Moral Discourse: Error-Ridden or Relatively Defensible?

James Norton

Table of Contents

Introduction

When a particular discourse essentially revolves around something one believes not to exist, one will likely be an error theorist about this discourse. If the putative properties or objects of reference fail to be instantiated or to exist, the propositions expressed by the discourse will be systematically false. Hence, most people are error theorists about Santa Claus discourse[1], unicorn discourse and witchcraft discourse. If Santa has no place in one's critically considered ontology, then propositions such as "Santa has a big white beard and a red suit" will be false, despite being commonly uttered and accepted without debate. Thus, what motivates an error theory about a particular discourse is that something which that discourse requires in order to successfully refer is, in fact, not present in the world.

Richard Joyce (2001) argues that moral discourse warrants an error theory. Following John Mackie (1977), Joyce understands moral discourse as truth-apt, and intending to refer to objectively prescriptive values. Thus, unless we can discover what such values are like, and that they exist in the world, moral propositions are systematically false. Joyce argues that an objective moral prescription implies that there are some actions for which a valid imperative[2] could be directed at any agent. Thus, in order for a moral assertion to be true, there must be some action for which a 'real' reason can be given to any individual to comply. The task of Chapter One is to explore the competing conceptions of what it takes for an agent to have a real reason, and to consider whether any of these could underwrite a universally valid imperative. I find that real reasons are necessarily rooted in the desires of the agent, and that these desires are divergent enough to ensure that no imperative is universally valid. Thus the error theory is entailed if one holds an objectively prescriptive conception of moral discourse.

---

[1] While there is room for a non-cognitivist analysis of Santa Claus discourse, or perhaps to suppose that Santa Claus does exist, but is an abstract object that comes into existence as a result of the Santa fiction, there is a good case to be made for an error theory.

[2] Joyce writes that "Mackie uses the word "valid" [...], I think, as a quality that stands to imperatives as truth stands to propositions." (2011:3, footnote 4).

Chapter Two investigates a response by Stephen Finlay (2008), who argues that moral assertions are not intended to refer to objectively prescriptive values, but instead to the standards and ends of the speaker. Thus, a moral imperative is valid if the prescribed action facilitates the realisation of the speaker's ends, and many moral assertions are true. Hence, Finlay's relational understanding of moral discourse escapes the threat of error theory. However, there remains controversy about whether this is a legitimate way to understand moral concepts. Indeed, Joyce (2011) insists that objectively prescriptive values are a necessary feature of our moral concepts. I analyse Finlay's re-interpretation of the behaviours surrounding moral discourse in the light of his theory, finding his explanation less coherent than one which assumes objectivity, but plausible nonetheless.

Finally, I evaluate whether Finlay's relational moral discourse can pragmatically 'pull its weight' and perform the same functions as the objectively prescriptive moral discourse. I find that relational moral imperatives often fail to provide authoritative reasons for action to the person being addressed. However, I see no principled way of deciding whether such use is a non-negotiable element of moral discourse. Thus, I conclude that this dispute remains underdetermined: what to make of it depends upon one's intuitions about the requisite strength of moral imperatives.

Chapter One: Objective Prescriptivity and Error

1.0 Introduction: Mackie

John Mackie's (1977) seminal book *Ethics: Inventing Right and Wrong* challenged users of moral discourse to review precisely what such discourse entails. He further questioned whether belief in moral properties is compatible with the way the rest of the world is construed; whether it has a place in a critically considered analysis of the world. Mackie denied a non-cognitive analysis of moral discourse, believing moral claims to be assertoric and truth-apt. However, Mackie also denied the realist thesis that there are moral claims that are true[3]. This is because moral claims imply the existence of values that are 'objectively prescriptive' and such entities are not to be found. Thus, he concluded that whilst moral claims 'aim' at truth, they are in reality systematically false. Hence, moral discourse is hopelessly committed to asserting claims that, by their nature, necessarily fail to refer. This chapter will examine the nature of objective prescriptivity (following in the footsteps of Richard Joyce's 2001 book, *The Myth of Morality*) and evaluate whether moral discourse, if committed to this thesis, is indeed fundamentally flawed. The following chapter will re-evaluate whether objective prescriptivity is a necessary feature of moral discourse (in the light of a 2008 paper by Stephen Finlay) and thus whether there is a call for a moral error theory.

1.1 Imperatives and Institutions: Kant and Foot

Imagine a murderer whose guilt is not in question. Furthermore, imagine that this murderer believes he has done as he should in every dimension of the act and is entirely

---

[3] Specifically, claims of moral requirement (e.g., you morally ought Ø) are never true. However, the denial of such a claim (it is false that you morally ought Ø) is always true, if vacuously so. Thus, as helpfully observed by Joyce (2001) it is only *atomic* moral claims that are necessarily false.

unrepentant. This is not merely a case where the murderer stands to gain something from his action which *outweighs* whatever loss he suffers: this is an unusual circumstance in which *every* outcome appears to satisfy his desires. Perhaps he hated his victim and everybody who cares about him. Perhaps he gains a special, guilt-free thrill from killing. Perhaps he wanted to be caught, even punished. Such a case is highly unusual, but surely not impossible. Nevertheless, despite the absence of a consequent not desired by the murderer, the *moral* injunction that he ought not to have killed his victim remains steadfast. No matter how poorly he judges it would have served his desires, morally speaking he still ought to have refrained. To give some credence to this story and agree that the moral imperative not to kill would apply to such an agent is to admit that moral imperatives apply to agents regardless of their subjective desires and ends; moral judgements are *objective*. This chapter argues that despite the purported desire-transcendence of moral imperatives, *all valid imperatives* provide reasons for action, and *real* reasons for action are those somehow rooted in desires[4]. Thus, to vindicate moral discourse it must be established that, on some level, the unrepentant murderer is frustrating his own desires.

The insistence that moral requirements are imposed upon agents regardless of their desires can be traced back to Kant's explication of the categorically imperative nature of morality. Thus, to understand what it means for morality to be objective in this way it is useful to discuss the nature of hypothetical and categorical imperatives. Hypothetical imperatives are those which ought to be done in order to attain satisfaction of an agent's desires. In Kant's words, "if the action would be good merely as a means to something else the imperative is hypothetical" (1785/1997: 4:414). For example, "If you want to feel better then you should see a doctor this afternoon" is a hypothetical imperative. As such, a hypothetical imperative functions as practical advice in order to achieve a goal. However, if

---

[4] This claim omits the possibility of externalism about reasons, which shall be dealt with in **1.3**. Furthermore, there are various ways to understand the connection between desires and reasons.

our agent suddenly rallies and is feeling fine, the applicability of the hypothetical imperative to go see the doctor dissolves, or 'evaporates', as the reason to comply is no longer present. Further, if you were to discover that this afternoon there is a football match on which the agent would much prefer to watch than to regain his health, the same evaporation occurs. Thus, hypothetical imperatives are highly susceptible to changes in the constitution of the agent and to his peculiar preferences.

Kant emphatically argued that moral prescriptions are *not* like hypothetical imperatives. They are *applied* to the agent in virtue of the circumstances, and an overwhelming desire to perform an action forbidden by morality will not make the moral requirement to abstain evaporate. Thus, moral prescriptions are categorical in nature. In Kant's words, "the categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end" (1785/1997: 4:414). This objective necessity is why Philippa Foot (1972[5]) describes the categorical understanding of morality as 'inescapable'. Hence, if Øing in circumstances C is morally required, Øing *must be done by any agent* who finds himself in C, despite any unusual desires he may have. However, morality is not the only institution which places demands upon agents regardless of their desires. *Games*, for example, require that everyone plays by the rules despite potential desires satisfied by cheating.

Following Foot, we shall now turn our attention to the institution of etiquette. Foot famously observed that the behavioural imperatives of etiquette share the 'non-evaporability' of morality. If it is true according to the regulations of etiquette that "one ought not speak with one's mouth full," this demand applies equally to an agent who *cares* about etiquette as to one who does not. It applies equally to an agent who is *aware* of the demands of etiquette as to one who is not. Significantly, it applies to an agent with a strong

---

[5] All further references to Foot are also to this paper.

desire to flaunt the demands of etiquette, as the unrepentant murderer has to moral requirement. For example, if one were in a position to win a large bet by annoying a friend over lunch, the desire to win the bet would provide strong personal reasons to talk with one's mouth full, thereby acting contrary to the requirements of etiquette. Nevertheless, the imperative of etiquette still applies, and as such the *institutional* requirement to refrain from speaking with one's mouth full remains steadfast despite the unusual desires caused by the bet. Likely a person who is dedicated to, or immersed in, the institution of etiquette (say, an etiquette teacher) would still say that one ought not speak with one's mouth full, despite the bet. However, one entirely uncommitted to etiquette might say there is in fact *no reason* to conform to its institutional requirements.

Despite the above considerations, I suspect the reader, in lieu of unusually strong allegiance to etiquette, would advise winning the bet over acceding to the rules of etiquette. This is in stark contrast with a potential equivalent case where one serves to win a large bet by torturing a baby. The categorical imperative that one ought not speak with one's mouth full did not evaporate, but was *outweighed* by the bet[6]. The categorical moral imperative that one ought not torture babies also does not evaporate, but moreover is not (*and cannot be*) outweighed by the bet, nor by other desires. Thus, Foot is uneasy that "we find 'should' used non-hypothetically in some non-moral statements to which no one attributes the special dignity and necessity conveyed by the description 'categorical imperative'." (1972:308). Richard Joyce, the philosopher who has most enthusiastically endorsed and expanded upon Mackie's moral error theory, derives from this that there is something special about moral imperatives that goes beyond their categorical nature (2001). Therefore, categorical imperativeness is a necessary but not sufficient characteristic of moral judgement. Moral prescriptions don't merely *apply*; they *demand*. Therefore we must

---

[6] Note, however, that to an absolute non-subscriber of etiquette, the imperative did not even provide a reason that needed to be outweighed.

differentiate weak and strong categorical imperatives. A weak categorical imperative is exemplified by the requirements of etiquette, and (whilst from the perspective of the institution never evaporating) can be outweighed by other desires to the extent that it presents no reason for action[7]. However, moral discourse purports to consist of *strong* categorical imperatives that always provide reasons. Hence it is essential to identify the 'extra ingredient' that differentiates the strong categorical imperatives of morality from the weak imperatives of etiquette, chess and gladiatorial combat[8].

Foot concludes that no such 'fugitive thought' is to be found - that the imperatives of morality are after all merely institutional, weak categorical imperatives (Joyce's terminology, not hers). She accuses Kant and his followers of artificially imbuing moral imperatives with a "magic force" (1972:315). Thus, Foot accepts that a moral imperative is only authoritative to one for whom the imperative provides reasons. Joyce agrees that all *valid* imperatives have this characteristic, but maintains that what it takes for an imperative to be *moral* is, *inter alia*, universally applicable authority: providing reasons *to any agent*. Thus, for Joyce, the troubling nature of the elusive fugitive thought is not only an issue for the deontological moral conception, but a challenge for moral discourse as a whole. Therefore, on behalf of the moral realist Joyce seeks to make some sense of strong, non-institutional categorical imperatives.

