontic concepts. But Nathanson does no such thing when he points out that the current system *fails to treat like cases alike*. All that Nathanson needs is the loosely deontic and eminently sensible claim that a system that fails to treat like cases alike when deliberating about the apportioning of serious harms on individuals cannot be considered just. This much can be captured with a very minimal concept **JUSTICE**, on which something is unjust if it violates the end of treating like cases alike. That is to say, even if the conception of **JUSTICE** with which we are familiar involves appeal to heavily deontic concepts and is therefore error-infected, it remains an option to

formulate a new, highly minimal conception on which treating like cases alike is all that is constitutive of justice. This places one and only one constraint on deliberation concerning the distribution of harms and resources: that where there is a difference in allocation, it must be supported by a difference in the cases.

Also, although Nathanson *uses the words* "punishing people by death is an unjust and immoral practice", if **JUSTICE** itself were to be targeted by an error theory, Nathanson would still have available to him the claim that "it is not the case that punishing people by death is just". In this way, Pojman's case *for* capital punishment would be undermined, and Nathanson would still have at his disposal the thread of his argument that is relativised to the end of respecting the value of human life in order to make his case against it.

In sum, then, the lesson to take is that Nathanson can easily make his argument without appealing to heavily deontic concepts, and that since his argument against capital punishment is successful, we have here an example of successful engagement in the ethical project without the use of heavily deontic concepts. So, heavily deontic concepts are not instrumentally indispensable to the ethical project. So, of the two arguments considered in §5.8, we should prefer Argument 2. So the ethical project is intrinsically indispensable. In other words:

> **Indispensability Premise 2**: For most of us, engagement with the ethical project is both possible and of overwhelming importance.

§5.10: The Objection from Unwitting Realism Revisited

It should now be obvious why the objection from unwitting realism fails. I have made good on my promise to develop my indispensability argument without recourse to strongly deontic concepts, of which irreducibly normative claims concerning justice and the value of human life must surely be an instance. Nothing I that I have said in the present part of this chapter goes beyond the "non-Humean instrumentalist" conception of practical rationality discussed in Chapter 1. So all the normativity I have appealed to here is reducible: the ethical truth that the death penalty is unethical and unjust reduces to the truths that: (i) the institution of the death penalty shows blatant disregard for the value of human life and routinely fails to treat like cases alike; (ii) no realistically achievable

modifications to the institution are likely to reliably eliminate these shortcomings; (iii) it is a matter of fact about us that we *are* committed to the ends of respecting the value of human life and treating like cases alike. There are no heavily deontic concepts here, and there is no irreducible normativity. The toils of the present part of this chapter have been against the UMET-proponent, but not for the robust realist.

§5.11: An Objection – Truth vs. Practice

The objection from loss clearly has the same form as the Moral Moorean Arguments discussed and rejected in Chapter 2. Recall that the Moorean Schema is:

> **Moorean Schema**:
> [1] If S then not-M
> [2] M
> [3] Therefore: not-S

My argument obviously fits this formula:

> **Indispensability (Moorean Schema)**
> [1] If UMET is true then it is not the case that for most of us, the ethical project is intrinsically indispensable.
> [2] For most of us the ethical project is intrinsically indispensable.
> [3] Therefore: UMET is not true.

Why should we suppose that this argument performs any better than the Moral Moorean Arguments? Answer: the original Moral Moorean Arguments failed because they put things in terms of epistemic rationality. These arguments compare two *propositions*, asking which is the more dubious. The indispensability argument does not look to epistemic rationality but instead to practical rationality. It asks: do we have overwhelming reason to engage in the ethical project, and can we make a success of it? The answers given are that we do have overwhelming reason, because people need to be able to rely on certain things which can be guaranteed only by some form of the ethical life, and that we can make a success of this project. So when it is claimed that *Intrinsic-and-Loose* is more plausible than *Exigency-and-Heavy*, the epistemic notion of 'plausibility' is being deployed in a way that is entirely parasitic on practical rationality. We are not dealing here with *truth* so much as we are dealing with *deliberation*; we are less interested in

which propositions to believe than in which actions to perform, policies to enact, and modes of behaviour to conform to.

