EXPRESSIVE FICTIONALISM

by

SIMON LOWE

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy

College of Arts and Laws

The University of Birmingham

January 2018

ABSTRACT

In this thesis I will present a theoretical, non-realist account of moral language which I call 'Expressive Fictionalism'. Expressive Fictionalism is a combined approach involving semantic content based on Joycean revisionary fictionalism and pragmatic expressivism with influences of projectivism and the quasi-realism of Simon Blackburn. The result is a marriage between the two which ultimately works towards mutual advantage.

The aim of this thesis is to provide a non-realist account of moral language in the form of expressive fictionalism, which, I posit, can explain a form of moral communication, on both a semantic and a pragmatic level, without compromising its own non-realism in the process, which avoids issues which are associated with the Frege-Geach problem and which is a non-error theoretic account of moral discourse. My methodology is a combination of semantics, pragmatics, thought experiment with some influences of empiricism, references to modern studies in behavioural & cognitive psychology as well as historical analogy. The thesis ultimately rests on a portrayal of moral language as a means of communicating emotions, as well as projecting these emotions onto the world through moral utterance, using a constructed 'morality narrative' as a contextual platform for self-expression.

Table of Contents

Section 1 – Non-Realism, Expressivism & Quasi-Realism

Section 2 – Quasi-Realism vs. Moral Fictionalism

Section 3 – Expressive Fictionalism

Section 1 – Non-Realism, Expressivism & Quasi-Realism

Chapter 1: Introduction(s)

"Kicking cats is wrong!" "It is good to stop people from doing things like that!" These are moral sentiments that (for the most part at least) we take for granted and are devoid of controversy, and yet the question of what exactly these kinds of utterances actually **mean** is impressively difficult to answer philosophically. It is equally difficult to answer what exactly is going on linguistically, semantically or pragmatically. What exactly am I doing when I declare with no small degree of passion and conviction that it is wrong to kick the cat? Am I stating a fact? Am I making people aware of some important feature of the world and about the nature of kicking living organisms? If not, then what am I communicating? Am I trying to express some attitude towards kicking living things? How does that work? Are these attitudes expressed semantically? Am I even making an assertion at all when I say things like that? If not, then why does it sound like I am? Am I deceiving people in speaking like this? Am I deceiving myself? Are these adjectives of right and wrong, of good and of evil referential? Do they refer to natural qualities inherent in the world, or are they more like projections, the result of psychological constructions in response to perceived realities? What exactly is going on?

There are two primary directions that we can go in order to answer questions like these within philosophy of language. We can examine the case for what we call 'moral realism', in which we take a stance of saying that these moral adjectives refer to qualities of right and wrong found in the world and in certain kinds of behaviour. This then leads us on a long and difficult quest of figuring out exactly how these kinds of qualities operate metaphysically, what constitutes them and how these moral 'facts' may be accessed epistemologically. Alternatively, we can examine the case for doubting the existence of qualities of right and wrong, proceeding on the basis that these kinds of qualities **either** simply do not exist or that we simply cannot have any epistemic access to them even if they do; this is what I call throughout the thesis 'non-realism'. If we take a non-realist approach to the question of meaning in moral utterances like 'kicking cats is wrong', then we must account for meaning without relying in any way on metaphysics or epistemology, but purely through philosophy

of language and on the psychology of moral speakers, and this is what I will be doing over the following three sections.

The aim of my thesis is to present a fully functioning, non-realist account of moral language. To do this, such an account needs to do the following things:

1. It must account for the semantic meaning of moral utterances like 'kicking cats is wrong!' That is it must be able to explain the content of that sentence and what the literal meaning of the words are.
2. It must explain what is being communicated in its entirety. This may involve going beyond semantics and into pragmatic discourse; that is to not only explain what is being said, but how, why and what the deeper meaning is in virtue of the context in which it is uttered.
3. It must be **non**-realist in the sense that it does not rely on any additional metaphysics or epistemology. This includes commitments towards nihilistic positions on moral ontology; that is it does not rely on the idea that moral qualities do not exist anymore than it relies on the idea that they do.

In this section, 'Non-Realism, Expressivism and Quasi-Realism', I begin by examining the case against moral realism. I do this by looking at arguments like Mackie's argument from queerness and Harman's argument from explanation, and whilst I do not set out to defend those arguments per se, I use them as notable examples of the fact that we do indeed have sophisticated philosophical reasons to doubt moral realism. I then delve into non-realism. In chapter three, I outline an important difference between non-realism and anti-realism and also what is meant by terminology which we will come to rely on later on, namely: non-cognitivism, expressivism and projectivism. In these two chapters, I will also highlight the contested division in philosophy regarding the complexity of making a clear distinction between moral language and moral ontology, arguing ultimately that we can indeed make that distinction.

I then briefly go over some of the non/anti-realist approaches to moral language. In particular, I look at error theory and explain some of the issues that I have with this kind of anti-realist approach. I then begin to explore expressivism. The version of expressivism that I am most interested in within section one is quasi-realism, specifically the work of Simon Blackburn.

This is because, on the face of it, quasi-realism seems to offer the non-realist everything that they need. Blackburn claims to circumvent the Frege-Geach problem (a serious problem for expressivism in general), and in addition to gain entitlement to give moral realists most or all that they want to hear about the nature of moral language as well (this will later prove problematic for Blackburn), including an account of moral fallibility, moral improvement and even moral truth conditions all within the non-cognitivist lens of projectivism. Throughout chapters five through to seven, I explore Blackburn's solutions to the Frege-Geach problem and how his special logic of attitudes mirrors the language of assertion and more formal forms of logic, including an interesting account of rules like modus ponens.

In chapter eight I begin to look at some of the problems with quasi-realism, in particular those argued by Andy Egan, who challenges Blackburn's account of moral fallibility. This will eventually become a recurring theme in objections to quasi-realism in which Blackburn is challenged on his ability to demonstrate attitude inconsistency in a way that is logically significant in the same way that contradiction in propositional or other forms of formal logic is. In chapter nine, we examine a relatively recent paper by Nicholas Smyth in which we see a defence of quasi-realism in the form of 'resolute-expressivism'. We then leave section one with a fairly positive outlook on quasi-realism as being the account that we are looking for, but with some worries, namely that it is not entirely clear that quasi-realism genuinely demonstrates attitude inconsistency, and if it does whether or not this is quasi-realism rather than merely a particularly abstract form of realism.

The following two sections rely a great deal upon this section. Throughout the beginning of section two, I examine some of the more problematic objections to quasi-realism including those of G.F. Schueler, Van Roojen and David Lewis, as well as some new objections of my own. I eventually conclude that there are some seemingly insurmountable problems for quasi-realism, which I outline in chapter five. However, there are some very important things we should take from quasi-realism (and from expressivism and projectivism more broadly) about the idea of communicating attitudes of approval and disapproval and about how moral speakers project qualities of right and wrong onto the world and onto people's actions through moral utterance. These may also be accounted for pragmatically rather than semantically. Nevertheless, I argue that we still need an accounting of the semantic **content** of moral utterances, and it is within chapter five that we are also introduced to another non-realist theory of great interest: moral fictionalism.

Throughout the second half of section two, I look at moral fictionalism as championed by Richard Joyce, in which we are introduced to more useful terminology and concepts like the tacit operator, pretence assertion and critical context, as well as the difference between revolutionary (or revisionary) and hermeneutic fictionalism. We are also given an interesting account by Joyce of some of the emotional/pragmatic motivations of fictionalist speakers. However, I argue that this account is underdeveloped and that it can be significantly augmented if we combine it with what we have previously learned in section one from expressivism, projectivism and quasi-realism, but only if we apply these pragmatically through an account of context, thus placing them **around** fictionalist semantics rather than replacing fictionalist semantics with expressivist semantics vulnerable to a lack of accounting for attitude inconsistency as well as the Frege-Geach problem. I demonstrate the strengths of this combined approach in chapter nine when I argue against Jonas Olson's objections to revisionary fictionalism. In doing so, I take an approach not unlike Smyth's style of defending quasi-realism, albeit by modifying it slightly in the process, but unlike Smyth I argue that these are not changes per se so much as **developments** of the original theory, and I call this combined approach '**Expressive Fictionalism**'.

In doing this I argue that there is little more to say on semantic content, but there is a great deal more to say on the pragmatic act of speaking in moral terms in the form of useful fiction. Throughout section three, I apply a pragmatic expressivist approach to fictionalist semantics. I begin by explaining what exactly I mean by 'pragmatic discourse' and the significance of context in communicating things other than what semantic content implies. I then attempt to provide some scientific basis for establishing a connection between emotional sensitivity and moral language throughout chapter three, in which I look at correlations between attitudes and behaviour with neurobiological phenomena in the ventromedial prefrontal cortex, observations of behavioural anomalies in humpback whales and case studies and testimonies of severe depression and anti-depressant medications, specifically of selective serotonin reuptake inhibitors, concluding that there is indeed enough empirical evidence to suggest a potent correlation and perhaps even a causal relation between emotions and moral thought and language. Throughout chapters four and five, I then discuss the role of emotions in the construction of moral language including a ten stage inferential process by which a fictionalist observes an event, reacts to this event emotionally and speaks out for or against it, communicating emotions and attitudes using a morality fiction as a contextual platform. This

process includes two separate models working in parallel with each other. The first is based in social positioning, and the role of emotions like pride and shame, and the formation of a concept of moral responsibility. The second model augments the first model by portraying the ten stage process mentioned earlier, using empathy or compassion as the primary mechanism.

The remaining chapters of my final section deal largely with critical contexts and the problem of deception and with the idea of applying expressive fictionalism to narratives of moral progress, in which I look at Catherine Wilson's account of unidirectional narratives and historical comparisons with progress within the natural sciences. Finally in chapter nine, I apply what we have learned by constructing an account of moral progress in human history as an example within the context of the morality fiction.