1.2 Objective Prescriptivity

Thus far, we have established that what differentiates the strong categorical imperatives of moral discourse from the weak categorical imperatives of etiquette is that moral imperatives are authoritative for any agent, and thus *objectively prescriptive*. Joyce's

---

[7] Say, the imperatives of halal meat preparation for a Buddhist.
[8] For useful explication of weak categorical imperatives, see the story of Celadus the unwilling gladiator in Chapter 2 of Joyce (2001).

(2001) preferred method of explicating this authority is through the provision of *reasons*. As such, an imperative is authoritative only if it provides a reason for action. Therefore, Joyce presents what he calls 'Mackie's Platitude': *"It is necessary and a priori that, for an agent x, if x ought to Ø, then x has a reason to Ø"* (2001:38). As such, the term 'ought' is taken to *mean* 'has a reason'. Mackie's platitude is supported by the common sense understanding of the need to extrapolate reasons from a valid 'ought'. Say one tells an agent he ought Ø, and his response is "why?". If no answer can be found for this fair request for reasons, there is no rational basis for the ought.

Of course, moral discourse can be construed so as not to imply objective prescriptivity. One such attempt is explored in detail in Chapter Two. However, Joyce argues this is not how competent users of the term *moral* would conceive and use it. Joyce seeks to "motivate a sense of unease" about the possibility that morality might *not* be objectively prescriptive (2001)[9]. Consider again the unrepentant murderer. If through his unusual constitution he truly has *no reason* to refrain, what sense can be made of demanding he ought? Yet moral discourse does not hesitate to make this demand. Joyce maintains that it is unacceptable that such use of moral imperatives might turn out to be insubstantial in this way. Rather, moral imperatives are deemed to be backed by the force of reason, and thus have authority for any agent.

However, it is essential to note that characterising moral judgements in this way need not imply that a moral prescription will necessarily *force* an agent to act morally. Empirical observation quickly demonstrates that, no matter how one conceives morality, there are agents who act contrary to its prescriptions. This occurs even in agents who care a lot about morality and make sincere moral judgements themselves. Other desires can be so powerful

---

[9] Joyce later clarifies that the idea behind objective prescriptivity is difficult to express, and best captured by "an extremely inchoate presupposition, using a term like "practical oomph" […] and possibly providing a bunch of examples […] in order to try and capture the "feel" of moral authority." (2011:6)

as to *outweigh* the desire satisfied by behaving morally[10]. *Therefore, to make sense of objective prescriptions is not to search for actions which everybody (in fact) performs*. Rather, the search is for actions which any given agent *ought* (or *ought not*) do when in the relevant circumstances. As agents often fail to do what they ought, all that is required is that the moral imperative provides a valid reason for any agent. An ought without an accompanying reason quickly begins to look lonely and irrational. Note, however, that there are various theories regarding what is needed in order to make a reason 'valid', which shall be explored throughout the rest of this chapter.

Nevertheless, if certain circumstances indicate reasons for a particular action to be performed by any agent, the accompanying imperative looks like a candidate for a *true, objectively prescriptive* moral claim. If such a circumstance and its accompanying imperative can be found, the central premise of the error theory that all moral claims are false is itself falsified, and the error theory is sunk. Note, however, that the class of agents for whom a valid moral imperative must give reasons includes all agents we consider bound by morality in all possible worlds. Therefore, even if all *actual* agents have a reason to comply, if this is not so for a *non-actual* but *possible* agent, the imperative is not universally valid. Moreover, if there are possible agents whose constitution ensures that *no* imperatives are universally valid, the error theory is necessarily true and moral discourse fails to refer in every possible world.

Furthermore, such an imperative would need to coherently fit within a folk conception of *moral* requirement. If, for example, it could be found that every agent actually has a reason to clean their teeth after eating, this would be at best an objectively prescriptive imperative of *hygiene*. In Joyce's words: "a theory of imperatives that managed to supply strong categorical imperatives - that located Foot's "fugitive thought" - but for

---

[10] Assuming there is such a desire; see the next section.

things like "Kill anyone who annoys you", "Steal when you can," etc., simply would not be a *morality*." (2001:67). Nevertheless, this appears to be a justifiable route through which to investigate the presuppositions underlying moral discourse. Thus, we shall explore the nature of reasons for action.

1.3 Internal and External Reasons

There are various competing accounts regarding what it takes for an agent to have a 'valid' or 'real' reason for action. The stance one takes towards these views will determine the nature of the project to find universally authoritative imperatives. Thus we shall briefly examine the various kinds of reasons. The first distinction is between subjective and objective reasons. Imagine the case of Wilfred, an agent who is currently very thirsty. There is a cup in front of Wilfred which he believes to contain water, and insofar as he is thirsty, he has a *subjective* reason to drink the water. However, unbeknownst to Wilfred the cup actually contains poison, and thus he has an *objective* reason to refrain from drinking. Hence, the subjectivity or objectivity of reasons is dependent upon how well informed the agent in question is. However, an imperative to Wilfred not to drink validly connects to his desires despite his ignorance. Thus, we must distinguish *internal* and *external* reasons as per Bernard Williams (1981). Williams explains that an internal reason is one *appropriately connected* to the agent's desires, which *could potentially* come to motivate the agent, were he to be perfectly rational and armed with true beliefs. An external reason, however, fails to have such a potential connection and thus applies regardless of the agent's desires. Hence, Wilfred has an *internal objective* reason not to drink by virtue of his *de dicto* desire not to drink poison.

There is a strong intuition that one is only rational to the extent that one is guided by subjective internal reasons. Thus, it would in fact be *irrational* (though unwittingly fortunate)

for Wilfred not to drink from the cup. However, if he were to become aware of the actual contents of the cup he would admit that there was a reason not to drink the whole time, and furthermore *no reason to drink*. This helpfully illustrates a potential disconnect between what it is *rational* to do and what one has a *reason* to do, depending on one's understanding of reasons and rationality[11].

Moreover, some philosophers such as David Brink (1989) have argued that, in fact, *externalism* about reasons is the best framework within which to understand moral prescriptions, as expounded by Cornell realism. As such, moral reasons for action are completely divorced from the psychology of the agents in question[12]. Brink's externalist moral realism ardently endorses conceiving morality as objectively prescriptive. Indeed, the notion that moral imperatives are not dependent upon agents' desires is easily explicable if reasons are separate from desires. However, this is at the expense of alienating the agent from his own reasons. If my 'real' reasons for action are, in fact, completely unrelated to my desires, what sense can be made of saying that they are *my* reasons? There is a real debate to be had here, though beyond the scope of this thesis. I will follow Joyce's (and Williams') example in rejecting external reasons as a valid way to authoritatively connect imperatives to agents.

Therefore, our search is now restricted to the realm of internal reasons: those which can be coherently connected to the desires of the agents in question. However, there are various ways of conceiving the precise nature of this connection. These views can be understood as different incarnations of the institution of *practical rationality*, the mechanism by which all rational agents evaluate reasons for action, and deliberate what to do.

---

[11] For more on this dichotomy see John Broome's (2007) *Does Rationality Consist In Responding Correctly To Reasons?* and Nicholas Southwood's (2008) *Vindicating The Normativity Of Rationality*.
[12] The Kantian (1785/1997) understanding of morality as an appeal to rational consistency also divorces moral reasons from an agent's desires.

1.4 Practical Rationality

Joyce (2001) has argued that the reasons alluded to by institutional categorical imperatives suffer from inapplicability if the agent in question cares little for the institution. When confronted with a demand of etiquette, one may reasonably ask "So what?" After all, why care about what etiquette requires if you do not subscribe to it as an institution? Why care more about etiquette than (say) kosher? However, the moral institution uniquely[13] purports to *require* subscription. When the unrepentant murderer breaks a moral prohibition, we say that he is (on some level) breaking a rule that is *his rule*. Thus, in order to justify this intuition, the challenge is to find an institution to which everybody (if not consciously, then tacitly) subscribes. Such an institution could potentially underwrite moral prescriptions through providing reasons connected in some way to every agent's desires and justifying the presupposition of objective prescriptivity.

Joyce posits that an institution which could underwrite moral imperatives by providing universally applicable reasons may find its roots in rational prescriptions. Thus, he alights upon the institution of practical rationality. Cleverly, Joyce notes that one who questions practical rationality externally (as one might question other institutions such as etiquette) has unwittingly demonstrated his allegiance to practical rationality itself. "Why should I care about practical rationality?" is a question asking for reasons to adopt a system of evaluating reasons, but to ask for such a reason implies that one is already in the business of evaluating reasons, and thus already subscribed to practical rationality. While an agent who never asks this question will not fall into Joyce's 'trap', the addition of idealised rationality which we shall see in Michael Smith's non-Humean instrumentalism neatly ensures subscription to practical rationality.

---

[13] Perhaps uniquely only in the practical realm. The epistemic imperatives of theoretical rationality, such as "Don't believe in things unless you have some evidence that they exist" also appear to require subscription.

Notably, Foot has argued that such a project is doomed from the outset, as irrationality necessarily involves frustrating one's own ends and immorality need not involve this. Nevertheless, having established practical rationality as an 'un-so-what-able' institution, let us examine whether it might somehow vindicate objectively prescriptive moral judgements. However, as hinted earlier, the precise nature of practical rationality remains debatable. Specifically, which reasons count as *validly* connected to the desires of the agent hinges upon how this institution is construed. Following Joyce, we shall examine instrumentalism as per Hume, followed by Michael Smith's broader account of instrumental reasons.

1.5 Humean Instrumentalism

The Humean (1739/1978) understanding of practical rationality derives reasons directly from the desires of the agent. As such, what one rationally ought do is what one most desires. In terms of reasons, the Humean thesis is that *if one desires X, and Øing is the best way to make X come about, one has a reason to Ø*. Thus an agent has a reason to perform any action which best furthers any of his desires. This implies the possibility of having real reasons to perform a variety of different actions in the same circumstances, which is intuitively the case. Moreover, these desires could correspond with actions which promote dramatically different outcomes. However, as we have seen, objective prescriptions only require the existence of a reason to act morally that can be validly connected to any individual's desires in order to be vindicated.

Within the Humean framework it is possible that a moral imperative could turn out to be true if every agent has a desire which is satisfied by the action morally prescribed, thus providing a subjective internal reason. For example, it could be that refraining from murder

satisfies some desire present in every agent[14]. Such desires could be scattered and varied, but if every agent has his own reason the objective prescription is justified. However, to investigate such a possibility is an empirical nightmare, and Joyce (2000, 2001) is pessimistic about the potential of this project. In any case, such a state of affairs could never be discovered from an armchair. However, perhaps there is a single, ubiquitous (perhaps *necessary*) desire that could ground the truth of a moral prescription. An intuitive place to search for a desire held by every agent is in the domain of self-interest. Surely every individual has reason to further his own ends. Thus, the project here is to ground an apparently selfless moral prescription such as "Don't steal" in a sophisticated self-interest.