But now this seems to open me up to a counter-objection. According to this counter-objection, because I have shifted my focus away from the epistemological and/or metaphysical considerations on the basis of which UMET-proponents argue for their view and towards considerations concerning deliberation and practicality, I have changed the subject and my objection fails to target error theoretical programmes focussed squarely on epistemological and/or metaphysical considerations. Hence the practical indispensability version of the objection from loss fails.

My response to this counter-objection is in two parts. My first line of response is that is not clear that all error theoretical arguments in the literature *are* focussed entirely on epistemological and/or metaphysical considerations to the exclusion of considerations concerning deliberation and practicality. Recall that the main prong of AfO as it was formulated in Chapter 1 was the obscurity of the notion of moral obligation, and the argument by which the notion of moral obligation was convicted of obscurity was based on the non-Humean instrumentalist view of practical reason. Here, the UMET-proponent is attempting to derive a metaphysical conclusion (viz., that there are no NICRs) from broadly the same sort of practical considerations with which I have been concerned in my formulation of the objection from loss (viz., those concerning the correct account of practical rationality). So it is not the case that arguments for UMET are quite as squarely focussed metaphysical and/or epistemic matters as the counter-objection appears to suggest.

To what extent does this first part of my response apply to the other strands of AfO? Recall that the second strand concerned the failure of philosophers to reach consensus on any foundational proposition about morality. Here an abductive inference was made: the idea was that part of the best explanation of this failure was that there were no truths in the area to be found. The *conclusion* here is squarely epistemic and/or metaphysical, but the evidence for it was made up of empirical observations about the details of disputes concerning foundational propositions about morality. Likewise, part of the evidence for *Indispensability Premise 2* concerned empirical observations about the details of disputes

concerning capital punishment. So the practical indispensability version of the objection from loss is very much in the same business as the second argumentative strand in AfO.

Does this mean, though, that instances of UMET based on the more metaphysical arguments I set aside in Chapter 1, like Streumer's, are immune to the objection from loss? I think not. For one thing, these metaphysical arguments concerning location problems and the prospects for naturalistic reductions of normative properties to non-normative properties have a little more to do with the practical workings of normative discourse than is sometimes admitted. Consider, for instance, that in his book-length defence of GNET, Streumer spends an entire chapter rejecting non-cognitivism on the grounds that it clashes with the constraint on normative language that "when two people make conflicting normative judgements, at most one of these judgements is correct" (Streumer 2017: chapter 6). More importantly, though, if UMET is to have any bite whatsoever, even only as a strictly metaphysical theory, it must be capable of rendering our ethical normative judgements unjustified. And this brings me to my second line of response to the counter-objection.

What *Indispensability Premise 1* says, essentially, is that if UMET is true then every time we make an ethical normative judgement like "capital punishment is ethically wrong" or "you oughtn't evade taxes unlawfully" or even "torturing human babies merely for fun is ethically reprehensible", we involve ourselves in error. According to UMET, we do this every time we enter into any social practice broadly continuous with existing ethical discourse. It doesn't matter how the UMET-proponent ends up with their metaethical theory; it will still be true that according to this theory, we involve ourselves in error every time we make an ethical judgement. Thus it is quite wrong to suggest, as the counter-objection does, that because the practical indispensability version of the objection from loss is concerned with practical matters of deliberation, it doesn't target UMET as a distinctly metaphysical thesis. For this metaphysical thesis has practical implications, as *Indispensability Premise 1* correctly states. And I take myself to have shown that these practical implications are false, as *Indispensability Premise 2* states.