In my last chapter, I conclude that the aim of the thesis has been satisfied and that the primary goals outlined in this chapter have been fulfilled.

Chapter 2: The Case against Moral Realism & the Complexity of the Language/Ontology Distinction

In this chapter, I will give as concise and comprehensive an account as I can of moral realism and why it will not be a central focus of my thesis. To put it succinctly, this boils down to the following issue: if we look at moral realism as a semantic claim (that moral sentences assert moral facts), then if there are issues with either the metaphysical or the epistemological foundation upon which this is based, then this will create problems for that claim. In other words, if there is a case to be made against either the existence or accessibility of moral facts, be it metaphysical or epistemological, (e.g. Mackie's argument from queerness or Harman's argument from explanation), then the semantic claim immediately becomes problematic, because we cannot prove that moral assertions can be **true** if the moral facts which they refer to are either inaccessible or non-existent. However, an additional problem lies in making the distinction in the first place between moral language and ontology. This is because a key point of contention between realism and anti-realism is how the linguistic and metaphysical issues of both may interact to constrain the shape of overall meta-ethical theories. This raises some extremely complex questions about which explanatory burdens exist regarding the metaphysics that arise for some accounts of the linguistic issues, and the linguistic issues that arise for some accounts of metaphysics. In this chapter I will use Harman's argument as an example of the kind of problems with moral realism that can arise and thereby use that as a springboard towards a conversation about moral language.

It should also be noted that this is an example; what I do not want to do is to **defend** Harman's view per se, and it should also be noted that I am aware that moral realists do have strategies to counter this kind of argument. Rather, my purpose is to use this example to reveal something very strange about moral language, and that is that whether or not the realist claim holds up, moral language almost always comes across as realist. What I mean by this is that moral sentences very rarely appear to take the form of opinions (e.g. 'that shirt is pretty cool, but that's just my opinion'); rather they far more often resemble statements of fact ('you cannot do this; this is wrong!'). If we assume for the sake of argument that, for example, Harman's argument holds up, then that leaves us with a very puzzling conclusion: given that we have reason to doubt the existence of moral properties, why do we continue to talk about morality in a seemingly realist way? This is, fundamentally, the question I wish my thesis to be an investigation into. I will later discuss in more detail the complexity making the

distinction between moral language and moral ontology in the first place, arguing ultimately from a post mackian point of view that we can indeed (and would do well to) make that distinction.

First, let us look at moral realism **very** broadly. In its broadest form, moral realism describes a general belief that moral sentences, that is sentences like, "kicking cats is morally wrong!" express propositions which refer (or attempt to refer) to objective features of the world, in other words they are independent of anyone's subjective attitude towards them[1]. Furthermore, at least some of those propositions are truth apt in that they have real truth conditions. In simpler terms, moral realism usually manifests as a general belief that one may say, "kicking cats is morally wrong!", and be objectively 'correct' or 'normative' (in point of fact) in saying so regardless of anyone's **opinion** on the matter, including the subject's (what exactly these notions of 'normativity' or 'correctness' mean, however, are a point of considerable debate between moral realists). By definition then, moral realism stands in stark opposition to all forms of non/anti-realism, such as non-cognitivism, moral nihilism, moral scepticism, ethical subjectivism and error theory. To make things very clear and simple here, let us then take moral non/anti-realism to be, by definition, that which does not embrace moral realism. In other words, it is a belief that either one cannot say that, "kicking cats is morally wrong!" with any objective accuracy, one cannot say it whilst genuinely referring to any objective features of the world, or both. To be a non/anti-realist is, therefore, to believe that at least one or both of these two essential components of moral realism are false, and to be a moral realist is to believe that both of these components are true. I will also, later on in this section, make an important distinction between 'non' and 'anti' realism.

To clarify to ourselves what exactly we are looking at, let us divide what we are looking at in moral realism into two separate claims. The first is a metaphysical claim about nature: there are things in nature that are intrinsically normative/correct and these features of the world are objective. The second is a semantic claim: when we talk about morality, we are, when we speak accurately/truly, directly referring to these objective features of nature[2]. Of these two claims, it is the second that we should be particularly interested in. The first claim is what we

[1] Objectivity and subjectivity are themselves famously problematic philosophically. For the purposes of this chapter, let us simplify things by treating 'objective' as meaning something which exists independently from the observer and 'subjective' as meaning something which results **from** the observer performing the observation.
[2] Once again, I would like to point out that this distinction is a contested issue in philosophy, something which I will address later on in this chapter.

may call an **independent** claim. In other words, it is a claim, no more or less, than about the way the world is. More interestingly, however, the second appears to be a much more **dependent** claim. Not only is it a claim about the nature of moral language, but it is a claim based on the assumption of the first claim (the one about the nature of the world). To avoid any unwanted confusion here, let us interpret the two separate claims like so: the first claim may be expressed like this, "There are objective moral facts", and the second claim, may be expressed like this, "Utterances about moral wrongness attempt to refer to objective moral facts". Under this interpretation, it seems that the second claim is dependent on the first claim. However, what happens then if we negate the first claim? For example:

1. "There are **no** objective moral facts"
2. "Utterances about moral wrongness attempt to **refer** to objective moral facts"

Already a problem arises. Under this analysis, it seems clear that there is something amiss in terms of moral language. If we are to believe J.L. Mackie's perspective for instance, we may be tempted to adopt a kind of moral nihilism, specifically 'error theory', in which we conclude that (1) is true, and that (2) betrays some kind of linguistic confusion or falsity. Before we get carried away with that, however, let us first consider why we might be tempted to adopt (1) in the first place.

To do this we could look at either Mackie's argument from queerness against moral realism, or alternatively we could look at Gilbert Harman's argument from his 1977 text, *The Nature of Morality: An Introduction to Ethics*. Here Harman uses abductive reasoning to argue against the existence of objective moral facts. According to Harman, we have little reason to suppose that objective moral facts do exist, because there is nothing about moral judgements that cannot be explained in non-moral terms (Harman, 1977, 4). To demonstrate why he thinks this, Harman asks us to consider the following thought experiment. Suppose a physicist is trying to test a scientific theory by observing a vapour trail in a cloud chamber. What you or I do not know is that the physicist is testing for the presence of a proton particle. As a result, what you or I see is a meaningless trail of vapour. In the absence of theory, evidence is meaningless. In order to prove that the proton is indeed there, and creating the vapour trail, we must first **hypothesize** that the proton exists and that this hypothesis **explains** the observation of the vapour trail (Harman, 1977, 6). This **abductive** reasoning, or inference to the best explanation, is a common feature of scientific method. It is not,

however, as Harman argues it, so obviously applied when explaining moral judgements (Harman, 1977, 6).

For example, suppose we observe the obscene act of a group of youths kicking a cat[3], react with revulsion, and conclude that the reason we react to it the way that we do is because the act is morally wrong. It is not so clear in this case that there is the same kind of explanatory work being done here (Harman, 1977, 5). Let us suppose for the sake of argument that there is. Let us suppose that we formulate the hypothesis that the reason we react negatively to this kind of behaviour is because this kind of behaviour constitutes something 'wrong'. In other words, something intrinsically or objectively 'incorrect' or 'anti-normative' about the way the world is or **ought** to be. Does the evidence fit the hypothesis? More importantly, does the hypothesis explain the evidence? Not really. Suppose for instance that there is something similarly 'wrong' (intrinsically/objectively), in the act of throwing a woman into an arena to be torn apart by lions. According to our hypothesis, if this is similarly wrong, then this should provoke a similar reaction, and perhaps today it would. However, we have plenty of historical evidence to believe that this behaviour has not always provoked this same negative reaction. To the contrary, execution by *objicĕre bestiis* (a defenceless victim devoured by beasts), or by *damnatio ad bestias* (where the victim is expected to fight the attacking beast) were, among others, common amusements to the fifty thousand or so Roman spectators who regularly flooded the Colosseum (Tacitus, AD 16-48, 44). So clearly this hypothesis is not **sufficient**, because clearly it does not tell us why we tend to **disagree** about what behaviours are right and wrong. More importantly, the hypothesis itself may be flawed, as we can provide numerous examples of seemingly 'wrong' or 'evil' behaviours that do not provoke any negative reactions whatsoever.

Suppose alternately then that we do as Harman does and come up with an alternative hypothesis, that the reason we react negatively to the cat kicking scenario is because A – we observe it, and B – we feel compassion for the animal that is suffering (Harman, 1977, 5). This hypothesis has an obvious advantage in that not only does it explain why we react negatively to the kicking of cats but also why we might similarly react negatively to a person being torn apart by a lion in the Colosseum, and yet an ancient Roman might not. A

[3] The example Harman actually uses is of them setting the cat on fire, but as I use the marginally less **horrific** 'cat kicking' scenario many times throughout the thesis as my go to example of a morally repugnant act, I shall stick with that example here as well.

behaviour might be reacted to negatively by one party, as it is observed and the object of the act may be empathised with; conversely a behaviour might not be reacted to negatively because either it is not directly observed, or, in the case of the Roman(s), the object of the act may not be cared about. Most importantly, however, this later hypothesis is notable in that it refers to **non**-moral features of behaviour and does not refer to any moral property. Returning directly to Harman's account, Harman argues that the existence or lack of existence of moral facts is in fact **irrelevant** to explaining our observing, responding and judging of behaviour (Harman, 1977, 9). To the contrary, it is the non-moral features of the behaviour that are all that are required in fitting the evidence with the theory. Given then that the question of the existence of moral properties is an irrelevant one in respect to us explaining why we react negatively to some behaviour(s) and positively to others, at least as far as Harman is concerned, why then should we suppose at all that these moral facts do exist?

Whether or not then there are any issues relating to the independent metaphysical claim of moral realism, there seems to be an epistemological issue to consider as well. Supposing, for the sake of argument, that moral properties do exist. Can we access the kind of moral knowledge we would need to explain why some actions are evil and yet others are not? In other words, can we know not only that, but **which** moral facts exist? If we believe Harman's argument, then it seems the answer to that question is no, because according to Harman, there is no abductive **inference** we can use to connect the existence of any alleged moral properties to the phenomenal evidence of moral judgement and moral discourse. However, this leaves us in a very confusing place.