Whilst this Hobbesian (1651/1924) moral conception has every reason to succeed when dealing with normally constituted agents, there is reason to doubt its applicability in certain cases. Initially, note that there are often *prudential* reasons to behave morally. If you are known as a liar, the community will not trust you. Thus the desire to be trusted is frustrated. If you steal and are caught, you will be imprisoned, frustrating the desire for freedom. Conversely, if you are generous, you will be rewarded and liked. However, despite this robust correlation between prudence and morally good action, Joyce argues that any attempt to ground reasons to behave morally in prudential self-interest will fail. This is because reasons to act morally must be universal, and thus validly connect to the desires of unusual individuals. While prudence might provide the *majority* of agents with reason to act morally in *most* situations, if the philosopher is allowed free rein to manipulate the details of the agent and the situation, the prescriptions of prudence and morality drift apart.

Joyce reminds us that this project makes little headway against Gyges, the Lydian shepherd from Plato's *Republic* (Cooper, 1997). Gyges stumbles upon a ring of invisibility, which he uses to fulfil his basest and most immoral desires anonymously *because of his self-*

---

[14] Perhaps a desire to be benevolent in some people, a desire to remain unpunished in others, etc.

interest! For any prudential reason to refrain that one cares to offer Gyges, the philosopher can tweak his desires and circumstances to make it inapplicable. Perhaps Gyges' rampant raping would foster a sense of fear and unease in the community which frustrates his desire to have happy and pleasant friends. However, if he exclusively inflicts harm in some nearby village, this desire need not be frustrated. Alternatively, perhaps Gyges is harming himself through his actions, sacrificing a desire for "inward peace of mind, consciousness of integrity, a satisfactory review of [his] own conduct" (Hume, 1983, quoted in Joyce 2001:33). To justify this claim, the onus is on the defender of morality rooted in Humean instrumentalism to argue that the nature of humanity is such that one who acts as Gyges does *necessarily* sacrifices the characteristics listed by Hume, which Joyce describes as "simply silly" (2000:6). To argue that human nature exhibits such homogeneity is a difficult undertaking; it is entirely possible that some people are simply depraved.

On Hume's view, there are two ways to be practically irrational, or fail to act in accordance with one's own practical reasons. Firstly, one can have a desire based on a false belief. For example, if you are convinced that doing a headstand for a full day (despite extreme discomfort) will bring you wealth that you greatly desire, you will deem yourself to have a reason to put yourself through a painful day of inversion. However, if the headstand is in actuality unrelated to wealth, Hume will say that in reality you have no reason to do so. Nevertheless, strictly speaking the irrationality is at the level of the false *belief*, not the action. Furthermore, as soon as the belief is corrected, the desire will evaporate[15]. Secondly, one can have a false belief about how to satisfy a desire, poorly weigh desires against each other, or perform an irrational action through weakness of will. These all revolve around

---

[15] "Since a passion can never, in any sense, be call'd unreasonable, but when founded on a false supposition, or when it chuses means insufficient for the design'd end, 'tis impossible, that reason and passion can ever oppose each other, or dispute for the government of the will and actions. The moment we perceive the falshood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition." (*Treatise* 2.3.3)

reflecting upon the rationality of one's action and finding it wanting. However, perhaps it is regret itself that is irrational, or perhaps one *must* have acted on one's strongest desire at the time (particularly if the strength of desires is determined by which ones lead to action). Thus, while most of Hume's explication of practical rationality reads as a *description* of practical deliberation, it is worth noting that there is room for rational error on this view.

However, the trouble here is that Gyges' base deeds appear to be perfectly aligned with his Humean reasons. The Humean understanding of the relationship between desires and reasons does not provide Gyges with any reason to refrain: it is quite possible that he has no false beliefs and is being practically rational. Nevertheless, his actions do not evade our moral judgement. What he did was morally wrong, and Joyce wants to 'motivate a sense of unease' that while we make a demand of Gyges that he ought not do these things, we are not supplying him with a reason to do as we claim he ought. When Gyges asks why he ought not do something he has no desire (or reason) to refrain from doing, if all we can say is "you simply mustn't!", it is Gyges who appears to be the reasonable one, and the repeated imperative begins to sound rather desperate. The defender of Humean instrumentalism can choose to concede that on his view one cannot validly claim that Gyges ought not commit his crimes. However, to do so is to admit the failure of the project to ground objectively prescriptive moral imperatives. As moral language is precisely that with which we would, as a matter of fact, censure Gyges, that it might turn out not to apply to him is unacceptable. Joyce summarises the situation as follows: "If a system of imperatives leaves the harmful wielder of the ring of Gyges uncensured as he wreaks havoc then it is not a moral system."(2000:6).

Therefore, if morality is underpinned by practical rationality and we endorse Humean instrumentalism we end up with an inescapable set of escapable propositions. This is because (as cleverly shown by Joyce) the institution of practical rationality is inescapable

(un-so-what-able), though any given prescription is escapable through citing particular unusual desires. This is an unsatisfactory result, as we have apparently demonstrated the inescapability of the moral institution without allowing it any specific, inescapable (and thus true) claims. Therefore, the exploration of Humean instrumentalism appears to be a dead end. In search for a more universal, convergent understanding of 'real' reasons we now turn to Michael Smith.

1.6 Michael Smith's Non-Humean Instrumentalism

It has been demonstrated that Humean instrumentalism fails to justify objective prescriptions. Hume and Foot are content to accept that moral talk simply fails to authoritatively address Gyges. However, Joyce is dissatisfied with this outcome; by his lights such failure indicates systematic error in moral discourse. Thus, to charitably bolster the case of the moral realist, we seek a theory of practical rationality which presents an alternative connection between desires and reasons, hopefully establishing some reasons which would speak authoritatively to Gyges.

In order to do so, Joyce turns to Michael Smith's (1994) account of practical rationality. Recall the case of Wilfred. Smith intends to capture the intuition that Wilfred has a real reason not to drink from his cup, and extends the domain of valid reasons to include all internal reasons (subjective and objective), minus those rooted in irrationality. Thus, Smith's understanding of instrumentalism removes epistemological inadequacy and poor evaluation of one's own desires from the understanding of practical rationality. As such, Smith would have us imagine, for every agent (S), an ideally placed equivalent of that agent (S+). Significantly, S+ is *not* circularly idealised in terms of morality. Instead, S+ retains all of the subjective desires and ends of S, but is also fully rational and has complete epistemic access to the state of the world and the consequences of his actions. Hence S+ is considers

all internal reasons, including the objective reasons S is unaware of. Moreover, S can be considered to have a normative reason to Ø by virtue of S+ 'recommending' Øing. Consider the following claim: "I accept that if I were entirely rational and fully aware of the state of the world I would have a reason to Ø, yet I do not have a reason to Ø". This declaration is unintelligible, as to be more rational and informed is to be in a *better* position to judge what actions one has a reason to do. In the case of Wilfred, his S+ would have recommended he not drink the poison, and pointed out that he has no reason to drink it. If our reasons to conform to a moral imperative are akin to Wilfred's reason not to drink, they are valid. Indeed, if each time one (for example) tortures a baby one is (though unaware of the fact) subtly 'poisoning' oneself, there is a real reason to refrain.

Therefore, the pertinent question becomes whether, stripped of irrationality and ignorance, there might be some circumstances in which every agent would have a reason to perform a given moral imperative. In other words, this is an analysis of whether the set of each agent's S+ would converge upon any prescriptions. If there is convergence (as Smith believes), we are approaching a valid objectively prescriptive claim, and vindicating objective moral realism in virtue of the existence of a *true* moral prescription. If some moral claims are true (even if very few) then the discourse does not fail to refer, and in fact we have learned something: there are very few moral truths. However, if the idealised agents remain divergent in the way S's are this argument has no traction against the error theory.

Joyce (2001) encourages us to imagine an agent (a la Gyges) who breaks a promise merely on the basis of inconvenience and the fact that he can avoid punishment, regardless of the detrimental impact it will have on others. Moral discourse does not hesitate to condemn this action. However, the question remains as to whether S+ would, in fact, advise S not to break the promise, despite a lack of negative consequences. In order to argue that S+ would provide a reason for S to refrain from promise-breaking it needs to be shown

precisely how S is necessarily either (a) failing to be rational, or (b) ignorant of some relevant fact, or both. Why should this be the case? In Foot's words, "The fact is that the man who rejects morality because he sees no reason to obey its rules can be convicted of villainy but not of inconsistency" (1972:310). S+ derives his desires directly from S, and is thus entirely dependent upon S's constitution. Therefore, it is intuitive to expect a similar kind of variation amongst ideal selves as we observe amongst agents. However, this is at heart an empirical question which remains unanswered.

This debate remains very much open and live. The objective moral realist is free to mount an argument about why each agent's S+ would, in fact, converge upon some moral prescription. Alternately, he could propose a different formulation of practical rationality, or another institution altogether from which to underwrite strong categorical imperatives. However, such discussion is beyond the scope of this thesis. Thus we shall abandon the search for objectively prescriptive reasons here, having elucidated the kind of values the error theorist thinks underpin moral discourse, and what it would take for them to exist.

1.7 Summary and Preview

In order to vindicate moral realism, valid objective prescriptions must be found, or their dispensability demonstrated. I conclude that this attempt to derive objectively prescriptive values from practical rationality is a failure. However, this does not necessarily preclude the possibility that such values could be rooted elsewhere. Nevertheless, until a more appealing strategy for finding objective values presents itself, an error theory of morality is called for. However, some philosophers have denied that objective prescriptivity is a necessary feature of moral discourse, which of course emancipates morality from the threat of this argument for error theory. One such philosopher is Stephen Finlay, and his objection to Mackie and Joyce is discussed in the following chapter.

Chapter Two: The Dialectic between Joyce and Finlay

2.0 Introduction

It has been shown that Richard Joyce's (2001) defence of the moral error theory can be understood as a two-step argument: firstly identifying the conceptual presuppositions of moral discourse, and secondly arguing that these commitments are philosophically indefensible. Thus, there are two different methods through which an opponent might criticise the error theory. The first, which Joyce dubs the 'head-on' strategy, is to disagree with the error theorists' ontological claims. Such an opponent accepts the error theorists' rich, objectively prescriptive conceptualisation of morality but promotes optimism about the existence of these values. Hence this strategy might be adopted by the objective moral realist. If there are real reasons underpinning universally valid imperatives, some moral claims are true. Perhaps such an opponent would criticise the rigour of Joyce's attempt to ground these imperatives in practical rationality. Alternatively she might mount her own attempt to appeal to another institution in order to underwrite objective prescriptivity. There are many ways such an argument might be formulated, and Joyce views this tack as the "only hope" for the opponent of error theory.