§5.12: Revolutionary Fictionalism, Conservationism, and The Objection from Loss

My argument against UMET is now at an end. I have defended both premises of *Objection From Loss – Practical Indispensability Version* and defended myself from the objection from unwitting realism. One question that might remain in some readers minds is this: how is my opposition to UMET different from revolutionary fictionalism and conservationism's support for retaining ethical discourse even after the error-infection affecting strongly deontic concepts has come to light.

To answer this, it is best to return to *Objection from Loss – Simplest Version*. The second premise of this version of the argument is equivalent to the claim that there is *no sense in which it is illegitimate* to adopt an ethical point of view. Because revolutionary fictionalists and conservationists are UMET-proponents, they cannot accept this claim, even though their U-NWP policy options are motivated by the closely related concern that there is *some sense in which it is legitimate* to adopt an ethical point of view. The claim that there is no sense in which it is illegitimate to adopt an ethical point of view is stronger than this, and once you accept it then you have rejected UMET. But it is the stronger claim that I have been at pains to demonstrate by interrogating Nathanson's convincing case against the death penalty and showing that it does not rely on any error-infected concepts.

Chapter 6:
Conclusion – LMET

One final matter needs tying up. In Chapter 1, I defended *Partial Escape* by pointing to the persuasive power of the various strands of AfO. It has turned out that AfO does not support UMET. UMET is not correct: the practical indispensability version of the objection from loss shows it to be wrong. But still I think we should find a way to accommodate the insights contained in AfO in our metaethical theory. Even as we participate in ethical discourse, happy that the threat to this project posed by UMET has been neutralised, we should still be alive to the fact that there is no sense to be made of NICRs, that there may well be no truths to be identified concerning foundational propositions about morality, that vast swathes of ethical discourse might be mistaken in presuming the validity of an ultimately implausible view of human agency, etc.

The most obvious way to do this is to accept a collection of LMETs. At the very least, we should accept an LMET targeting NICRs. The practical indispensability version of the objection from loss is entirely consistent with such an LMET: the whole point of the objection is that we can get by perfectly well in ethics using only weakly deontic concepts, which are entirely free of any commitment to validity of NICRs. I also think we should accept LMETs targeting the notion of truths about foundational propositions about morality, and targeting the notion of moral responsibility. But I have said much less in this thesis about foundational propositions about morality than I have said about NICRs, and I have said practically nothing about questions concerning moral responsibility, which are vast and complex. So with respect truths concerning foundational propositions about morality, I rest content with a tentative proposal that we should adopt an LMET with this target, rather than an outright assertion. With respect to the notion of moral responsibility, the details and tenability of various LMETs in this area will have to keep for another day.

References

ALSTON, W. (2005). *Beyond Justification: Dimensions of Epistemic Evaluation*. Ithica, NY: Cornell University Press.

ANSCOMBE, G. E. M. (1958). 'Modern Moral Philosophy'. *Philosophy*, 33:124, pp. 1-19.

BAMBOROUGH, R. (1979). *Moral Scepticism and Moral Knowledge*. London: Routledge and Keegan Paul.

BENTHAM, J. (1792). 'Anarchical Fallacies; Being an Examination of the Declaration of Rights Issued during the French Revolution'. Reprinted in J. Bowring (ed.) (1843), *The Works of Jeremy Bentham, Volume 2*. Edinburgh: William Tait.

BERKER, S. (2019). 'Mackie Was Not An Error Theorist'. *Philosophical Perspectives*, 33, pp. 5-25.

BLACKFORD, R. (2019). 'After Such Knowledge – What?' in Garner & Joyce (2019).

BOYD, R.N. (1988) 'How to be a Moral Realist' in G. Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press.

BRANDT, R. (1959). *Hopi Ethics: A Theoretical Analysis*. Chicago, IL: University of Chicago Press.

BRINK, D. (1989). *Moral Realism and the Foundation of Ethics.* Cambridge: Cambridge University Press.

BROWN, C. (2011). 'A New and Improved Supervenience Argument for Ethical Descriptivism' in R. Shafer-Landeau (ed.), *Oxford Studies in Metaethics, Volume 6*. Oxford: Oxford University Press.