When addressing moral language, we do not speak in a way that reflects a hypothesis about observation or empathy. We do not say, "I disapprove of kicking cats", we say, "Kicking cats **is wrong!**" This clearly seems to be a realist sentence. J.L. Mackie famously responded to this issue in 1977 in *Ethics: Inventing Right and Wrong*, when he coined the term 'error theory' (Mackie, 1977, 18). To explain what Mackie means by this, let us look again at an earlier example:

1. "There are **no** objective moral facts"
2. "Utterances about moral wrongs attempt to **refer** to objective moral facts"

Suppose someone is at odds with moral realism on the basis that whilst they see no special problem with it as a semantic claim (moral sentences refer to moral facts), they do have issues with moral realism as a metaphysical claim, due to an epistemological argument e.g. Harman's explanation argument. There is surely some kind of error here, because the semantic realist claim is dependent on the independent metaphysical claim, which we have epistemological reasons to doubt. Does it make sense to say that moral sentences refer to moral properties and to say that moral properties do not exist to be referred to? According to Mackie, the only conclusion we can draw here is that our second claim must also be **false**. When we use a realist sentence in a moral discussion, this is either a reflection of the fact that we believe the metaphysical claim but are mistaken, or, we do not believe it and are simply speaking falsely. Either way, we are in error in using this kind of moral language. This is what is meant by '**error** theory' (Mackie, 1977, 18).

However, can we even make this distinction? Throughout this chapter, I've separated moral realism into two claims[4]:

1. Moral language is descriptive.
2. Something in the world corresponds to that description.

Anti-realists like Mackie see these as two separate claims, accepting (1) and denying (2), the first being about language and the second being about ontology. Many realists, on the other hand, do not think that these two claims are as separable as philosophers like Mackie make them out to be. Whilst it is generally agreed that the two claims are distinct, a key point of contention is how the linguistic and metaphysical issues may interact to constrain the shape of overall meta-ethical theories. In other words, how do some accounts of the linguistics bear on certain metaphysical issues, and vice versa? What explanatory burdens regarding the metaphysics arise for some accounts of the linguistic issues? For example, say we look at this within the context of something like direct reference theory (Almog and Leonardi, 2012, 79). Already we run into serious problems, because according to this theory, meaning attributes its origin to what the word or the expression points to in the world. In other words, meaning is inescapably linked to reality, at least on some level. How can we make sense of language

[4] Notice that I have now switched the two claims around, this is because previously, we were examining how the metaphysical claim is independent and the linguistic is dependent. However, under certain ways of looking at them, the reverse can also be true; 'something in the world corresponds to a description of x' **depends** on 'x is being described'.

which fails to refer to reality in the first place? [5] One of the most substantial problems we face here is that realists want to connect how moral language is used with what morality ultimately is (metaphysically). This, of course, places certain explanatory burdens upon the realist to then provide for this connection with additional layers of theory based in metaphysics. Someone who claims, for instance, that moral predicates denote (non) natural moral properties, and that there are such properties, will need a meta-semantic story about how moral predicates come to have such properties, as opposed to other properties as their meanings. As a result, we ought to be careful here when discussing realists (in general) as like many other fields, how this actually manifests is both vast and varied. Reference magnetism (as found in Dunaway and McPherson, 2016) for instance, attempts to account for how moral properties come to be the meaning of moral predicates. There are objections to non-naturalism in the direction of there are no plausible meta-semantic stories of how moral predicates could come to denote non-natural, causally non-efficacious properties, and so on[6] (Baldwin, 1985, 23-45).

However, things are by no means easier for the anti or non realist either, because on this side of the spectrum we see all of the linguistic issues associated with language about 'something' (morality) which is not regarded as real in the first place. Again, whilst realists want to connect how moral language is used with what morality ultimately is (metaphysically), non and anti-realists want to do something different, and what that is can get extremely confusing both metaphysically and linguistically. So even for this side of the spectrum, it is highly controversial to what extent facts about the conventional meaning of language have implications regarding the metaphysics of moral properties/truths, and vice versa. For instance, if we remember our Bertrand Russell and the subject of definite descriptions or the conundrum of whether or not the present king of France is bald (Russell, 1905, 479) or if we delve even further into the works of Stephen Neale and existential quantifiers (as found in Neale, 1990), or if we look at Strawson's work with Fregean analyses (Ostertag, 1998, 3-4),

[5] This is exactly the kind of issue which will come back to haunt us in section two when we discuss issues like the problem of fictional discourse in respect to moral fictionalism (see section 2 chapter 8).

[6] Ethical non-naturalism is a non-definist form of moral realism as well as a form of cognitivism. It claims that:

1. Moral predicates express propositions
2. Some of those propositions are true
3. Those propositions are determined by objective features of the world
4. These features are not reducible to any set of non-moral features

Pioneers in ethical non-naturalism include G.E. Moore, one of the founders of the analytic tradition in philosophy alongside Bertrand Russell, Ludwig Wittgenstein and (prior to them) Gottlob Frege (Van Der Schaar, 2013, 8). This definition can be seen as compatible with non-reductive naturalist realism.

then we will quickly realise how talking about things which either by definition do not exist, or in the case of moral non-realism where the existence of moral properties is, at the very least, highly suspect, then this can be a pretty substantial can of worms in and of itself. If 'wrong' is not real, then what is my sentence about kicking cats being wrong doing exactly? In the case of certain kinds of non-realism, for instance quasi-realism, which we will discuss later on in this section, it can get even more confusing. For a quasi-realist for instance, 'wrong' is described as a kind of psychological construct, projected onto the world through moral utterance rather than any property of our environment. However, this surely demands some kind of semantic explanation concerning its reference, which proves to be extremely problematic as the sentence is not even asserted but expressed, as highlighted by the Frege-Geach problem (see section one, chapter five). In sections two and three when we deal with moral fictionalism as well, issues like this will become even more important when we attempt to deal with things like presupposition failure and the problem of fictional discourse (see section two, chapter eight). Ultimately, this will become a recurring issue which will be faced time and again throughout this thesis. Suffice it to say, for now, that we do need to be extremely careful when delineating linguistic and meta-semantic issues about moral language vs. metaphysical issues about the nature and existence of moral properties. We would also do well to outline, in no uncertain terms, what commitments on the linguistic issues, the metaphysical issues, and the dialectical connections we are defending and why.

Unfortunately, what I cannot do is give both the anti-realists and the realists **everything** that they want, because that would be somewhat akin to violating the law of the excluded middle[7]. Ultimately, either moral properties are real or they are not. Quasi-realists like Simon Blackburn try very hard to find some middle ground on this issue, incurring very substantial explanatory burdens regarding the meta-semantic and metaphysical stances which they hold. By contrast, I do ultimately want to make clear where on the spectrum of realism and anti-realism my account lies in relation to existing literature. Fortunately, referring to this realism/anti-realism divide as a spectrum is useful here in that it allows for some leeway to say that whilst my account very much does lean towards the anti-realism side, my approach is nevertheless different from most in that I leave the door open for the realist by committing to

[7] The law of excluded middle states that for any proposition, either it is true or its negation is true. It is one of the three laws of thought dating back to Plato (Boole, 2003, 3-4), alongside the law of non-contradiction and the law of identity. It is, however, not without considerable controversy. For example, criticisms in logic and mathematics contest this, claiming rather that a proposition is either true or it is not able to be proved as true (Clark, 1978, 292-322).

avoiding the claim: 'there are no moral properties'. There is an absolutely vital distinction that must be made here. There is an important difference between claiming that an explanation is wrong vs. providing an alternative explanation. In other words, where anti-realists like Mackie accept (1) (moral language is descriptive), and reject (2) (there are moral properties), I do something entirely different. I reject (1), and withhold judgement on (2). In terms of the linguistic component of moral realism, that moral predicates refer to moral properties, this is something which I do **reject**, at least in terms of when we are talking about a particular kind of moral utterance used by a non-realist[8]. In this sense, my account stands opposed to error theory in many respects, as I do claim that there is **not** a pervasive error being committed by the non-realist who uses realist 'sounding' moral predicates. In terms of the metaphysical component of moral realism, however, that there are moral properties, that is something which my account withholds judgement on. However, the very way I approach this issue is nevertheless through that post mackian paradigm of treating moral language and moral ontology as two separate issues, because it is only through making that distinction that what follows throughout the rest of this thesis ultimately works.

In the following sub chapter, '2.1. In Defence of Non-Realism', I discuss my own concerns about moral realism as a metaphysical claim which mirror those of Harman's argument from explanation. Like Harman, my concerns are less with the pure metaphysics than they are with the epistemology surrounding them; I see no abductive inference at least to suppose the existence of something explaining observations for which we already have much simpler and more evident explanations for. However, where I differ from Harman is that I do not then actively suppose that moral properties do not exist, but rather leave the question unanswered. I do not violate the law of the excluded middle either, because in order to do so I would need to answer the question of whether or not moral properties are real with something other than a yes or a no, which I do not do; I simply do not answer such a gigantic question as this, leaving it firmly in the hands of other philosophers. The distinction that I make between moral language and ontology is, thus, also a theoretical one. Assuming, for the sake of argument, that we cannot rely on real moral properties to help us out; can we still explain moral language in a meaningful way? I argue that we most definitely can. However, as mentioned before, this still leaves us with a problem, because even if we make that

[8] Later on, when we discuss moral fictionalism, I will discuss the phenomenon of 'pretence assertion'; this means that moral language is treated 'as if' it is descriptive, but in reality, it is not. Later on than that, when I discuss my own 'expressive fictionalism', I will explain how emotion states may be expressed and communicated through this phenomenon of moral fictionalism (see section 3 chapters 4 and 5).

distinction, if we have reasons to doubt the existence of moral reality, then how can we make sense of moral language, which sounds realist, in anything other than a theory of pervasive error?