The second, 'concessive' strategy, accepts the error theorists' sparse ontology, endorsing the ontological denial of objectively prescriptive values. Instead, this opponent attacks the presupposition that these values must exist to validate moral discourse. If such values are, in fact, not integral to morality, the apparent failure of practical rationality to provide universally authoritative reasons as per Chapter One is insignificant. Stephen Finlay notably[16] utilises the concessive strategy in *The Error in the Error Theory* (2008), attempting

---

[16] In fact, Finlay's paper won the 2008 Australasian Journal of Philosophy Best Paper Award.

to neutralise the error theorists' project before it gets off the ground. This chapter evaluates Finlay's criticisms, Joyce's (2011) rebuttal, and Finlay's (2011) re-rebuttal.

2.1 Relativistic Moral Discourse: A Legitimate Threat to Error Theory

Mackie (1977) denied the existence of objective values. However, he never went so far as to deny that there are values at all. Rather, he argued that all *genuine* value is relative to some standard or end[17]. As such, if refraining from speaking with one's mouth full is a value dictated by etiquette, one should refrain according to (say) the end of not disgusting one's tablemates. If you think this end is valuable, you have a reason to refrain. Hence, rightness-according-to-an-institution and rightness-relative-to-an-end notably share the feature of escapability. That is, if one does not subscribe to the institution or end, one will not have a reason to act as the standard requires. Hence, imperatives about what someone ought do, (construed as appeals to real reasons that validly connect to the agent's desires) will fail to be universally applicable, as humanity is far from homogenous in terms of subscription to institutions and ends.

Section **1.1** intended to motivate a sense of unease that moral imperatives might be merely institutional, or contingent upon subscription to certain ends. If the reason-bringing power of morality is limited in this way then there is no difference in authority between (say) telling Gyges that morality demands he stop raping and telling him that the laws of kosher demand that he not eat ham afterwards. Nevertheless, Stephen Finlay (2008) argues that *even moral value is relative to standards or ends*, and is thus metaphysically defensible[18]. As such, moral imperatives imply the existence of authoritative reasons only for

---

[17] A judgement that something is *big*, for example, is always relative to a comparison class. Mount Everest is big relative to me, but not relative to the sun.

[18] In this respect, Finlay's paper can be understood as a rejoinder to Joyce (2001) on behalf of Foot (1972), who contended that it was only Kant and his followers who imbued moral discourse with false objectivity.

agents who share the standard on which the imperative is based. Joyce and Finlay agree on this understanding of the limited reason-bringing power of imperatives. However, Joyce insists that a genuinely *moral* imperative must provide authoritative reasons to *any agent*, and thus avers an error theory. In contrast, Finlay argues that relativistic moral discourse is not only plausible, but represents moral thought and speech more realistically than an objective conception.

Finlay's denial that moral discourse presupposes objective values targets the incarnation of Mackie's argument that Joyce (2001) has identified and expanded upon as the most promising avenue through which to argue for error. Nevertheless, arguments towards error theory can be launched from multiple starting points. Joyce alludes to various other potentially flawed presuppositions of moral discourse: "It might be that moral discourse presupposes a view of human autonomy which we discover through empirical methods to be flawed [see, e.g., Libet 2004]; perhaps it presupposes a view of human character traits that we discover through experimental psychology to be wrong [see Doris 2002]; or perhaps it presupposes the truth of a kind of internalism that neuroscience and psychopathology reveal to be mistaken [see Roskies 2003]."(2011:2). Joyce criticises Finlay's failure to locate and rebut other flawed presuppositions as a "major strategic weakness". However, Finlay (2011) legitimately responds that by these standards *any* journal-length attack upon the error theory will inevitably be inadequate. Furthermore, Finlay's (2008) paper explicitly admits that the error theorist can viably posit other flawed presuppositions. As arguments from objective prescriptivity to error underwrite most of Joyce's publications on the topic (2000, 2001, 2007) if Finlay can demonstrate its dispensability to moral discourse, Joyce's argument requires a complete overhaul. Therefore, Finlay's paper remains a legitimate threat to moral error theory through attacking its most popular and effective argument.

2.2 Locating Finlay's Objection within Joyce's Master Argument

There is a terminological dispute that runs through this argument which requires clarification. For the sake of brevity I shall skip any petty squabbles about 'correct' terminology or what Mackie did or did not say and briefly sketch how the terms are used here. Finlay refers to '*absolute* values' and '*absolute* authority', whilst Joyce prefers '*objective* values' and '*objective* prescriptivity'. I shall take them to be talking about the same thing, though primarily utilising Joyce's terminology here. Whilst lamenting Finlay's choice of the term 'absolute', Joyce points out that within metaethics, absolutism stands opposed to relativism differently than objectivity stands to non-objectivity. Yet in the same breath he admits that "the vernacular usage of 'absolute,' however, need have nothing to do with relativism." (2011:2). I contend that the extent to which Finlay's objections succeed or fail in challenging the error theory rests not upon his choice of words.

Joyce's (2011:6) rebuttal provides a useful clarification of the structure of his master argument for error theory. He accuses Finlay (2008) of understanding his argument as the following:

> *F1) Morality is conceptually nonrelativistic.*
> *F2) In fact, a nonrelativistic morality is indefensible.*

Thus, the moral error theory is established.

Whilst his argument in actuality is (2011:7):

J1) Morality conceptually involves non-institutional categorical imperatives (NICIs).
> *Hypothesis: Moral non-institutional categorical imperatives are rational requirements.*
> *i) Rational requirements are relativistic (in a certain way).*
> *ii) But moral requirements are nonrelativistic (in that way), hence*
> *iii) Moral requirements cannot be rational requirements.*
J2) In fact, non-institutional categorical imperatives are indefensible.
Thus, the moral error theory is established.

Joyce argues that Finlay's defence of moral relativism, through aiming to disprove F1, actually targets the dispensable proposition (ii). However, despite Joyce's claim that his J1 has "no obvious connection to *moral relativism*" (2011:4), I argue that Finlay's attack remains significant, and this reiteration helpfully allows us to locate it in the context of the overarching argument. As Joyce interprets Finlay's arguments as an attack upon (ii), he finds them innocuous, for (ii) exists only in the service of rejecting *Hypothesis* as a plausible way of defending NICIs. Furthermore, *Hypothesis* functions merely to show that the most likely candidate for denying J2 is unsuccessful. As Finlay follows Joyce in rejecting *Hypothesis* and accepting J2, were his arguments for relativism targeted at (ii), this would be a poor strategy indeed. In actuality, Finlay's relativistic proposal intends to deny J1: he disagrees that morality conceptually involves non-institutional categorical imperatives. If moral values are, in fact, end-relational in the way Finlay argues then they are merely institutional: J1 is false and Joyce's master argument fails. Thus, Finlay's attack remains a legitimate threat, and we shall examine his re-interpretation of the behaviours surrounding moral discourse.

2.3 Evaluating the Evidence

A primary goal of Finlay's (2008) paper is to re-evaluate the apparent evidence that morality is understood as objectively prescriptive. He intends to demonstrate that moral thought and speech is not only *validly* interpreted as reflecting a relational understanding of morality, but actually *better explained* by such an understanding. As ultimately Joyce argues that the content of a concept is intimately related to its *use(s)*, it is crucial for us to understand Finlay's competing explanation of empirically observed behaviours. Finlay identifies seven ostensible sources of evidence for the objective characterisation of morality.

The first of these is *reflective evidence*. This is an appeal to the ways in which ordinary users of moral discourse consciously conceive of their concept. Finlay wishes to dispel the

intuition that most people would, upon reflection, say that their moral values are objective. He points out that many prominent thinkers throughout history have espoused relativistic moral discourse[19], and that a relativistic bent is observed in undergraduate philosophy students. Finlay presents an analogy from morality to motion in order to demonstrate that reflective evidence can be misleading[20]. Ancient users of motion discourse likely would have described their motion judgements as non-relative. However, we now view 'absolute' motion as absurd, as the idea of *movement* only makes sense relative to some other object or frame of reference. Nevertheless, we do not uncharitably accuse ancient users of motion-talk and motion-thought of utilising a concept that is fundamentally flawed. Instead, we judge that the reflective evidence fails to accurately depict the concept at play.

Thus, folk can be genuinely unaware of the content of the concept they are actually employing. Joyce and Finlay agree upon this. Hence, even an agent who enthusiastically avows objectivity as central to moral discourse may be charitably understood as sincerely misunderstanding his own concept. Likewise, an agent convinced morality is relativistic can belie this stance through (say) believing that Gyges is doing wrong *by his own standards* or believing he has a reason to stop his rampant raping. Perhaps all we can derive from this is that reflective evidence counts for little. A relativistic or objective reflective understanding of moral discourse could be ubiquitously accepted whilst, in fact, the other is actually at play. Thus reflective evidence tips the balance in favour of neither conception[21].

The second appeal is to *linguistic evidence*. Finlay posits that a prominent reason objectivity is attributed to moral imperatives is that their relativistic nature is never vocalised. Thus, moral imperatives may be hypothetical after all, with the "if you

---

[19] From ancient thinkers such as Protagoras (Cooper, 1997) to current metaethical relativists such as Gilbert Harman (1975).

[20] Pre-relative motion discourse shall become one of our paradigmatic discourses for which an error theory is *not* appropriate, despite a flawed presupposition.

[21] As famously expressed by Hilary Putnam, "'meanings' just ain't in the head." (1975:227)

want/value…" left tacit. Finlay contends that relativity may remain unsaid for the sake of *rhetorical flourish*. It is well accepted (Mackie, 1977; West, 2010) that a primary function of moral discourse is to extract certain desirable behaviours from other agents. Thus, as *explicitly* relativised imperatives are less effective in achieving this goal, there is compelling reason for omission. Moreover, perhaps constantly vocalised relativisation would become tedious and long-winded. Finlay imagines the captain of the All Blacks rugby team, during a huddle, relativising his advice with "If you want to win the game…". Clearly, it would be unnecessary, frustrating, and even silly. Nevertheless, he is not saying anything new or different when he adds the relativising prefix, but merely making explicit that which need not be: the shared goal of wanting to win is *assumed*, and when verbalised makes the speaker seem foolish. Perhaps, then, in the moral domain, we fail to prefix our claims with "relative to standard S…" for two reasons. Firstly, because this omission increases the apparent power of the imperative, and secondly because a shared standard is assumed.