BUCKLEY, D. (2020). 'Varieties of Epistemic Instrumentalism'. *Synthese*, 198, pp. 9293-9313.

COPPLESTONE, F. (1965). 'The Existence of God: A Debate' in P. Edwards & A. Pap, *A Modern Introduction to Philosophy*, Revised Edition. New York : The Free Press.

COWIE, C. (2014). 'In Defence of Instrumentalism about Epistemic Normativity.' *Synthese* 191: 16, pp. 4003-40017

----(2019). *Morality and Epistemic Judgement: The Argument from Analogy*. Oxford: Oxford University Press.

CUNEO, T. (2007). *The Normative Web*. Oxford: Oxford University Press.

DANCY, J. (2004). 'On the Importance of Making Things Right'. *Ratio*, 17, pp. 229-237.

----(2005). 'Nonnaturalism' in Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.

DARWALL, S. (2017). 'Ethics and Morality' in T. McPherson & D. Plunkett, (eds.), *The Routledge Handbook of Metaethics*. New York: Routledge.

DUNAWAY, B. (2015). 'Supervenience Arguments and Normative Non-Naturalism'. *Philosophy and Phenomenological Research* 91, pp. 627-655.

DWORKIN, R. (2011). *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.

ENOCH, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

FINLAY, S. (2008). 'The Error in the Error Theory'. *Australasian Journal of Philosophy*, 86:3, pp. 347-369.

----(2014). *A Confusion of Tongues*. Oxford: Oxford University Press.

----(2009) 'The Obscurity of Internal Reasons'. *Philosopher's Imprint*, 9:1, pp. 1-22.

FIRTH, R. (1952). 'Ethical Absolutism and the Ideal Observer'. *Philosophy and Phenomenological Research*, 12, pp. 317-45.

FITZPATRICK, W. (2008). 'Robust Ethical Realism, Non-naturalism, and Normativity' in R. Shafer-Landeau (ed.) *Oxford Studies in Metaethics, Volume 3*. Oxford: Oxford University Press.

GARNER, R. (2007). 'Abolishing Morality'. *Ethical Theory and Moral Practice*, 10:5, pp. 499-513. Reprinted in Joyce and Kirchin (eds), *A World Without Values*, Springer Dordrecht, 2010.