I will conclude this chapter then with a simple question. Is this really the conclusion that we want to arrive at? If not, then it seems we have the following options. We can either attempt to refute Harman and Mackie's arguments in an attempt to save moral realism, or we can attempt to tackle the question of realist moral language in the (assumed) absence of real moral content **directly**. In other words, we can attempt to answer the following question: assuming we do not believe that there are any moral facts, and assuming we are not confused or lying, why do we continue to talk about morality 'as if' we do believe that there are moral facts? In other words, why do we continue to say, "Kicking cats is wrong!" instead of "I disapprove of kicking cats!" To do this, we should see if there is an account of moral language that can explain **why** we speak about morality in a seemingly realist way that is **independent** (does not depend on a separate metaphysical claim), and which retains explanatory power (retains the ability to explain why we react negatively to certain behaviour(s) and why we disagree about them), and does all this without a reduction to error theory. In so doing we should find an account which is immune to Mackie and Harman's arguments. So to recap, what we are looking for is an independent, linguistic account of seemingly realist moral language, which has explanatory power and is not reducible to error theory. Throughout the rest of this thesis, I will look at two candidates for this in particular, both of which are non-realist and non-cognitive: expressivism (specifically in the form of Blackburn's quasi-realism), and later in section two, moral fictionalism as championed by Richard Joyce.

Chapter 2.1. In Defence of Non-Realism

In chapter two I made reference to Harman's argument from explanation as a go to example of the kinds of argument that may provoke us to doubt moral realism as a metaphysical claim. In the interest of solidifying this premise, I would also do well to offer up some of my own concerns about moral realism as a metaphysical claim. Like Harman, my concern is less with metaphysical arguments for the existence of moral properties per se and far more with the epistemological dilemma of how we may have access to them even if they do exist, and whether or not we really need to do so. For me the dilemma seems to be rooted in moral disagreement. There are numerous theories about how we may 'define' what is right, and what is wrong, for instance from utilitarianism, deontology or from Aristotelian virtue ethics. Ironically, this is not unlike some of the deeper levels of material science where there are multiple models that may explain the nature of the universe both at a cosmological and at a subatomic level, and like debates in philosophy about the 'true' nature of morality, there are similarly lengthy debates among physicists about whether or not string theory is an accurate model of subatomic particles, or whether the big rip is a probable outcome of dark energy pushing galaxies further apart. Like Harman, however, I would argue that there is an important difference between these kinds of debates and that is the process of fitting theory with **evidence**, and about the use of **abductive** reasoning.

For example, take the idea of 'moral progress'; say we attempt to explain why people have disagreed about what constitutes 'good' and 'decent' behaviour throughout history by arguing that some cultures have had a closer, or more sophisticated understanding of the nature of morality than others. For me this is extremely problematic for the simple reason that there is a much simpler explanation, which fits the observable evidence much better, and that is that a society's system of values is shaped not by eternal truths wrought by the laws of nature, but by human **attitudes & sentiments**, which are highly suggestible. In other words, 'right' and 'wrong' resemble less a collection of mathematical constants, like the speed of light or the laws of thermodynamics, which (assuming our current model of physics is accurate) physically cannot be broken, but rather more a collection of mutually (or not) agreed upon principles, which people are encouraged to uphold, but of which the underlying reality is far more questionable, and which are easily violated by individuals who either fail to understand them or simply choose to break the rules.

One seems to accurately reflect the way the universe **is**, whether we like it or not; in other words, whether or not we agree with the laws of thermodynamics, those are the laws that we are subject to in this universe. We may have inaccurate scientific models of course, at which point we rework our theories in the light of new evidence, but the underlying principle is that there exists a set of laws by which the universe operates, laws we **cannot** disobey, and which we access and gain knowledge of by applied mathematics, empirical evidence and abductive reasoning. The other seems to accurately reflect human needs and emotions, whereby we are driven by certain biological imperatives necessary for our survival, such as food, shelter or reproduction, but which must also be applied in a **social** environment. Human beings are not solitary animals, but rather live in often very large, highly complex social networks and systems built on interdependency. As such, we shape our psychological attitudes to ensure not only our own survival but also the survival of those that we depend upon. What is also very interesting is that in multiple animal species that live in social groups rather than as individuals, the needs of the group is often as important, or even completely overrides the needs of the individual[9]. This theory far better seems to fit the **evidence** observed in the biological sciences rather than it does moral realist explanations about moral disagreements.

Now, where does my view differ from Harman's? In chapter two I discussed the complexity of separating moral realism into one linguistic and one metaphysical claim via that kind of post mackian paradigm. By doing this I do place myself firmly towards the anti-realist side of the spectrum, but what I also hasten to add is that unlike Mackie or Harman, what I do not seek to do is to use my scepticism of moral properties to single handily take down moral realism. I would claim that it would be either difficult or even impossible to do so. In the following chapter, I will describe my non-realism as akin to a kind of **agnosticism** on the issue of whether moral properties are real or not. I would claim that moral properties are unknown, and in all likelihood unknowable. This does not mean that they are not there, but it does mean that we do not have access to them. Therefore, we are left with the following options. Either we can attempt to find access to them via epistemology, or, we can attempt to

[9] Humans are by no means the only species to exhibit personal sacrifice, nor does it seem entirely accurate to say that reproduction of the individual's genes is the only biological imperative. It is also clearly evident from individuals who choose not to reproduce that there exist other motivations besides continuing one's own genes. However, what does appear constant (in all social animals) is the need to preserve the continuation of the social group, even if that requires the sacrifice of the individual. In particular, the needs and survival of offspring seems to be of paramount of importance. An example I use in section three on this very issue of animal 'altruism' is a documented case of a pod of orcas hunting a grey whale calf and some fascinating behavioural anomalies in some nearby humpback whales (see section 3 chapter 3).

see if we can make sense of moral language **without** moral properties, the latter of which is what I am doing.

So in summary, my concern is very similar to Harman's argument from explanation in that there simply does not seem to be any special need to commit to the metaphysical claims of moral realism when there is a much simpler explanation for moral disagreements, one which far better fits the observable evidence provided in the biological sciences, and which has much more direct epistemological access to it by way of abductive reasoning. In other words, I am applying not only Ockham's razor (choosing the simpler explanation over the more complex one), but also inference to the best explanation in lieu of what we can **directly observe** in nature. However, my account is also markedly different from both Harman and Mackie in that this is as far as I go in this particular direction, committing as I am to avoiding the claim 'there are, therefore, no moral properties'. In other words, I attempt to leave the door open for moral realism by rejecting that moral predicates refer to moral properties (in the case of the fictionalist at least), but entirely withholding judgement on whether there are (or are not) moral properties. Much later on, in section three, I will develop this line of thought much further, and attempt to provide some theoretical accounts of how psychological attitudes may be translated into the kinds of 'moral' imperatives that we use in everyday situations. In my next chapter, however, I will do some useful sign posting as well as introducing something which has been alluded to more than once thus far, but has yet to be fully explored, and that is this notion of 'agnosticism' in regards to moral ontology.

Chapter 3: Non-Cognitivism, Expressivism, Projectivism and the notion of Agnosticism

Having now looked a little at moral realism, let us now look more closely at both non-realism and anti-realism. The purpose of this chapter is not to provide any arguments for or against any particular, pre-existing philosophical perspective as such, rather it is to offer some useful sign posting for what is to come, as from this point on I will be relying on a lot of non-realist terminology, and, therefore, I would do well to illustrate what these terms mean (as I will be using them). Before I delve into the terminology, however, I will first explain a very important difference between two critical concepts which I will also come to heavily rely upon in later chapters, and that is the aforementioned difference between an **anti**-realist perspective and a **non**-realist perspective. Anti-realism, in addition to standing in stark opposition to moral-realism, attaches claims about moral ontology (there are no moral properties). By contrast, non-realism differs from this slightly. Rather than declaring that the claims made by moral realists are false, a non-realist perspective is one which simply **withholds judgment** about the metaphysical claims (there are moral properties), creating a theory without relying in any way on those kinds of claims being true or false. In other words, rather than taking an 'x is incorrect' position, the non-realist instead takes a position akin to saying, 'we have not, or cannot, establish that x is correct'. One makes an active counter claim about moral ontology in response to the ontological claims of moral realism, whereas the other merely regards those claims as unverified. In other words, anti-realism is the claim that, 'morality does not exist'; non-realism is the claim that, 'we have not, or cannot, verify the existence of morality'.

There is a similar difference between theological positions like atheism and agnosticism. For instance, an atheist may claim, 'I do not believe in god', and in most cases, the negation is also true, 'I believe there is no god', thereby maintaining the law of excluded middle. God either does or does not exist, there is no middle ground. However, the spectrum of belief is not without varying degrees of certainty and confirmation. An agnostic, for example, may claim, 'I am **undecided** as to whether or not there is a god'. Again, this is not the same as answering a binary proposition (is god real or not?) with something other than a yes or a no, thus violating the law of excluded middle. Instead, this position leaves the proposition entirely **unanswered**. In other words, there is an important difference between rejecting a

conclusion and embracing its negation vs. simply not embracing either the conclusion **or** its negation by simply walking away from the proposition. Nevertheless, I would argue that agnosticism still leans heavily on the atheistic side of the spectrum of belief, in this case, for the simple reason that to arrive at agnosticism (towards god) one has to begin from a position of scepticism; one has to at least **doubt** the existence of god in the first place. So even though agnosticism exists separately from atheism, and is distinct from it in the sense that one believes in the absence of something, and the other does not commit to any belief one way or the other, the two are nevertheless in a similar camp, as they both approach the same proposition from a position of doubt. The atheist, for instance, may well also claim that 'there is no god', is the default position accepted in lieu of an absence of any data one way or the other in regards to the existence of god. The agnostic, on the other hand, may counter this by claiming that the default position from such a situation is not to leap to any conclusions whatsoever in the absence of conclusive evidence, but rather accept that the truth is ultimately unknown.