However, Joyce (2011) questions the coherence of tacit relativity and the relevance of the All Blacks analogy to moral discourse. Imagine the judge at the Nuremberg trials, while sentencing a Nazi war criminal, relativising his moral condemnation with "…according to our standards, and furthermore I demand you subscribe to these standards." If Finlay's understanding is valid the meaning of the claim has not changed; the tacit elements have merely been verbalised. However, would we expect from his audience mild frustration akin to the All Blacks example because he made the tacit explicit? Joyce asserts that the audience would be outraged, and demand *more* from the judge. Such a reaction makes no sense if he is actually saying the same thing. By changing the 'locus of moral condemnation' from a simple connection between the criminal and his crimes to a tripartite connection between *us* (and our standards), the criminal and the crime, we are struck by the contingency of the judgement. If moral truth is merely institutional in this way, this moral condemnation is akin to condemning an agent who violated some standard of etiquette. Understandably, we are

uncomfortable *punishing* violating such an institutional imperative, and we want the moral case to be different. Ultimately, Joyce's case against tacit relativism can be summarised as follows: "It would be astonishing if adding "…irrespective of standards" raises no eyebrows, but adding "…relative to our standards" prompts outrage, while in fact it is the latter kind of suffix that is *really* in play tacitly" (2011:9). Quite so; if making objectivity explicit fails to alter the response to a moral imperative whereas explicitly relative imperatives are treated entirely differently, Finlay's tacit relativism prompts suspicion[22]. Thus the linguistic evidence in favour of objective prescriptivity remains robust.

The third source of evidence Finlay dubs *appraisal evidence*. This is a nod to Joyce's (2001) insistence that moral imperatives are not sensitive to changes in the subjective constitution of the wrongdoer. As such, moral requirement cannot be avoided through citing unusual desires. This inescapability is central to the notion of a categorical imperative. Joyce further argues that for any given circumstances, if there is an action which is morally required it presents authoritative reasons for any agent. However, Finlay derives a different conclusion form the same observations. His preferred theory relativises the truth of moral claims according to the *standards of the person judging*. Thus, if Ted believes Øing is wrong in C according to standard S he will not retract that judgement on the basis of another's desires. However, Ted's claim can be true whilst Julie also truly claims (relative to S*) that there is nothing morally required in C, or that another action is required, or even that Øing is morally *required* in C. Significantly, neither Ted nor Julie will alter their moral judgement upon learning of an unusually constituted agent.

Thus, appraisal evidence is also explained by Finlay's theory. If moral wrongness is relative to the judger's standards, it follows that changes in the constitution of the '*judgee'*

---

[22] In fact, Finlay himself admits that it is a serious objection that "since people don't ordinarily take themselves to be asserting end-relational propositions when they utter ought-sentences, it is most unlikely that they are" (2009:335, footnote 41).

are irrelevant. As such, while it can be legitimately said of the Nazi war criminal that his actions are wrong, this wrongness may fail to have any authority or 'practical oomph' for him, depending upon his standards. There is a fundamental disagreement here about whether this is acceptable, as Joyce believes that we have no business telling someone (say, Gyges) what to do if our demand cannot provide him with reasons. In contrast, Finlay thinks we can say that Gyges did wrong-according-to-an-end and thus separate the wrongness of his action from his psychology. If moral truth is relative such an imperative is licensed, but is merely a valid institutional output, with as much authority for Gyges as the imperatives of kosher.

Nevertheless, we clearly differentiate the "special dignity and necessity" (Foot, 1972) of moral imperatives from other institutional requirements. Thus, it might appear that Finlay fails to distinguish weak from strong categorical imperatives. However, he is well aware of the distinction, and like Joyce rejects strong categorical imperatives as indefensible. Rather, he argues that weak, institutional categorical imperatives characterise moral discourse. Finlay believes that the unique respect for moral imperatives stems from the great value placed upon the moral domain. Thus, *moral* standards and ends are simply the most important ends and do not differ *in kind* from the ends of other institutions. This serves to explain why we are particularly precious about violations of moral standards in a way we are not in the case of etiquette. Uniquely, perhaps, compliance with moral standards is deemed more important than an agent's happiness. In Finlay's words, "… we care more about (eg) the welfare of children than we do about the happiness of those who may be abusing them, and for this reason we do not withdraw our moral appraisals of a person or his actions in response to recognizing his personal reasons" (2008:11). Thus, the relativist can accept Joyce's J2 (that non-institutional categorical imperatives are indefensible), as he maintains that our peculiar insistence about moral imperatives does not stem from their non-institutional nature. This raises some questions addressed in section **2.6**.

Jonas Olson (2010) argues that there is a paradigmatic use of moral discourse, not considered by Finlay, which requires objectivity in order to make sense. This is the sense in which a moral speaker intends to communicate the existence of reasons to Ø that exist independently of any psychological facts about the judgee *or the judger*. Some uses of moral imperatives imply not only a lack of contingence upon the subjective constitution of the person judged (appraisal evidence), but also not upon the constitution of the judger. This is hinted at by the Nuremberg judge example. Explicitly relativising a moral judgement to one's standards, such that it is clearly incidental that one happens to have said standards, leads others to react as though a judgement less powerful than a full-blooded moral one has been made. Thus, perhaps a paradigmatic use of moral discourse is to make demands upon others without invoking anything about one's own views. This use of moral imperatives, by which the appraiser intends to communicate something about the *other agent's reasons*, or the ends he is rationally obliged to value, simply makes no sense on Finlay's account[23]. However, the question remains how central this particular use of moral discourse actually is. Furthermore, the relativist will maintain that, charitably understood, moral imperatives never intend to indicate the presence of such reasons.

The following three sources of evidence Finlay treats as different facets of the same observation, as they all revolve around interactions between agents with different moral standards. *Address evidence* notes that moral imperatives are often addressed to non-subscribers despite this fact. The thought here is that it would be strange for people to morally condemn others with different standards if moral value is relative to standards, which are themselves beyond reproach[24]. *Expectation evidence* is a reflection upon the fact that moral imperatives addressed to non-subscribers are often accompanied by an expectation that the imperative will motivate, or remind the agent of pre-existing

---

[23] Chapter One lead us to the conclusion that this particular use is metaphysically indefensible.
[24] What the relativist *cannot* say is that, whilst morality is relative to standards, some standards are better than others, or some standard is *correct*.

motivation. *Disputation evidence* notes that disagreement occurs between agents with fundamentally different moral standards. Moreover, it is assumed by both parties that they are disagreeing about something substantial, and univocally utilising moral terminology. This assumption that fundamental moral disagreement is at the level of "the fact of the matter" reflects a substantial difference between moral dispute and (say) disagreement about ice-cream flavours. If moral truth is end-relational, these agents should realise that they are speaking equivocally, and that the dispute is not genuine. However, the observed behaviours reflect no such understanding, and appear to indicate an objective concept at play.

Finlay's primary reason for rejecting these observations as evidence that moral discourse implies objectivity is that such occurrences are "nowhere near as common or characteristic of moral discourse as one might think" (2008:13). Finlay sketches what moral discourse would look like in a homogenous society: an understandable lack of address, expectation or disputation evidence because everybody subscribes to the same ends. He then argues that, while our own society is probably the most morally fragmented in history, our moral practises are 'largely continuous' with those in the homogenous society. Four reasons are given. Firstly, even if society encompasses individuals with widely varying moral beliefs, most moral discourse will take place between individuals with similar views. Secondly, agents *assume* their interlocutor shares moral standards, such that even their relative-to-standards judgements remain authoritative and reason-bringing. Furthermore, any apparent absence of shared moral standards can be attributed to one's opponent misrepresenting (purposefully or otherwise) his actual views. Thirdly, Finlay claims that, if there is an overt lack of agreement upon moral standards, such dialogue would in actuality be abandoned as a vain pursuit, as the equivocation of moral terminology would be accepted. For example, in the abortion debate, a pro-life supporter might assume the shared standard that killing innocent persons is wrong, and proceed to argue that a foetus is an

innocent person. However, if this fundamental standard is not accepted, it is at least not *obvious* that he would feel justified continuing the dialogue. Finally, imbuing an imperative with unjustified moral authority (through keeping relativity tacit) provides the illusion of additional strength, rhetorically demanding that the implicit standards be adhered to. This is not because they are the *correct* standards, as this begs the question in favour of an objective moral conception. Rather, the demand is to comply with the standards one has chosen, and *values more than the happiness* of one's interlocutor. Thus, Finlay concludes that these behaviours are pragmatically explicable, and compatible with the relativist's story.

However, Finlay here fails to justify his claim that moral discourse in our highly fragmented society is largely continuous with the equivalent discourse in a homogenous society. Why, for example, assume that most moral disputes occur between agents who subscribe to the same standards? Jonas Olson points out that "Even a cursory glance at public political debate in many countries will reveal fundamental moral disagreements between conservatives and feminists; socialists and neo-liberals; cosmopolitans and nationalists; etc." (2010:18). Even Finlay's own example of the abortion debate indicates that deep-seated moral dispute is very much alive. Despite evidence of fundamental disagreement, these debates continue. Perhaps the appearance of divergent standards in one's opponent *is* interpreted as a misrepresentation, but this can only go so far. After lengthy debate, insistence that your opponent actually, deep down, *agrees* with you starts to look suspiciously non-relativistic. Moreover, Olsen encourages us to consider metaethical disputation. The debate between utilitarians and deontologists, for example, indicates that even academic philosophers (whom, one would hope, are very careful about such things) assume the univocal nature of moral terminology. There is no assumed standard here, as the *argument itself is about standards*. Finlay asks us to recall the last time we engaged in moral discourse with the likes of Charles Manson, but why insist on such extreme divergence?

Address, expectation and disputation evidence is abundantly available, in favour of objective prescriptivity.

The final source of evidence is *reactive attitude evidence*. This encompasses the observation that we *blame* agents who have erred according to our own moral standards, not theirs. This behaviour signifies that you have done something wrong and punishable when you subvert *my* moral standards in a way entirely different to that of subverting my standards of etiquette. Thus there appears to be an assumption that moral standards are shared and apply equally to anybody. As such, this evidence is aligned with Joyce's proposal that competent users of moral discourse assume that one who acts immorally has failed to respond to reasons that are authoritative *for her*. Furthermore, this implies allegiance to strong categorical imperatives. Finlay's response here is to admit that this blame does imply such allegiance, and thus an error theory is called for in the case of 'second order' blame judgements. However, he claims this would fail to contaminate the semantic truth conditions of first order judgements of right and wrong. For these second order judgements to contaminate the concept itself, it would have to be the case that without such reactions of blame, one could not coherently hold the action to be wrong.