GARNER, R. & JOYCE, R. (eds) (2019) *The End of Morality: Taking Moral Abolitionism Seriously*. Abingdon: Routledge.
GOLDMAN, A. (2002). 'The Unity of the Epistemic Virtues' in Haddock, Miller & Pritchard (eds) *Epistemic Value*, Oxford: Oxford University Press.
GREENE, J. (2002). *The Terrible, Horrible, No Good, Very Bad Truth about Morality and What to Do About it*. PhD Dissertation, Princeton University.
GRIM, S. R. (2009). 'Epistemic Normativity' in Haddock, Miller & Pritchard (eds) *Epistemic Value*, Oxford: Oxford University Press.
HARMAN, G. (1986). *The Nature of Morality*. Oxford: Oxford University Press.
HINKFUSS, I. (2019). 'To Hell With Morality' in Garner & Joyce 2019.
HINTZ, H. (1958). 'Some Further Reflections on Moral Responsibility' in S. Hook (ed.) *Determinism and Freedom in the Age of Modern Science*. New York: New York University Press.
HONDERICH, T. (1988). *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes.* Oxford: Oxford University Press.
HORGAN,T. & TIMMS, M. (2009). 'Analytical Moral Functionalism Meets Moral Twin Earth' in I. Ravenscroft (ed.) *Mind, Ethics and Conditionals*. Oxford: Oxford University Press.
HORWICH, P. (2006). 'The Value of Truth'. *Noûs*, 40: pp. 347-360.
ISSEROW, J. (2017). *What to do when the World Doesn't Play Along: Life After Moral Error Theory*. PhD Dissertation, Australian National University.
----(2019). 'Minimising the Misuse of Morality' in Garner & Joyce (2019).
JACKSON, F. (1992). 'Critical Notice of Susan Hurley, *Natural Reasons*'. *Australasian Journal of Philosophy* 70, pp. 475-488.
----(1994). 'Armchair Metaphysics' in Michaelis. M. & O'Leary-Hawthorne, J. (eds.), *Philosophy in Mind.* Dordrecht: Springer.
----(1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
JACKSON, F. & PETTIT, P. (1995). 'Moral Functionalism and Moral Motivation'. *Philosophical Quarterly*, 45, pp. 20-40.
----(1996). 'Moral Functionalism, Supervenience and Reductionism'.
JOYCE, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
----(2005). 'Moral Fictionalism' in M. Kalderon (ed.) *Fictionalism in Metaphysics*. Oxford: Oxford University Press.
----(2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.
----(2011) 'The Error in "The Error in The Error Theory"'. *Australasian Journal of Philosophy*, 89:3, pp. 519-534.
KELLY, T. (2003) 'Epistemic Rationality as Instrumental Rationality: A Critique'. *Philosophy and Phenomenological Research*, 66: 3, pp. 612-640.
KIM, J. (2003). *Supervenience and Mind*. Cambridge: Cambridge University Press.
KIRCHIN, S. (2017). *Thick Evaluation*. Oxford: Oxford University Press.
KITCHER, P. (2011). *The Ethical Project*. Harvard University Press.
KORNBLITH, H. (1993). 'Epistemic Normativity'. *Synthese*, 94, pp. 357-376.
----(2002). *Knowledge and Its Place in Nature*. New York: Oxford University Press.
KORSGAARD, C. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
----(1997). 'The Normativity of Instrumental Reason' in G. Cullity and B. Gaut (eds), *Ethics and Practical Reason*. Oxford: Oxford University Press.
----(2009) *Self-Constitution.* Oxford: Oxford University Press.
KRAMER, M. (2009). *Moral Realism as a Moral Doctrine*. Chichester: Wiley Blackwell.
----(2017) 'There's Nothing Quasi about Quasi Realism'. *The Journal of Ethics*, 21, pp. 185-212.
KVANVIG, J. (2003). *The Value of Knowledge and The Pursuit of Understanding*. Cambridge: Cambridge University Press.
LOEB, D. (1998). 'Moral Realism and the Argument from Disagreement'. *Philosophical Studies*, 90, pp. 281-303.
LEITE, A. (2007). 'Epistemic Instrumentalism and Reasons for Belief: A Reply to Tom Kelly's "Epistemic Rationality as Instrumental Rationality: A Critique"'. *Philosophy and Phenomenological Research*, 75:2, pp. 456-464.