Now how does this relate to moral realism and anti-realism? As mentioned in the latter half of my previous chapter, moral realism seeks to connect moral language with moral ontology, and contests the treating of the two subjects as distinct issues, claiming rather that moral language is **intrinsically** linked to moral reality. This is similar to how the theist believes in god, claiming that the existence of God is **intrinsic** to how the universe works[10]. The anti-realist is similar to the atheist, claiming that arguments for the existence of god do not hold up, and so the more logical explanation is that god does not exist, or, in the case of moral anti-realists like Mackie, whilst there is no issue with portraying moral language as descriptive, the notion that it is **accurately** descriptive (of moral reality) is far more doubtful; therefore, the much simpler explanation is that moral speakers are in **error**. My non-realist position, however, is a little different to both of these approaches. Whilst I agree with Mackie that moral language and ontology should be treated as separate issues, as complex as that may be, and whilst like Mackie I approach moral realism from a position of scepticism and doubt, placing myself firmly on that side of the spectrum of belief, I nevertheless believe that the more intellectually honest conclusion (on the reality of moral properties) is not to have one,

[10] Unsurprisingly, the proliferation of philosophical arguments, in all fields of the subject, on the **necessary** existence of God throughout history is truly staggering, from Anselm's ontological argument, Thomas Aquinas, Pascal's Wager or William James's 'Will to Believe', philosophers have never, and in all probability never will universally or even consistently agree on this subject; it is one of **the** most controversial issues in all of philosophy, much like the issue of whether moral properties exist or not (as found in Jordon, 2018).

not ultimately having any **conclusive** evidence with which to make one. In other words, just as the theological agnostic doubts the existence of god, but is ultimately undecided about whether to accept either the hypothesis of god's existence or the null hypothesis, the experiment being incomplete, moral agnosticism comes from a similar position in regards to the existence of moral properties. To put things as basically and succinctly as I can, non-realism and moral agnosticism come from a position of doubt in regards to moral realism; they do, therefore, place my own account (of expressive fictionalism) in a firmly separate camp from it, and unlike something like quasi-realism, the account does not try to give the hard realist everything that they want. What expressive fictionalism also does **not** do, however, is stand in direct opposition to moral realism, but rather stands entirely **alternative** to it. In other words, the thesis proceeds as follows:

1. We have epistemological reasons to **doubt** metaphysical accounts of moral properties.
2. Assuming, for the sake of argument, that we have not, or cannot, verify the reality of moral properties, moral **language** nevertheless sounds distinctly realist in nature (ergo, we are treating the two as separate issues).
3. Given these two observations, can we still come up with a theory which explains how this works, **without** simply concluding, as the anti-realist does, that there are no moral properties, or that, therefore, moral language is altogether flawed?

Without further diversion then, let us look at three more terms associated with non-realism that will come to be very useful in the next few chapters: non-cognitivism, expressivism and projectivism. First, let us look at **non-cognitivism**. All three of these concepts, as I will be treating them at least, will be specific to philosophy of language. Non-cognitivism, for instance, I am taking to mean a position that claims that moral sentences **do not express propositions** in the same way that more descriptive, fact oriented sentences do[11]. For most non-cognitivists, this also has the side effect of implying that ethical sentences are non truth apt (or at the very least, not truth apt in the same way as conventional propositional logic is). Another potential side effect of non-cognitivism, therefore, is also the preclusion of moral knowledge[12]. This is not to say at all that non-cognitivism claims that moral utterances are

[11] This is, of course, how non-cognitivism usually manifests. However, there are contemporary non-cognitivists that do accept a notion of propositions on which moral sentences express propositions (as found in Plunkett and McPherson, 2018, see ch. 1-2 and 8).

[12] This is assuming that what we are taking to be 'knowledge' requires that the thing that is known is 'true' in a propositional sense.

meaningless, but it does portray those utterances as non-declarative speech acts rather than fact asserting propositions. What these speech acts are exactly depends on the variety or flavour of non-cognitivism we are looking at, but one of the most notorious ones, and the one we will be looking at very closely from this point onwards, is **expressivism**.

Other varieties of non-cognitivism include **emotivism**, which we can find in the works of A. J Ayer (Ayer, 1936, 106) and C. L. Stevenson (Stevenson, 1937, 83), which claims that moral sentences express emotional attitudes, and is colloquially known as the 'boo! hurray! theory of ethics', or its close cousin **universal prescriptivism** developed by R. M. Hare (Hare, 1963, 10ff.) which likens ethical sentences to universalizable imperatives (those who make moral judgements are committed to those same judgements in other situations where similar facts obtain). Emotivism is particular, and as we discover later with Blackburn's **quasi-realism**, can also be seen as a form of expressivism, so let us for the moment focus on expressivism in its simplest form. Expressivism is the view that sentences which employ moral terminology, such as, 'kicking cats is wrong', are non-descriptive and that these moral adjectives are non-referential to any metaphysical or ontological properties or facts. The primary function of the moral speech act is not to make the listener aware of any particular information about the nature of things or of the 'immoral' act in question, but rather it is to express the evaluative attitude of the one speaking. In laymen's terms, it is to more akin to expressing a feeling about something rather than describing how the thing really is in (purely) objective terms. Obviously, this has a variety of potential effects on the truth value and truth conditions of a given sentence using moral terminology. Some expressivists simply claim that moral sentences are not truth conditional and/or lack any truth conditional content. Others, Blackburn in particular, seek to provide alternative criteria for special, **moral truth** conditions, and there will be more on that in later chapters. The most important thing we want to remember about expressivism, however, is that it is non-cognitivist in the sense that it claims that ethical sentences are non-propositional, rather, they are expressive, and that this is a non-realist theory of moral language. Expressivism claims something about what ethical sentences do not do in relation to theories about the existence of moral properties (i.e. refer to them in any way), and this stands in stark contrast with more anti-realist theories like error theory, which do claim something about what ethical sentences do in relation to such theories (i.e. they do attempt to refer to them, and this is an **error** on the part of the speaker due to a lack of existence or lack of access to any real moral properties).

Another important concept and one which will be vital to understand later on is the theory of **projectivism**. To simplify for future reference, this is what I call the 'as if' theory; in relation to moral language for instance, projectivism is the idea that speakers using moral terminology use those terms to 'project' certain qualities onto something 'as if' those qualities actually belong to it, even if actually they do not. As Hume describes it:

> Tis a common observation, that the mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses. (Hume, 2016, 167)

Projectivism comes in many forms and varieties as well, but in particular we will focus on the form of projectivism, as applied in meta-ethics, championed by Simon Blackburn and how he uses it in **quasi-realism**. Quasi-realism, as we will discover later, walks a very thin line between realism and non-realism. Rather than saying outright, for example, that moral sentences lack truth conditions or that there are no real moral qualities involved in the evaluation of an act as moral or immoral, Blackburn claims that moral qualities and moral truth conditions are actually constructed by the speaker, and that they are projected onto the world in a manner not unlike a light source projecting light onto a surface and creating a reflection[13] (Blackburn, 1984, 174). So quasi-realism is not a realist theory, because it does not claim things about the actual nature of ethics so much as it claims things about the people making use of ethical terms, but it is also not strictly anti-realist either, because it does not claim that those people are not talking about something which in some alternative senses could be interpreted as 'real', hence 'quasi'-realism. In laymen's terms, I would describe quasi-realism as the theory that morals are real not because they exist out there in the world, but because we create them, and the method of that creative act is projectivism. In other words, moral language acts like a projector projecting images onto a blank wall for all to interpret. So to refine our light on the wall analogy, Blackburn does present this projection (i.e. the light) as a linguistic phenomenon rather than a purely cognitive one; it takes on the

[13] Blackburn never actually uses this analogy to my knowledge, but as I understand the concept of projectivism, the analogy seems appropriate. In actuality, the analogy is not my creation either, but rather inspired by a scene in Michael Strasczynski's Babylon 5, in which a character called G'kar takes a seemingly projectivist stance on his explanation of the nature of God; a lantern represents man's search for God, creating a reflection on a wall (representing the universe) which the seeker interprets as God, not realising that the source of the reflection comes not from the wall but from seeker himself. Similarly, a person who does not carry a lantern, who does not search for God, sees nothing. (*Babylon 5*, series five episode fourteen ("Meditations of the Abyss"))

shape of realist moral language, but the ultimate source of the (moral) utterance comes from the speaker, not what is being spoken about. What is interesting, however, is in how this ties in with expressivism. Recall that expressivism is the theory that the communicative speech act of ethical discourse amounts to expressing attitudes towards things we either approve or disapprove of. This may be achieved in a variety of ways, but according to quasi-realists like Blackburn, the method used is a form of projectivism, whereby we speak 'as if' things were right or wrong, thereby creating those kinds of qualities psychologically and projecting them from within ourselves out onto the thing that we are looking at.

Later on in chapter five, I will explain this in greater detail, but before I progress onto that, I would also do well to outline some of the other forms of non-realism out there, and this will be the subject of my next chapter. I would also do well to talk a great deal about a particular threat to expressivism, and which is in no small way a contributing factor in the creation of quasi-realism, and this is the Frege-Geach problem. This, along with Blackburn's response to it, will be the subject of a great deal of discussion in chapters six and seven in particular.