Nevertheless, there is something intuitive about the necessity of blame. If I believe Øing is morally wrong, and you disagree and insist upon Øing, yet I do not blame you for your actions, this attitude should lead me to ask some important questions about in what sense I judged the action to be morally wrong[25]. The moral sphere is undeniably one with great value placed upon it, and judgements of wrongness tend to go hand in hand with judgements of blame. Perhaps in the case of *unintentional* immoral acts wrongness and blame come apart. Nevertheless, in such cases blame is the natural response, and this is tempered by whatever mitigating circumstances may apply. Finlay observes that we do not

---

[25] Note that a sincere relativist could coherently make such a judgement and yet withhold blame. However, empirically speaking, these judgements tend to coincide.

blame agents for failing to respond to authoritative reasons in general. Thus, unresponsiveness to authority is insufficient to explain our blame; there is some other factor at play. However, the special nature of *moral* authority is accepted by both Joyce (through strong categorical imperatives) and Finlay (as we care more about the institution than happiness). Reactive attitude evidence still supports an objective understanding of moral discourse.

Finlay here has shown that the behavioural evidence indicating that users of moral discourse employ an objective conception *can be explained* given an assumption of relativism. While the evidence is *better* and more coherently explained through an appeal to objective prescriptivity, Finlay's relational interpretation can (for the most part) account for these behaviours. Nevertheless, Finlay goes on to argue that even hypothetically assuming universal acceptance of objectively prescriptive moral discourse, it is uncharitable to assume that this contaminates the truth-conditions of moral claims. In other words, people may genuinely be unaware of precisely what their moral assertions and thoughts actually mean. Note again, however, that this works both ways: even if the assumption of relativism is ubiquitous on a conscious level, if morality is *used* in ways that imply objective prescriptivity, this element may be essential to the concept. In any case, we shall now move on to consider Finlay's positive argument that the error theory is false.

2.4 Absolutism and Meaning

Thus far, Finlay (2008) has argued defensively that moral thought and behaviour *can be* legitimately interpreted as relativistic. However, Finlay believes that even if, in fact, most users of moral terms aver objective morality this would fail to systematically infect the truth conditions of the discourse. Thus, for the sake of argument he hypothetically concedes to the error theorist that most people understand moral discourse as objectively prescriptive.

However, as he agrees that there is no sense to be made of non-institutional categorical imperatives, this position necessarily involves attribution of genuine conceptual ignorance on a wide scale. The case of ancient absolute motion discourse has already provided a convincing example where, instead of attributing widespread conceptual error to a discourse in virtue of what people *think* they're talking about, we charitably assume them to be actually talking about defensible relative motion. Thus we have a paradigmatic case where the meaning and truth conditions of sentences *are not* what the speakers think they are. Hence there is space for Finlay to argue that the same applies to moral discourse.

Consideration of *why* ancient people espoused an indefensible notion of absolute motion may foster sympathy for Finlay's argument. Whilst there is no sense to be made of an object simply *moving* without some frame of reference, or another object to compare it to, it is easy to see why, *prima facie*, motion judgements might appear to be objective. The surface of the earth itself provides a frame of reference so enormous and ever-present that judgements which in actuality are relative-to-the-surface-of-the-earth appear to be non-relative. Thus, one who speaks of (tacitly) relativised motion can confer with a believer in absolute motion without either realising their disagreement. Similarly, moral judgements made relative to commonly held standards may give the illusion of being objective. However, this is simply a chimera caused by abundant convergence in moral ends. Hence, what motion and moral judgements have in common is the near-universal constancy of one parameter. If the fact that this parameter is merely 'near'-universal remains unnoticed, its indispensability may go unnoticed. Thus we have a reasonable causal story to tell about how metaethical reflection and behavioural observation could result in belief in objective values whilst in actuality moral judgements are tracking relative values.

It is essential that the error theorist's argument, when applied to ancient motion discourse, does not find it guilty of systematic error. Thus the error theorist must have some

story to tell about the relevant difference between the two discourses. Motion he is happy to allow relativistic truth-conditions, and maintain that these were the actual truth-conditions *the whole time*, whilst morality remains inextricably tied to objectivity such that a relativistic revision simply *would not count as moral discourse.* In Finlay's words, "The difficult question we cannot avoid here is: how is the (semantic and conceptual) content of our language and thought determined?" (2008:22). If somebody's concept is precisely what they think it is, this provides unintuitive results in the motion case. Finlay implies that the error theorist's argument turns on the equivalent thought about morality; that the *reflective evidence* mentioned above is the substantial basis of the argument. However, while there is room for the relativist to argue his case *despite* reflective evidence, the same goes for the error theorist. Indeed, Joyce (2001, 2011, 2012) conceives conceptual commitment as involving interplay between reflective understanding and *function* - the ways the discourse is *used* and the behaviours surrounding it.

Thus, there remains space for the error theorist to drive a wedge between motion discourse and moral discourse: if relativistic motion discourse can perform all the functions of absolute motion discourse, but relativistic moral discourse cannot perform all the functions of objective moral discourse[26], a disanalogy appears. However, this is an empirical claim which remains undecided, as is the *necessity* of whichever elements of objective moral discourse fail to survive the conceptual revision, both of which shall be discussed in the final sections. Nonetheless, the analogy to motion is a useful tool to evaluate the aptness of moral error theory. Therefore, the next section involves further comparisons with *other discourses infected with false presuppositions*. Given that the appropriate reaction to discovering the infected presupposition appears to vary between discourses, perhaps we can convincingly argue for categorising moral discourse as either irreversibly error-laden or resurrectable through revision.

---

[26] Authoritatively censuring Gyges comes to mind here.

2.5 Discourses

Ultimately, both parties to this dispute agree about whether an error theory or conceptual revision is appropriate in the four paradigmatic discourses of witchcraft, phlogiston, water and motion. Thus, evaluation of the reasoning behind the strong, convergent intuitions in these cases provides a useful framework within which to consider morality. We shall investigate the rationale behind the categorisation of these four discourses, and then examine the nature and function of moral discourse.

Witchcraft and phlogiston discourses are deemed irrevocably error-laden because of a false presupposition. Thus these discourses have been largely abandoned. However, water and motion discourses were, once upon a time, also infected with such a presupposition. Yet the consensus in these cases is that the flawed presupposition was legitimately extracted without altering the concept itself, changing its truth-conditions, or warranting a new label. In fact, we insist that the amended discourse is what we were using the whole time. Therefore, mere commitment to a false presupposition is not a *sufficient* condition to warrant an error theory. Consequently, we must further investigate the nature and necessity of the commitment in order to illuminate what other factors are at play.

The flawed presupposition in 'witch discourse' is the use of magical powers. A woman who bears all the other hallmarks of a witch (solitary, black hat, disruptive of the patriarchy etc.) but nevertheless lacks magical ability is not a witch. Such is the *necessity* of magic-use to the concept of witchcraft. Thus, the less-than-superstitious individual who denies that magic powers exist in the world finds the extension of the term *witch* to be empty. Hence, the need for an error theory. Witch discourse as it stands fails to refer, leaving witch-assertions systematically false. Furthermore, a modified version of witch discourse which abandoned talk of magic appears to not be about witches after all. We deem labelling women who use magic to be a *non-negotiable use* of witch discourse, such that the 'second

best contender' for the concept is simply not good enough. Thus, if morality's commitment to objectivity is akin to witchcraft's commitment to magic, the moral error theory is vindicated.

'Phlogiston' was a term used to refer to a substance stored in flammable materials and released during combustion. Prior to a chemical understanding of the role of oxygen in combustion, phlogiston was thought to coherently explain observations such as changes in the weight of a burning substance. However, it has been discovered that there simply is no such stuff universally to be found in flammable materials and that the observations are best explained through understanding the role of oxygen. Thus, the extension of the term *phlogiston* is empty. Furthermore, revision of the concept such that 'phlogiston' actually refers to oxygen is illegitimate. The property of being stored in flammable materials is deemed essential to phlogiston, and oxygen does not share this property. Thus, sensible people espouse an error theory about phlogiston, and do not deem oxygen discourse to be a resurrected version of phlogiston discourse. Hence, once more, if morality's commitment to objectivity is akin to phlogiston's definitive commitments, the moral error theory is vindicated (Joyce, 2001).

We shall now move on to discourses which, on the face of it, have been *legitimately* revised. Water, for example, was once thought to be an element. It is now known to be a compound. However, when this was discovered, the reaction was *not* to accuse all prior water-claims of being false. Rather, it was surmised that this compound was what we had been talking about *the whole time*. Perhaps the really central, non-negotiable features of water discourse involve referring to stuff that is clear, drinkable, falls from the sky, fills the oceans etc. Nevertheless, the commitment to water's elemental nature appears to have been rightfully revised. However, why is water discourse charitably altered whilst phlogiston gets the uncharitable boot? There are some intuitive answers here. There is something in

the nature of water as a macroscopically observable substance which ensures conceptual continuity. One can point to a bucket of water and say "that's the stuff I'm referring to - right there!" (Putnam, 1975). When one discovers that the nature of the 'stuff' isn't precisely how one thought it was, this doesn't warrant a new concept and label, for it is clearly the same stuff. Hence the disanalogy to phlogiston. Competent phlogiston speakers would have pointed to flammable objects and said "it's the stuff in there that is released during combustion" and thus could hardly claim they were speaking of oxygen in the air the whole time.

However, when Finlay and Joyce alight upon the 'point' of a discourse as decisive in evaluating the legitimacy of a certain revision, they are not referring to the above method of point*ing* to the thing under discussion. Clearly such a methodology fails when discussing the nonmaterial concepts of motion and morality[27]. Rather, Finlay calls the primary intention of the discourse to refer its 'referential point'. In light of this, perhaps motion discourse can be understood as primarily intending to refer to the changes in the spatial positioning of objects over time. This (entirely plausible) understanding of the central intention of motion talk is not undermined by the conceptual revision from absolute to relative motion, as this goal is still achieved. In addition, it appears that there are no uses of motion-talk that *do not* survive the conceptual revision[28]. Thus this revision is valid, allowing the same verbal output and use. Finlay interprets Joyce as being committed to claiming that the referential point of moral discourse is to refer to objectively prescriptive values. If this is the case, the extension of the phrase "X is morally right" remains empty without universally authoritative reasons, as was argued in Chapter One. In contrast, we have already seen Finlay argue that the

---

[27] Arguably this is not the case, as we can still point to cases of motion and moral judgement and say that is what we are talking about.
[28] The intention to *describe changes in absolute location* may not survive translation, but such use has no apparent advantages over the relational equivalent.

apparently objective features of moral discourse are better explained by a *non-referential point*, i.e., rhetorical flourish.

Finlay presents another way of explicating the referential point of a discourse: *essential application conditions*. For example, the term 'witch' is appropriate if, and only if talking of a woman with supernatural powers. Moreover, the term motion is applicable if, and only if an object has changed position over time, relative to some frame of reference. However, Finlay goes on to claim that *Øing is morally wrong if, and only if, it violates certain standards or ends*. In contrast, the error theorist will likely maintain that Øing is morally wrong if, *inter alia,* there is an authoritative reason for any given agent to refrain from Øing. Unlike the error theorist's formulation, Finlay's provides a *justifiable* framework within which to morally denounce an action; in lieu of objective values a *metaphysically defensible* thing to which moral claims could refer are such standards. This is a point Finlay has already argued. However, it is central to the error theorist's case that moral claims are, in fact, *not* referring to something defensible. Finlay is being overly charitable here in assuming that people *must* be referring to the closest instantiated relative of their intention: "Assuming there is no genuine absolute motion, or genuine absolute moral properties, the absolutist's judgements could not be responsive to these fictional properties" (2008:26).