LEITER, B. (2005). 'Nietzsche's Theory of the Will'. *Philosophical Topics* 33:2, pp. 119-137.
----(2014). 'Moral Skepticism and Moral Disagreement in Nietzsche'. *Oxford Studies in Metaethics*, Volume 9, pp.126-151
LILLEHAMMER, H. (2004). 'Moral Error Theory'. *Proceedings of the Aristotelian Society*, 94:1, pp. 95-111.
---(2013) 'A Distinction Without a Difference? Good Advice for Moral Error Theorists'. Ratio, 26, pp. 373-390.
LUHMANN, N. (2008). *Law as a Social System*. Oxford: Oxford University Press.
LUTZ. M. (2014). 'The Now What Problem for Moral Error Theory'. *Philosophical Studies*, 171, pp. 351-371.
LYCAN, W. G. (2013). 'On Two Main Themes in Gutting's What Philosophers Know'. *Southern Journal of Philosophy*, 51, pp. 112–120.
LYNCH, M. (2004). *True to Life : Why Truth Matters*. Cambridge, MA: MIT Press.
MAJORS, B. (2005). 'Moral Discourse and Descriptive Properties.' *Philosophical Quarterly* 55, pp. 475-494
MARKS, J. (2011). 'Moral Pornography'. *Philosophy Now*, 87, p. 52.
----(2013). *Ethics Without Morals*. New York, NY: Routledge.
----(2019). 'Beyond The Surf and Spray' in Garner & Joyce (2019).
MARQUIS, D. (1989) 'Why Abortion is Immoral' *Journal of Philosophy*, 86, pp. 183-202. Reprinted in Kuhse & Singer (eds), *Bioethics: An Anthology*, Oxford: Blackwell, 1999, pp. 46-57.
McCLUSKEY, C. (2017). *Thomas Aquinas on Moral Wrongdoing*. Cambridge: Cambridge University Press.
McGRATH, S. (2008). 'Moral Disagreement and Moral Expertise'. *Oxford Studies in Metaethics*, Volume 3, pp. 87-108.
McNAUGHTON, D. & RAWLING, P. (2003). 'Naturalism and Normativity.' *Proceedings of the Aristotelian Society, Supplementary Volume*, 77, pp. 23-45.
McPHERSON, T. (2009). 'Moral Arguments and Moral Revisionism'. *Journal of Ethics and Social Philosophy*, 3:2, pp. 1-25.
MACKIE, J.L. (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.
MOODY-ADAMS, M. (1997). *Fieldwork in Familiar Places*. Reprint, Cambridge, MA: Harvard University Press.
MILLGRAM, E. (1995) 'Was Hume a Humean?'. *Hume Studies*, 21:1, pp. 75-93.
MURPHY, J. G. (1988). 'Hatred: A Qualified Defence' in J.G Murphy & J Hampton (eds.) *Forgiveness and Mercy*. Cambridge: Cambridge University Press.
NOLAN, D. (2005). 'Fictionalist Attitudes About Fictionalist Matters' in M. Kalderon (ed.) *Fictionalism in Metaphysics*. Oxford: Oxford University Press.
NOLAN, D., RESTAL, G. & WEST, C., (2005). 'Moral Fictionalism Versus the Rest'. *Australasian Journal of Philosophy* 83: 3, pp. 307-330.
NOORDHOF, P. (1999) 'Moral Requirements are Still Not Rational Requirements'. *Analysis*, 59:3, pp. 127-136
OLSON, J. (2011a). 'Getting Real about Moral Fictionalism' in R. Shafer-Landeau (ed.) *Oxford Studies in Metaethics Volume 6*. Oxford: Oxford University Press.
----(2011b). 'Error Theory and Reasons for Belief' in Reisner and Steglich-Peterson (eds), *Reasons for Belief*, Cambridge: Cambridge University Press.
----(2014). *Moral Error Theory*. Oxford: Oxford University Press.
PAAKKUNAINEN, H. (2018). 'Doubts About "Genuinely Normative" Epistemic Reasons' in McHugh, C, Way, J. and Whiting, D. (eds) *Metaepistemology*, Oxford: Oxford University Press.
PAPINEAU, D. (2013). 'There Are No Norms of Belief' in Chan (ed.), *The Aim of Belief*, Oxford: Oxford University Press.
PEREBOOM, D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.
----(2009). 'Free Will, Love and Anger'. *Ideas Y Valores* 141, pp. 169-189.
----(2014) *Free Will, Agency and Meaning in Life*. Oxford: Oxford University Press.