Chapter 4: Alternative Non/Anti-Realist Approaches to Moral Language

Throughout section one I will be focussing on a non-realist approach to moral language called 'expressivism', in particular Simon Blackburn's quasi-realism, and in section two I will look at another approach called 'moral fictionalism', championed by Richard Joyce. Before continuing on with these approaches, however, it is worth briefly noting some of the alternative approaches that exist. In this chapter I will briefly explain some of these alternative approaches and attempt to place them within the context of the overall thesis. This is to eventually end up with a complete theory of moral language which is independent of additional philosophy in metaphysics or epistemology, and within that context why quasi-realism and moral fictionalism are much more favourable candidates for this. What I mean by this is that the **language** theory should be adoptable not only by people who are hard anti-realists in the sense that they do make the metaphysical commitment that morals are not real but also by people who are morally **agnostic**. In other words, to those who make no attempt to verify whether morals are real or not on a metaphysical level, the theory should be adoptable. However, just because an approach may **style** itself as anti-realist, does not necessarily mean that it needs to make that metaphysical commitment; what may appear an anti-realist approach may in fact be a non-realist approach in disguise, or at least open to adaptation towards this. With this in mind, I will focus on two very different kinds of self acclaimed anti-realist approaches in particular: error theory, as championed by J.L. Mackie, and "non-objectivism", as described by Richard Joyce.

First, let us take a closer look at error theory. This is the position championed by moral sceptics such as J.L. Mackie, and consists of three essential components:

1. There are no moral features of the world – right or wrong.
2. Therefore, no moral judgements are true; however,
3. Our sincere moral judgements try, and always fail, to describe the moral features of things. (Shafer-Landau, 2010, 292)

So there is a definite commitment here to an ontological claim, where all sincere moral sentences that assert moral facts about the world are mistaken, as there are no moral facts to which those sentences refer. This also leads to an epistemological commitment, whereby the

claim is that to have moral knowledge there must be some kind of moral truth. This comes from the cognitivist claim that moral sentences are truth apt assertions and combines it with the nihilistic claim that there is, ontologically, no such thing as a moral fact. As I will discuss later on in the thesis, conflating moral truth with moral fact can be a basis for objection to error theory, most notably from quasi-realists like Simon Blackburn. However, for now let us assume that there is a connection between an assertoric truth apt sentence and what is and is not considered **factually** accurate, constituted by what is real and that if that fact is absent, then the assertion proclaiming it to be true is by definition both false and mistaken. So for instance, 'Hilary won the 2016 U.S. general election' is a truth-apt sentence, but the **fact** of this occurring is absent (it did not happen), and this seems to have a clear effect on whether or not the sentence is **true**. However, there are alternative ways of looking at error theory.

The most common form of error theory is what Joyce describes as the Global Falsity approach, which consists of the cognitivist stance that moral assertions are false in that they claim that moral facts exist when in reality they do not (as found in Joyce, 2007). J.L. Mackie in fact goes further and claims that moral assertions can only be true at all if there are moral properties that are intrinsically motivating (Joyce, 2007). This is something I mentioned earlier in my chapter on moral realism when I mentioned Mackie's 'Argument from Queerness'. However, there is also an alternative form of error theory that Joyce calls 'Presupposition Failure' (Joyce, 2007). This means that moral sentences are neither true nor false but that the sentence is still thought to be somehow truth apt. The presupposition that moral facts exist leads to a semantic failure to communicate a **normal** proposition, in the same way that someone claiming that, 'the present king of France is bald' might do[14]. This is to say that a sentence may be truth apt in the sense that it has the logical **form** of an assertion, but cannot be properly determined as true or false in the same way as a normal assertion can be, because it **presupposes** facts that do not exist. This presupposition is in fact very similar to what Joyce talks about in his work, with the obvious difference that rather than such presupposition sentences leading to semantic failure, they are he argues a form of "fictionalism", which of course I will be discussing at great length later on in section two.[15] However, presupposition failure, whether applied to moral discourse or not, is not without

[14] This example is something of a can of worms, however, as it is not well established, philosophically, what someone who utters something like 'the present king of France is bald' is actually doing semantically.

[15] 'Spiderman's real name is Peter Parker' can be interpreted either as a contextual, truth apt assertion about the canon Marvel Universe, or, as Joyce would put it, it can be seen as a non-truth apt (but still **treated** as such) 'pretence' assertion, which is dependent on a shared, cooperative understanding of that universe and its fictitious nature (see section 2 chapter 7 for my account of Joycean 'fictionalism').

considerable debate. Strawson, for example, argues that there are grounds to determine that such sentences as, 'the present king of France is bald', are false (as seen in Beaver, 2007).

The issue I have with any kind of error theory of course is that it constrains us to nihilistic ontological commitments, where my aim in this thesis is to construct a non-realist position which can be championed by the moral agnostic, in other words, one who does not attempt to verify whether or not moral facts do or do not exist. My aim here is to see if a non-realist argument can indeed be constructed **independently** [through the language]. As such, error theory, whilst interesting, is lacking in the specific characteristics required for doing this.

Having said this, there are also alternative non-realist accounts to be found in the debate over whether ethical sentences are cognitivist or non-cognitivist, or whether or not moral claims are truth apt assertions that can be used to express the subject's moral beliefs (cognitivism) or whether they have a different, non-cognitivist function. Both forms of expressivism and moral fictionalism that I will be focussing on are portrayed as non-cognitivist. Whilst moral sentences may appear to be and even be **treated** like or 'as' assertions, they ultimately are not, rather they are some other means of communicating some other kind of important information. This is because most theories which use a cognitivist approach, such as error theory, constrain us to an ontological commitment, in this case that when ethical sentences are uttered, assertions are being made about how the world 'really' is.[16] This is not to say that the cognitivist necessarily does this. For example, as we will discover much later, one brand of fictionalism, the 'tacit operator approach', cleverly avoids this by saying that the semantic content of a fiction statement translates to "within the context of x, x is true"; I will discuss this at considerable length in my second section. Aside from such exceptions, however, most cognitivists on moral language do appear to agree that moral assertions do **attempt** to refer to real phenomena, successfully or not. This is with one possible exception: what Joyce describes as "non-objectivism".

In Joyce's words: "..."moral non-objectivism" denotes the view that moral facts exist and are mind-dependent (in the relevant sense), while "moral objectivism" holds that they exist and are mind-independent." (Joyce, 2007) Whilst this seems to give us a succinct answer to what

[16] If we want to create an **independent** non-realist moral language theory, that is one which does not supervene upon metaphysics, then most forms of cognitivism are automatically unattractive, because they rely on the premise that when moral sentences are uttered, they refer to **real**, metaphysical phenomena.

non-objectivism is, it does also leave it open to debate what exactly is meant by a 'mind-dependent' fact.

As a non-realist account, non-objectivism is also dubious, because it is not strictly speaking non-realist per se but rather a form of minimal realism (proposing that moral facts do exist albeit mind dependently). In other words moral features are not "objective" in the sense of them being (mind) independent, rather they are, at least partly, constituted by features of the mind. This is, however, not to say that they are not in some senses **real**. The basic idea here is to emphasise the dependence of moral 'goodness' on mental activity, for instance:

X is good iff John approves of *X*
X is good iff John would approve of *X* (in such-and-such circumstances)
X is good iff *X* merits John's approval

According to Joyce, it is difficult in this sense to find a successful description of *X* that is in no way dependent on mental activity, and this catalogue can conceivably be much longer depending on whether we treat 'iff' as necessary or contingent, a priori or a posteriori (Joyce, 2007).

The question I would ask is what exactly is the semantic content of these kinds of moral sentences and how are they interpreted propositionally. Can we make inferences from them? The cognitivist, for instance, subscribes to the theory that moral sentences like 'kicking cats is morally wrong', can be **true**. If non-objectivism says this also, then this becomes very confusing. As we will discover later on, quasi-realists like Blackburn describe moral discourse in a similar way, including an account which describes how moral sentences may be interpreted as true in virtue of mental state(s) of the speaker rather than on any objective features of the world. The difference, however, is that Blackburn's account is distinctly non-cognitivist; it involves attitudes being 'projected' onto the world via moral utterances (projectivism), and as we will discover towards the end of section one and the beginning of section two, it leads him into multiple difficulties. It is not entirely clear what non-objectivism has to account for this question.
What is important for us to decipher here is whether or not non-objectivism is viable as a non-realist metaethical theory, and even if it is whether or not, on inspection, it merely collapses into a slightly different, more cognitivist form of quasi-realism. To do this, we must

first defuse any contention of this by drawing a distinction between the objective and the non-objective. Unfortunately, if dealing with anything in philosophy leads us to opening up a can of worms then it is this. There are innumerable different ways to draw a distinction between the objective and the non-objective, and none of them are without objection(s). What we can do, however, is to look at the kind of argument that some philosophers have put forward towards this.

For example, one fairly novel argument from Crispin Wright involves discriminating between phenomena that play what he calls a "wide cosmological role" and those that play a more narrow one; this is also similar to Harman's argument which we have previously cited (Wright, 1988, 197-198). This argument is largely based on an analysis of the contextual nature of explanatory power that the subject matter possesses. For example, there is a great deal that the rectangular shape of a door explains, including our tendency to see it and think of it as rectangular, the shadow it casts and the absence of draft in the office when shut. In this case it appears that the object of the discussion, the geometric shape of the door, determines how we perceive it. By contrast, a value, such as funniness, seems to play a much narrower cosmological role in that it explains very little about any object so much as it explains how we react to it. The main difference seems to be that whilst we may refer to objects which have this property of funniness; it is difficult to imagine that the funniness of something can explain the occurrence of any other phenomena without our tendency to see it as funny playing an intermediary role (Wright, 1988, 197-198). In other words, whilst the geometrical shape of an open door way may help to explain the draft in the office, and thereby other physical phenomena besides, the funniness of a joke does not seem to explain anything if we do not **perceive** the joke as funny.[17] Wright believes that morals have a similarly narrow cosmological role; morality does not explain any phenomenon without the intermediary explanation that people react to things in a particular way, and so it is not, in the sense that Wright portrays it, mind independent, and so it is portrayed as non-objective. (Wright, 1988, 197-198)

[17] To begrudgingly use a cliché thought experiment here: if the door is open, but there is nobody in the office, is there a draft in the office? Alternatively, if a joke is uttered by a machine, but nobody, including the machine (lacking self awareness), is listening, is the joke funny? Whilst this is not exactly what Wright is driving at, it may help to conceptualise that there is a difference between the two occurrences if we look at it in terms of explanations of things happening. We might explain physical phenomenon, such as somebody's tea getting cold, via the fact that there was a draft in the office, because the door was open. This could be true whether there was somebody in the room or not. However, if we were to explain that somebody spilled their tea because they were laughing at an amusing joke, then this is far more **person oriented** in the sense that this could not have happened were a person's reacting to something not a contributing factor.