Alarmingly, this charitable interpretation can be used to defend revision of concepts such as witchcraft where an error theory is unanimously accepted. Whilst judgements are always in response to some genuinely perceived stimuli, if the agent infers from this stimuli a further, non-existent, property, his concept may still be error-ridden. As per Hume's *pathetic fallacy*, the mind can 'spread' itself upon external objects and take what *is* there as evidence for that which is not (Mackie, 1977:42). Hence, Finlay's charity appears to mistakenly vindicate witch discourse. If users of moral and motion discourse could not have been talking about the uninstantiated objective properties, charitably witch-talk could not

have been referring to uninstantiated magical powers. Rather, the essential application condition for witch-hood must be (say) a woman who challenges patriarchal values. Thus, counter-intuitively, there are witches!

However, Finlay believes he can drive a wedge between this parody argument to justify witch discourse and his sincere argument to justify moral discourse. In order to demonstrate that witch discourse remains indefensible, he points out that even that which is necessarily coextensive need not be part of the essential application conditions of a concept. In other words, just because every person referred to as a witch happens to also be a woman who challenged the patriarchy, this co-extensionality doesn't mean that *that's what witches are*. For example, we can imagine a case of a woman who disrupts the patriarchy so as to be a witch according to the parody argument, and yet is not, and (suspending disbelief about magic) vice versa. Thus the evidence that such women exist does not vindicate witch discourse. Finlay's view in the moral case is that moral judgements do not imply sensitivity to objective values - there are none - but what a person *takes to be evidence* of the realisation of such values: his judgements relative to his own moral standards. The fact that people take their relational judgements to be evidence of objective values will not undermine the validity of the judgement as long as it is the relational judgement is the essential application condition. Thus, the error theory uncharitably supposes that because people extrapolate objective judgements from their genuine relative judgements, the relative judgements themselves are false.

Nevertheless, Finlay fails to differentiate moral discourse from witch discourse. The flaw in the argument to validate witch discourse is that the evidence and its associated instantiated property are non-essential to witch-hood. However, it is entirely reasonable for the error theorist to insist that the equivalent holds for moral discourse. One can imagine a case where "the standard evidence of moral wrongness obtains, but the action (say) is not

morally wrong, or of a situation where none of the standard evidence obtains (no apparent suffering, no hurtful intentions) but in which the action is wrong nonetheless." (Joyce, 2011:12).[29] Thus, despite there being (perhaps trivially) some standard against which any given moral judgement can be understood, this doesn't imply that this relational property is the essential application condition for a moral judgement. Furthermore, Finlay's assertion that *relational judgements* are taken as *evidence* for objective moral truth can be resisted. Far more likely is that suffering, harm, and selfish or cruel intentions are taken as direct evidence for objective moral wrongness.

Thus Finlay has failed to expose an asymmetry between witch discourse and moral discourse. From the evidence for witches (confessions, testimony, patriarchal challenging etc), superstitious medieval people extrapolated the property of witch-hood, which is in fact uninstantiated. Perhaps this evidence, in fact, reliably constitutes the instantiated property of disrupting the patriarchy, but the witch error theorist has other reasons to think that this property will fail to adequately serve as the essential application condition of witch discourse. Analogously, whilst the evidence for morality (selfishness, hurtful intentions, suffering etc) may reliably constitute various instantiated relational properties, the moral error theorist has reasons to think the relational conception inadequate to transparently serve as moral discourse. Hence, the charitable argument that we should assume people to be referring to whatever instantiated property best fits the evidence fails.

2.6 The Crux of the Argument: Evidence Revisited and Some Unusual Implications

Insistence that the essential application conditions for moral judgements are relational is the crux of Finlay's (2008) positive argument that the error theory is false. His

---

[29] Consider, for example, that many view consensual sex between a brother and sister as immoral despite a viable absence of suffering, harm or hurtful intentions. Conversely, despite the pain he inflicts, I do not morally condemn my dentist.

re-evaluation of the evidence was intended as a dispensable prelude to this argument. However, in order to resist Finlay's relational theory, it is incumbent upon the error theorist to expand upon her 'other reasons' for thinking relativistic moral discourse an inadequate contender for morality. Inevitably, this leads us full circle back to an appraisal of the evidence. Therefore, let us now reconsider several of the more decisive points previously mentioned, beginning with linguistic evidence for objective prescriptivity.

Finlay writes that if people were genuinely employing an objective conception of morality, they would likely be sceptical that relational judgements count as moral. However, this is precisely the thought motivated by the case of the Nuremberg judge - that *explicitly* relational judgements *do* strike the folk as failing to capture something essential to the authority of morality. It is difficult to ascertain whether tacitly relativised moral statements invoke similar scepticism because we may be entirely unaware of the tacit elements in another's speech - indeed, Finlay would have us believe that these relational elements are tacit even in the thoughts of the speaker. The thought experiment revolving around making explicit what Finlay believes is tacit was intended to surmount this difficulty. If you agree with Joyce's estimation of the folk's probable reaction, you will likely be suspicious of Finlay's tacit relativity. Thus, the link between linguistic evidence and objective moral discourse is far simpler and more coherent than the relativistic explanation.

Secondly, observation of moral address, expectation and dispute provides the error theorist with further reason to be doubtful of the relational interpretation. Contra Finlay, moral judgement and dispute between those with fundamentally different standards *is* a common occurrence, as we have seen argued by Jonas Olson (2010) in section **2.3**. Finlay's insistence that our moral discourse is largely continuous with that of a homogenous society is further undermined by the increasing prevalence of such debate *online*, where convergence of standards in one's geographical region is irrelevant. These fundamental

disputes are undertaken with a passion that belies the possibility that moral truth is equivocal and merely determined by one's own standards.

Thirdly, Finlay formulates the difference between weak and strong categorical imperatives in terms whether we value the demand more than the happiness of the agent in question. This is a defensible position to take, but may have some unusual consequences. Imagine that I care very much about the kind of music that people listen to. I will interfere (at the expense of people's happiness) to ensure that they listen to the right music. I will punish people for listening to the wrong music. If valuation above happiness is *sufficient* to make values moral, my musical views are moral. Finlay may choose to bite the bullet on this and admit that listening to jazz is a moral end for me, and listening to pop is morally wrong. This is coherent, but some might find it an unintuitive consequence of Finlay's theory that my musical judgements are moral, as they are aesthetic, and unrelated to harm[30]. Alternatively, Finlay might argue that whilst outweighing considerations of happiness is necessary for moral value, there are further conditions required. In this case a convincing story needs to be told about the other requisite factors. Consider, for example, that in Western culture judgements about what to wear are primarily aesthetic, whereas in Muslim culture decisions about clothing are heavily morally loaded. These considerations do not weigh decisively, but there are challenges with conceiving the peculiar nature of moral value this way.

Fourthly, there is a crucial disanalogy between revision of absolute motion and revision of objective morality. In all likelihood, no competent user of motion talk who is aware of the options of a relational or absolute understanding of the discourse, and the reasons why the absolute form is defective will insist upon defending absolute motion judgements. There is nothing to be achieved by absolute motion judgements which cannot

---

[30] Jonothan Haidt (1992) has notably argued that the *content* of moral judgements and the domains considered moral varies across cultures.

be achieved by relational motion judgements, and the arguments for their insensibility are convincing. This debate is long settled. However, amongst the folk (and even metaethicists) who are aware of the options of relational or objective moral judgements, there are many who insist upon asserting non-institutional categorical imperatives, who insist that relational judgements are *too weak*, that they cannot do the same jobs and express the same meaning. The difference in attitude between those aware of the two forms of each discourse should encourage us to question the strength of the analogy here.

In addition to these reconsiderations, Jonas Olson has recently argued that there are other reasons to think Finlay's relativistic project fails to capture something essential to moral discourse. Olson (2010) points out that Finlay's system of relativising moral claims according to standards necessarily applies to *all* moral claims, including claims about standards themselves. For example, to assert that utilitarianism is the correct metaethical theory is a moral claim. However, such metaethical assertions can only be relativised to themselves; they are their own standard. This reasoning indicates that every claim about a moral standard is actually a tautology, such as 'in order to maximise overall utility, one must maximise overall utility'. While Finlay (2009) is prepared to bite the bullet on this, it remains a surprising result which fits incongruously with observed behaviours relating to moral standards. Olson identifies several strange implications of thinking of moral standards as tautologies. Remarkably, there is no correct or incorrect metaethical theory. Agents making fundamental claims about moral discourse do not *intend* to utter tautologies, and engage in sincere debate about which standard is correct, which is bizarre behaviour if all contenders are trivially correct.

Olson goes on to point out that competent users of moral discourse intend their fundamental moral claims to be informative, sometimes controversial, and not obviously true. To discover, for example, that the morally correct thing to do is that which maximises

overall utility might change one's behaviour, beliefs and judgements. How a tautology could perform these belief- and action-guiding roles is quite mysterious. Finlay (2009) argues that assertions of moral standards function as 'conversation stoppers' which demand and motivate action, rather than conveying semantic content. Whilst such assertions likely sometimes perform these tasks, agents also take one another to be conveying content, and are even prepared to *review* their moral standards. As agents do not use fundamental moral claims as though they are tautologies, it is most unlikely that this is in actuality what they mean. Olson thus concludes that attributing widespread false beliefs (and absurd behaviours) relating to the nature of fundamental moral discourse is a less charitable understanding of moral discourse than the objectively prescriptive (though error-ridden) conception favoured by Joyce and Mackie. Significantly, however, whilst this makes most people's metaethical views and behaviour look strange, the conclusion that there is no uniquely correct moral standard is quite compatible with Finlay's relational theory. Nevertheless, the mere fact that Finlay's theory comfortably tolerates a consequence that, if commonly accepted, would drastically alter fundamental moral discourse indicates that his understanding of moral discourse fails to fails to capture the prevailing view in this case.

In the light of this scepticism about relational moral judgements, the error theorist will likely maintain that the essential application conditions for morality revolve around reasons that are universally authoritative. Hence, the impasse between the moral relativist and error theorist. In the case of Gyges the error theorist claims we have no business demanding he refrain from his debauchery unless we can provide a valid reason to refrain. However, Finlay disagrees, claiming "I think we should deny that the ordinary concept of wrongness entails any such thing" (2008:29). The absence of a valid reason for Gyges to refrain is simply explicable for the moral relativist. The demand that he refrain is made relative to certain standards and ends of the speaker which Gyges does not share – end of story. Thus, morality construed this way can make sense of Gyges' indifference, whilst licensing moral

condemnation of his behaviour. However, this conception cannot explain the intuition that Gyges is *breaking his own rules*, captured in Joyce's discussion of strong categorical imperatives: "because our moral framework is categorical we can carry on legitimately saying 'Gyges, you ought not do that!' But if our utterances are merely a verbal output that has been validated by an institution of our own creation, it all begins to sound rather shrill" (2011:5).