PIGDEN, C. (2007). 'Nihilism, Nietzsche, and the Doppelganger Problem'. *Ethical Theory and Moral Practice*, 10:5, pp. 441-456. Reprinted in Joyce and Kirchin (eds), *A World Without Values*, Springer Dordrecht, 2010.
PLANTINGA, A. (2010). 'Naturalism, Theism, Obligation and Supervenience'. *Faith and Philosophy* 27, pp. 247-272.
PRINZ, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
RAILTON, P. (1986). 'Moral Realism'. *The Philosophical Review* 95:2, pp. 163-207.
ROBERTS, D. (2011). 'Shapelessness and the Thick'. *Ethics* 121: 3, pp. 489-520.
----(2013). 'It's Evaluation, Only Thicker' in S. Kirchin (ed.) *Thick Concepts*. Oxford: Oxford University Press.
ROWLAND, R. (2013). 'Moral Error Theory and the Argument from Epistemic Reasons'. *Journal of Ethics and Social Philosophy*, 7:1.
RYCHLAK, J. (1979). *Discovering Free Will and Personal Responsibility*. New York: Oxford University Press.
SCHAFER-LANDEAU. R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.
SCHROEDER, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.
SCHROETER, L. & SCHROETER, F. (2009). 'A Third Way in Metaethics'. *Nous* 36, pp. 249-275.
SHARADIN, N. (2015). 'A Partial Defence of Permissivism'. *Ratio* 28, pp. 57-71.
SMILANSKY, S. (1994). 'The Ethical Advantages of Hard Determinism'. *Philosophy and Phenomenological Research* 54:2, pp. 355-363.
----(2000). *Free Will and Illusion*. Oxford: Oxford University Press.
----(2005) 'Free Will and Respect for Persons'. *Midwest Studies in Philosophy* 29, pp. 248-261.
SMITH, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishers.
SOBEL, D. (1999). 'Do the Desires of Rational Agents Converge?'. *Analysis*, 59:3, pp. 137-147
STEGLICH-PETERSON, A. (2011). 'How to be a Teleologist About Epistemic Reasons' in A. Reisner & A. Steglich-Peterson (eds.), *Reasons for Belief*, Cambridge: Cambridge University Press.
STRAWSON, G. (1986). *Freedom and Belief*. Oxford: Oxford University Press.
----(1994) 'The Impossibility of Moral Responsibility'. *Philosophical Studies* 75: 1/2, pp. 5-24. Reprinted in G. Watson (ed.) *Free Will* (2nd edition), Oxford: Oxford University Press.
STREUMER, B. (2017). *Unbelievable Errors*. Oxford: Oxford University Press.
STURGEON, N. (1985). 'Moral Explanations' in D. Copp and D. Zimmerman (eds.), *Morality, Reason and Truth*, Totowa NJ: Rowman and Allenheld. Reprinted in G. Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press.
SUIKKANEN, J. (2010). 'Non-naturalism: The Jackson Challenge' in R. Shafer-Landeau (ed.), *Oxford Studies in Metaethics, Volume 5*. Oxford: Oxford University Press.
SVOBODA, T. (2015). 'Why Moral Error Theorists Should Become Revisionary Moral Expressivists'. *Journal of Moral Philosophy* 14:1, pp. 48-72.
VAN INWAGEN, P. (2006). *The Problem of Evil*. Oxford: Oxford University Press.
---(2008). *Metaphysics*. Boulder, CO: Westview Press.
VAN ROOJEN, M. (1996). 'Moral Functionalism and Moral Reductionism'. *Philosophical Quarterly* 46, pp. 77-81.
VAYRYNEN, P. (2013). *The Lewd, the Rude and the Nasty*. Oxford: Oxford University Press.
WEGNER, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
WILLIAMS, B. (1981). *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.
----(1985). *Ethics and the Limits of Philosophy*. Fontana Press. Reprint, Abingdon: Routledge, 2011.
----(1993) *Shame and Necessity*. Berkeley: University of California Press.
WILLIAMSON, T. (2001). 'Ethics, Supervenience and Ramsey Sentences.' *Philosophy and Phenomenological Research*, 62, pp. 625-630.
WOLF, S. (1981). 'The Importance of Free Will'. *Mind* 90: 359, pp. 386-405.
YABLO, S. (2000). 'Red, Bitter, Best'. *Philosophical Books* 41, pp. 13-23.
ZANGWILL, N. (2000). 'Against Analytic Moral Functionalism'. *Ratio* 13, pp. 183-203.