However, we would do well to remember that what we are looking for is an account of moral **language** and not of morality itself per se. In terms of moral language, how does non-objectivism translate? If, as Wright claims, it is a form of **anti**-realism, then we ought to take a moral statement like, 'kicking cats is morally wrong', as **not** referring to or explaining any real moral properties. That the act **is** morally wrong is far more open to interpretation if we believe that moral qualities, including condemnation, are not intrinsically linked to these facts. However, if this is true, then why does the sentence **appear** to resemble any other assertion, e.g.: there does not appear, at face value, to be any distinct difference in **semantic form** between 'kicking cats **is** morally wrong' and 'the door **is** rectangular'. There are differences in how we perceive these properties and how they relate to how we explain other phenomena, but as regards the semantics of these sentences themselves, merely **saying** that the content of one is more 'objective' than the other does not strike me as quite enough to explain why they are formed and treated so similarly and yet are viewed so differently by anti-realists, unless there is more to this.

Ultimately, this kind of approach to moral discourse does in fact seem to, as far as I can tell, eventually collapse into a kind of quasi-realism (which I will outline in my next chapter) when we attempt to explain how it is that these statements can be **treated** as truth apt in the absence of truth conditional content. This is because the more we press "non-objectivism" on how these seemingly truth apt sentences are being treated, the more it seems that there is a resemblance in the account of speakers 'projecting' their attitudes by making the utterance than the utterance genuinely reflecting the world it appears to describe. For example, if we look again at statements like "X is good iff John approves of X", and this emphasis on moral quality being conditional on approval or disapproval, then striking parallels between this, and what I will be discussing at length in my next few chapters about expressivism and quasi-realism begin to surface. Eventually, both appear to boil down to the same premise, which is that a moral sentence may allude, or even refer to real phenomena, but the moral component(s) of these utterances are ultimately constituted not by these phenomena but by mental activity in reaction to them.

To summarize, error theory is useful in that it gives us grounds to doubt moral realism and thus provide a platform for non-realist moral discourse. However, it is lacking in the specific criteria that I am looking for, that being a non-realist thesis that can be adopted by the moral

agnostic, and it lacks this because it is by its very nature committed to the theory that there are no moral facts. This dependence on additional ontology also applies to most forms of cognitivism, this is with the possible exception of presupposition failure (a subdivision of error theory), but which may easily collapse into a form of moral fictionalism when pressing for a more thorough explanation. Similar issues are present when we look to non-objectivism, which similarly seems to collapse into quasi-realism.

Chapter 5: Simon Blackburn & Quasi-Realism

The origins of quasi-realism can be found the 1984 book *Spreading the Word: Groundings in the Philosophy of Language* by Simon Blackburn. Quasi-realism is a position in meta-ethics which holds that ethical statements do not express propositions; rather they project emotional attitudes onto the world as we speak, '**as if'** they were objective properties. Blackburn describes quasi-realism as a **project** which aims to explain why we appear to use moral language in an objective manner (Blackburn, 1984, 170-171). In *STW*, Blackburn introduces this project during his discussions on key issues within the philosophy of language, namely his discussions on language and realism in his fifth chapter, which he expands upon in his sixth and seventh chapter. In his fifth chapter, 'Realism and Variations', Blackburn discusses moral realism as he provides evidence for and against various forms of both this and anti-realism, as well as cognitivism and non-cognitivism. Recalling back to chapter three of this section, some questions we would do well to remember here are the following: do moral sentences express propositions? Can such propositions be true? If so, are such true propositions made true by objective features of the word? As we also discussed in chapter two of this section, we are treating 'moral realism' as a claim about moral language (that moral utterances are assertions about objective moral features of the world). Blackburn appears to treat these terms in much the same way as well.

The most important factor in the quasi-realist project is Blackburn's 'projectivism', the process by which we project attitudes onto states of affairs or objects. Blackburn argues that moral claims can be best understood as projections of our emotional attitudes (Blackburn, 1984, 170-171). Blackburn begins to defend projectivism in his sixth chapter, 'Evaluations, Projections and Quasi-Realism', as he explains how and why qualities, such as right or wrong, may be projected onto certain actions '**as if'** they were objective properties. He then outlines a particular crisis that arrives from this. In order to maintain this view, the projectivist seems committed to declaring that all moral discourse is subject to a pervasive error. Why do we say "ω is wrong", rather than, "I disapprove of ω" if that is what we mean, unless our moral language is somehow broken? The quasi-realist project's aim is to demonstrate that a projectivist can avoid committing to either moral realism, or an **error-theoretic** account of moral language. The phenomenon that we must then investigate is why, when we use language within the context of ethics, we use seemingly truth-conditional claims in the absence of truth conditional **content**. This is the central issue Blackburn deals with in

the end of his sixth chapter, and in his seventh chapter, 'Correspondence, Coherence and Pragmatism'. In my following chapters in this section, I will also discuss how Blackburn confronts the Frege-Geach problem and then goes on to use that line of reasoning to account for notions of moral sensibility, moral improvement, and the beginnings of a notion of moral truth. Blackburn's aim is to provide evidence for the quasi-realist project's potential for being a complete and comprehensive account of the **semantics** of moral language. Unlike conventional expressivism, however, quasi-realism is also aimed at providing the moral realist with everything that they want to hear without supervening in any way on any metaphysics or epistemology regarding the objective existence or accessibility of moral truth(s) or moral properties. As such, quasi-realism looks like a tempting candidate as the kind of non-realist account of moral language that I am ultimately searching for.

Towards the end of his fifth chapter, 'Realism and Variations', Blackburn introduces us to the kind of non-cognitive expressivism he favours, i.e. regarding some kinds of utterance as expressive (of emotion) rather than descriptive (of fact). Blackburn claims that moral views are best understood not as beliefs, but as attitudes. As A.J. Ayer puts it:

> The presence of an ethical symbol in a proposition adds nothing to its factual content. Thus if I say to someone, "You acted wrongly in stealing that money", I am not stating anything more than if I had simply said, "You stole that money". In adding that this action is wrong I am not making any further statement about it. I am simply evincing my moral disapproval of it. (Ayer, 1971, 107)

This is sometimes dubbed the 'boo-hooray' theory of ethics. Where realism must account for the metaphysical nature of alleged 'moral properties' present in the world, expressivism does not, rather it simply makes the case that certain facts or actions evoke certain emotions within individuals. Something that Blackburn hastens to add, however, is that a moral claim may not form a factual account of an event, but it does not, therefore, follow that it is devoid of truth conditions.

> There is no inference of the form 'this attitude is expressed, *so* these remarks have no truth conditions', but only 'this attitude is expressed; if we see the remark as having no truth conditions the philosophy improves; so let us see the remark as expressive rather than descriptive. (Blackburn, 1984, 170)

Furthermore, it does not matter if an utterance is descriptive as well as expressive if its distinctive **meaning** is expressive. Next, Blackburn introduces us to his take on projectivism,

> Suppose that we say we project an attitude or habit or other commitment which is not descriptive onto the world, when we speak and think as though there were a property of things which our sayings describe, which we can reason about, know about, be wrong about, and so on. (Blackburn, 1984, 170-171)

In the context of moral language, let us say that I condemn Erwin Schrödinger for putting a real life cat in a real life box to test his theories about quantum super positioning. Were he to put his cat in a box, then by Blackburn's account this is answered by my projecting the quality of 'wrongness' onto that action, rather than that action holding that quality as an objective property irrespective of my view. To the contrary, one might say that putting cats in boxes is wrong specifically **because** I think that it is wrong. However, as J.L. Mackie points out, this seems to imply that our ordinary use of moral predicates involves a pervasive error, because whilst we claim one thing, another is the case: I claim that putting cats in boxes is wrong, even though it is not (Mackie, 1977, 30-35). This alone reveals how expressivism can lead to an error-theoretic account of moral language, or that our language is somehow broken. It falls then to quasi-realism to account for why moral language is used in the way that it is, which Blackburn begins to explore in his sixth chapter, 'Evaluations, Projections, and Quasi-Realism'. According to Blackburn, there is no error being committed on my part when I accuse Schrödinger of doing something immoral with his cat. Blackburn's projectivist account has three significant promises. 1) It has metaphysical and epistemological simplicity; projectivism offers an explanation for moral discourse that is not reliant on any complex metaphysical account(s), or any special epistemological faculties for knowing of them, and allows for a naturalistic account of reality that does not contradict its account of morality. 2) It is an explanation of the supervenience of moral qualities. (I will discuss this in a moment) 3) It accounts for the intimate connection between moral belief, and motivation towards that belief (Blackburn, 1984, 183). This all sounds promising, provided that quasi-realism can actually deliver on what it promises.

In the interest of clarity, I should briefly discuss the notion of supervenience. "Supervenience" is the name of a certain kind of relation between two sets of facts, which we

can call A-facts and B-facts. A-facts supervene upon B-facts if any change in the A-facts requires that there must have been a change in the B-facts. For example, it is sometimes asserted that facts about mental phenomena supervene upon facts about physical phenomena: if there is a difference between two sets of mental facts, then there must also be a difference between two corresponding sets of physical facts. No change at the mental level is possible unless there is a change at the physical level. Just how Blackburn thinks projectivism is at an advantage in explaining the supervenience of the moral facts on the non-moral facts is complicated, and I will return to this element of his argument later. In a language context, however, projectivism suggests that moral properties are directly related to our linguistic interactions with our perceptions of actions. For example, the action of putting a cat in a box I may describe as having the quality of 'wrongness'. This is directly related to my linguistic utterance of 'wrong' being a projection of my perception of that action. In saying that putting a cat in an air tight box is 'wrong', I indicate that I understand the meaning of 'wrong' by whatever criteria I use to judge this action, and that I am projecting that property onto that action.