Finlay's explication of the error in the error theory vitally turns upon his claim that the essential application conditions for Øing to be morally wrong or right are judgements relative to some standard. The error theorist, in contrast, maintains that objectivity is essential to morality, and that such a conception makes better sense of observed behaviours surrounding the discourse. This is justified, as the objective interpretation involves less self-deception and a more transparent relationship between what people think and say and what they mean, as well as licensing the stronger meaning desired by moral speakers. Thus objective moral discourse is a superior candidate for the concept of morality. However, Finlay's tacit understanding of relational judgements is logically coherent, and is certainly not *disproved* by virtue of less parsimoniously explaining the evidence. Furthermore, it allows for a metaphysically defensible understanding moral discourse, which may count in its favour.

Conclusion: An Underdetermined Dispute

Chapter One follows Richard Joyce (2001) in concluding that objective moral prescriptions require universally valid reasons. As such reasons are not to be found, objective moral claims are systematically false. Chapter Two considers Stephen Finlay's (2008) contention that moral discourse is best understood as relational, and the subsequent debate with Joyce (2011) about whether moral discourse is actually in the business of providing objective prescriptions.

I contend that this dispute will likely remain underdetermined insofar as the thoughts and behaviours of competent users of moral discourse will not settle questions about the commitments of the discourse itself. Indeed, the substantial role that tacit content plays in Finlay's interpretation of the discourse severely limits the possibility of empirically resolving these questions. Furthermore, it is possible that the relativistic and error theoretical metaethical stances are genuinely equally valid in a more significant sense than mere empirical underdetermination: moral discourse may actually have no privileged meaning.

This notion of 'metaethical pluralism' is espoused by Joyce in his forthcoming paper of the same name[31]. Joyce takes his point of departure from David Lewis' (1989) discussion of moral value. Lewis argues that "values are those things which we are disposed, under certain idealized conditions, to desire to desire" (Joyce, 2012:2). However, he finds it unsettling that according to this conception, moral values *might have been other than they are commonly considered to be*, and considers demanding those values which we are *necessarily* disposed to desire to desire. This move soothes worries about the contingency of value and provides a more powerful moral conception, but at the price of making the set of moral values empty, as there is likely *nothing* which we are necessarily disposed to desire to

---

[31] While Joyce writes regarding the dichotomy between the moral naturalist and moral sceptic, the notion of pluralism remains applicable to this dispute.

desire. Lewis concludes that while the moral conception that includes necessity best captures what it would take to *perfectly* deserve the name moral value, "loosely speaking, the name may go to a claimant that deserves it imperfectly" (1989:136-7).

Lewis infers that what one makes of this situation is primarily a matter of temperament. One can fret about the shocking discovery that there are no moral values, or accept that the nature of value is not how one thought it was. Lewis and Finlay prefer the conservative adjustment of the nature of value, whilst Joyce takes the former route. However, assuming good (but inconclusive) arguments can be given for both sides as we have seen in this dispute, perhaps nobody is, in fact, *right.* Lewis says "When it comes to deserving a name, there's better and worse but who's to say how good is good enough?" (1989:137).

Joyce has proposed two ways to consider this question, though neither is decisive. In *The Myth of Morality* (2001), we imagined an anthropologist who discovers a culture that utilises a certain discourse that shares many characteristics with our 'morality'. Thus the question of whether a discourse is a good enough deserver for a term becomes one of *translation*: would we translate said discourse to *morality* in English, or would it be a mere *schmorality* (Joyce, 2007)[32]? However there is no definite answer to this question, which is still loaded with difficulties, not least of which is determining who 'we' are[33]. Subsequently, Joyce has appealed to how we *use* the discourse in question. If discourse Q is used in ten different ways, and the imperfect claimant Q* can only be used in nine of these ways, then Joyce argues that Q* is an inadequate claimant, as it cannot "pull its pragmatic weight in everyday life" (Joyce, 2012:5). Joyce argues that this way of conceiving the issue may bypass considerations of temperament: if Finlay's relational values can play all the practical roles of

---

[32] A similar thought experiment is proposed in West (2010).

[33] Note that this was not intended to be a usable decision procedure, but merely a way of conceiving the task at hand.

objective moral prescriptions then his imperfect claimant is vindicated, if it cannot then it is inadequate. However, the uncertainty is here reiterated anew. Perhaps Q* needn't fulfil *every* role Q is used for, but merely *enough* roles. However, who's to say how many is enough, or which are necessary?

The primary role of moral discourse that Finlay's relational theory cannot fulfil is that of asserting the existence of reasons that are universally authoritative, thus ensuring that moral prescriptions provide authoritative reasons to the *judgee*. Error theorists and objective moral realists both believe that this is a paradigmatic use of morality such that a revision which would not pull its weight in this area will be inadequate. This is because advocates of these theories believe that this is what moral imperatives *mean*, and thus to assert such reasons is a *necessary* function, or essential application condition. However, as Finlay's theory reinterprets the semantic meaning of moral claims and the truth conditions thereof, he is entitled to insist that his understanding of moral discourse *can* pull its weight in terms of performing the functions required of morality.

The significant way these theories come apart can once again be illustrated through Gyges. Undoubtedly it is a central use of moral imperatives to censure one who rampantly rapes and pillages. However, both theories licence this censure. They only differ insofar as an objective moral prescription necessarily would *authoritatively* censure Gyges, such that he would actually have reason to stop, whereas a relational moral imperative merely tells Gyges that according to a standard endorsed by the speaker he has a reason to stop. If you are inclined to think that a moral imperative requires the power of the former conception, and agree with the arguments of Chapter One towards the systematic falsity of objective moral prescriptions, this will lead you to embrace the moral error theory. However, if you are convinced of the adequacy of Finlay's interpretation of the meaning of moral claims, you might aver moral relativism.

If, however, one remains indifferent about the requisite strength of the imperative against Gyges (or remains indecisive for other reasons), there is a final appeal that can be made to the pragmatic benefits of espousing either theory. It might appear that pragmatic considerations would support the relativistic reinterpretation, as this allows us to maintain moral discourse. However, even if you would prefer a world that includes moral talk, thought and dispute, the options of moral fictionalism (Joyce, 2001, 2005), and non-cognitivism (Carnap, 1935) remain equally viable contenders, each exchanging the metaphysical baggage of realism or relativism for a new set of challenges. Furthermore, some philosophers such as Garner (2007) and Hinckfuss (1987) have controversially argued that, pragmatically speaking, the benefits of abolishing moral talk and speech outweigh those of keeping them. Discussion of these positions is beyond the scope of this thesis, but perhaps such pragmatic analysis is the way forward for metaethical thought[34].

References

Brink, D. (1989). *Moral Realism and the Foundations of Ethics.* Cambridge: Cambridge University Press.

Broome, J. (2007). Does Rationality Consist In Responding Correctly To Reasons? *Journal of Moral Philosophy, 4,* 349-374.

Carnap, R. (1935). *Philosophy and Logical Syntax*. London: Kegan Paul, Trench, Trubner & Co. Ltd.

Cooper, J. M. (Ed) (1997). *Plato: Complete works.* Indianapolis: Hackett Publishing Company.

Finlay, S. (2008). The Error in the Error Theory. *Australasian Journal of Philosophy, 86*(3)*,* 347-369.

Finlay, S. (2009). Oughts and Ends. *Philosophical Studies, 143*, 315-340.

Finlay, S. (2011). Errors Upon Errors: A Reply to Joyce. *Australasian Journal of Philosophy, 89*(3), 535 - 547.

Foot, P. (1972). Morality as a System of Hypothetical Imperatives. *Philosophical Review, 81*, 305-316.

Garner, R. (2007). Abolishing Morality. *Ethical Theory and Moral Practice, 10*(5), 499 - 513.

Haidt, J. (1992). *Moral Judgment, Affect, and Culture, or, is it wrong to eat your dog?* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Harman, G. (1975). Moral Relativism Defended. *The Philosophical Review, 84*(1), 3-22.

Hinckfuss, I. (1987). The Moral Society: Its Structure and Effects, in *Discussion Papers in Environmental Philosophy*. Canberra: Australia National University.

Hobbes, T. (1924). *Leviathan.* London: J.M. Dent. (Original work published 1651).

Hume, D. (1978). *A Treatise of Human Nature.* Oxford: Clarendon Press. (Original work published 1739).

Hume, D. (1983). *Enquiry Concerning the Principles of Morals*. Cambridge: Hackett Publishing Company. (Original work published 1751.)

Joyce, R. (2000). The Fugitive Thought. *Journal of Value Inquiry, 34*(4), 463-478.

Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.

Joyce, R. (2005). Moral Fictionalism, in *Fictionalism in Metaphysics* (ed. M. Kalderon). Oxford: Oxford University Press.

Joyce, R. (2007). Morality, Schmorality, in *Morality and Self-Interest* (ed. P. Bloomfield). Oxford: Oxford University Press.

Joyce, R. (Accepted for Publication, 2011). The Error in 'The Error in the Error Theory'. *Australasian Journal of Philosophy.*

Joyce, R. (Accepted for Publication, 2012). Metaethical Pluralism: How both moral naturalism and moral skepticism may be permissible positions, in *Ethical Naturalism: Current Debates* (eds. S. Nuccetelli & G. Seay). Cambridge: Cambridge University Press.

Kant, I. (1997). *Groundwork of the Metaphysics of Morals* (M. Gregor, trans.). Cambridge: Cambridge University Press. (Original work published 1785).

Lewis, D.K. (1989). Dispositional theories of value. Part II. *Proceedings of the Aristotelian Society,* supplementary vol. 63: 113-137.

Mackie, J.L. (1977). *Ethics: inventing right and wrong.* Harmondsworth: Penguin Books.

Olson, J. (2010). In Defense of Moral Error Theory, in *New Waves in Metaethics* (ed. M. Brady). Basingstoke: Palgrave MacMillan.

Putnam, H. (1975). The meaning of 'meaning', in *Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge: Cambridge University Press.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Southwood, N. (2008). Vindicating the Normativity of Rationality. *Ethics, 119*(1)*,* 9-30.

West, C. (2010). Business as usual? The error theory, internalism, and the function of morality, in *A World Without Values* (eds. R. Joyce & S. Kurchin). Springer Press.

Williams, B. (1981). Internal and External Reasons, in his *Moral Luck*. Cambridge: Cambridge University Press.