Moral sentiments are by no means the only projections that people make; sentiments relating to aesthetic taste may work in a very similar way by this logic. For example, if I was to compare Beethoven's Pastoral Symphony with, for example, the song 'One Time' by Justin Bieber, then I would judge Beethoven's music to be of significantly higher quality in a wide variety of ways. For example, one way of judging the two songs would be to compare the level of emotional impact that the two convey to the listener. As the two are perceived by me, the Pastoral Symphony seems to take the listener on an emotional journey, venturing through joy, excitement and wonder in some places to profound tranquillity and serenity in others. In contrast, Bieber seems to completely ignore the emotions of his audience in favour of repeatedly telling them what he claims he is experiencing, which is utterly impossible for me to take seriously, let alone identify with. Simply put, Beethoven demonstrates a far more profound empathy with his audience, not to mention a much more sophisticated understanding of human emotion and artistic expression. However, it is only because my perception of what qualifies as 'good music' supervenes on these kinds of sentiments that I feel at all qualified to make linguistic utterances such as, "Beethoven's music is better than Bieber's". This, of course, makes it questionable whether my perception of the two songs would be different if I had had a totally different set of experiences.

As stated earlier, this refers us back to the dilemma that expressivism faces when talking about subjective qualities in an objective manner. Why do I say that putting a cat in a box is wrong, '**as if**' it was wrong no matter what I or anyone else thought, when in fact 'wrong' refers to my **perception** of the act, rather than an objective property of the act? What happens if I change my mind? What happens when I cease caring about the plight of cats and begin to see no problem with putting them in boxes? Is it still wrong then? One way we could answer this is by considering the following, "I believe that energy is equal to mass times the speed of light squared. I believe this to be a universal constant. I believe that one day; a scientist will validly prove this principle to be false." Surely, at least one of these statements must be false. If one sincerely believes that something is **true**, one cannot believe it is possible to prove it false, any more than one could prove something true when it is in **fact** false. When it comes to things like 'rules', if one believes a rule to be overly susceptible to change, then one surely does not believe in the 'ultimate correctness' of that rule. I may change my mind with the acquisition of additional evidence, but that does not change the fact that originally, I did not foresee that I would. For instance, I accept the **possibility** that at some point in the future, someone may find evidence to doubt that $E = mc^2$. However, as of **this moment**, I do not believe that that will happen. Surely this is what it means to believe that something is ultimately 'correct' in virtue of **fact**.

When dealing in quasi-realist moral discourse, however, we are working not with moral facts but with something else entirely. For example, for the majority of history, it was not generally held that women should have the right to vote[18]. Similarly, it was not held that a desk consisted of mostly empty space. However, with an increased understanding of the structure of atoms, it is a view that is now commonly held, just as increased considerations of women's rights, at least partially, led to the suffragette's movement. That a majority of history consists of not believing that women should have a vote, does not, and should not change the fact that I currently believe that they should, any more than it should change my belief that energy equals mass times the speed of light squared, when for most of history, nobody believed that. The difference, however, as the quasi-realist portrays it, is that whilst I **assert** that $E = mc^2$, and this is a universal constant, a fact, one that would be the case even if I did not believe it, there is something very different about when I seemingly assert that, 'women should be

[18] In fact, for a majority of human history, neither men nor women were allowed to vote, as most historical civilisations have not been democracies at all, and most of those that have been have limited democratic rights only to minority social elites.

allowed to vote', '**as if**' it were a reference to a fact of that same, scientific variety. In the latter case, that quasi-realist claims that my belief/attitude towards it is fundamental to its truth value. To a quasi-realist this is a very different kind of assertion, if indeed an assertion at all. Where fact meets truth is a contentious issue at the best of times in philosophy, but where expressivist moral utterance meets moral truth for the quasi-realist is even more so, and sure enough it meets opposition. The most notable objection to the quasi-realist project's approach to these matters is the Frege-Geach problem, which I will discuss in my next chapter.

Chapter 6: The Frege-Geach Problem

The Frege-Geach Problem is an issue related to expressivism in moral discourse and an issue which Simon Blackburn attempts to deal with in *STW* (1984), by developing a more sophisticated version of expressivism that he calls **quasi-realism**, which we began to look at in the previous chapter. In this chapter I will describe in detail what exactly the Frege-Geach Problem is, why it is seen as so problematic for expressivists and explore Blackburn's responses to this problem.

In its most basic form, the Frege-Geach Problem demonstrates a hole in the expressivist account of moral language, where expressivists are unable to account for how expressive utterances are able to function in unasserted contexts. It was famously proposed by Peter Geach who developed it from the works of Gottlob Frege in 'Assertion', *Philosophical Review* (1964) (Blackburn, 1984, 189). The Frege-Geach Problem is as follows: if expressivism is true, then moral claims, e.g. "stealing is wrong", function as expressions of attitudes, similar to saying, "boo! stealing!". In speaking like this, the speaker expresses their disapproval of stealing. However, there are some situations in which these same utterances appear to work differently, and it is this peculiarity that the Frege-Geach Problem focuses on. For example, imagine I was to say, "**if** stealing is wrong, **then** arresting a thief is right", in other words I use the same expressive utterance as the antecedent of a conditional claim. However, in doing so, I reveal a significant problem. If we break the claim down and focus on the antecedent, "**if** stealing is wrong..." then we should notice that I am not actually asserting that stealing is wrong or right nor am I expressing any attitude whatsoever, condemnatory or otherwise. I am merely saying that **if** stealing is wrong, then arresting a thief is right. This is what is called an unasserted context. The expressivist has an account of the function of such expressions when they are asserted, (e.g. 'Stealing is wrong' = 'boo! stealing!'), but does not have an account of how they can function in unasserted contexts like these. So there is a mismatch here. According to expressivists, moral judgements, e.g. "stealing is wrong", express emotions, yet in conditionals, they do not, or at the very least if they do then we do not know how they work like we do with asserted contexts, and this mismatch is but one example of the Frege-Geach Problem. More broadly, the problem is defined by a general discontinuity associated with how moral judgements function on their

own compared to how they function in unasserted contexts. Conditionals are a useful example of this. For example, if we were to propose a simple modus ponens argument:

1) **if** {stealing is wrong}, **then** arresting a thief is right
2) {stealing is wrong}
C) **therefore**, arresting a thief is right

...then we have a serious problem when we also claim that the moral judgement, "stealing is wrong", is really an expression of an **attitude**. To help explain this dilemma, notice that I have placed the significant utterance in curly brackets {}. The problem is that the phrase within {} appears to hold two different meanings in (1) and (2), leading to a fallacy of equivocation. This is akin to the kind of problem encountered in arguments like: laws imply lawgivers, there are laws in nature, and therefore nature must have a lawgiver. The argument is fallacious, because a key term or phrase within that argument (in this case the use or meaning of 'law') is being used in an ambiguous way, in other words it holds one meaning in one part of the argument and a different meaning later on. Another example might be to say something like, nothing would be better than having world peace, a ham sandwich is better than nothing, and therefore a ham sandwich would be better than having world peace. To explain why this is so problematic in the Frege-Geach problem, when using inferences like modus ponens, it is important that all of the premises match up, and that meanings of words are not changed mid-way through the argument, otherwise we get exactly this kind of fallacy of equivocation. In (2), 'stealing is wrong' holds a different meaning to how it is used in (1) by simple virtue of the fact that in (1) it is a conditional proposition, conditional on whether or not **it is the case that** 'stealing is wrong', which (2) does not provide. (2) does not assert that 'stealing is wrong' **is the case**, it merely translates to 'boo! stealing!', which in (1) it does not, or at the very least not clearly so. This creates a **discontinuity** between the two utterances as they are used in the standard modus ponens argument.

Another problem lies in how the argument can be interpreted as valid or invalid. In other words, there seems to be a problem with the truth aptitude of what lies within {}. In order for modus ponens to work as a valid argument, it is imperative that what lies within the brackets in premise (1) is the same as what lies in the brackets in premise (2). In a valid modus ponens argument, were we not to use an expressivist analysis but rather treat all premises as moral realists would have us do (i.e. 'stealing is wrong' is truth apt), it cannot on pain of logical

contradiction be false that arresting a thief is right if the first two premises are true. This is a good truth functional analysis of logical connectives (**and, if-then** etc.) However, if we do use an expressivist analysis and assume that 'stealing is wrong' = 'boo! stealing!' or something similar, then these logical connectives become immediately unusable, because on the expressivist account, (2) is not truth apt, and the status of (1) is unclear at best.

Therefore, the expressivist must give an account of what is going on, or else their attempt to explain moral discourse is either incomplete or simply flawed. When I say, "if {stealing is wrong}...", then no matter what content lies within the brackets, I am neither asserting nor expressing anything whatsoever. However, if the terminology is that of a realist, then there is clear, truth conditional content to work with, but if the terminology is that of an expressivist, then it is simply not clear what someone making such an utterance in a conditional is doing. One possible way an expressivist might try to analyse the utterance that I can suggest could be to look at premise (1) in its entirety and then claim that the holistic sentence comes across as an assertion regardless of the content of the brackets. If I were to literally say, "if {Boo! Stealing!}, then {Hurray! Arresting a thief!}", this could still be taken as an assertion, for I am still communicating information about a state of mind that is conditional on another state of mind. However, this still makes premise (1) too different from premise (2) for modus ponens to properly function, in other words we are still left with the equivocation fallacy. In premise (2), "stealing is wrong", is an expression of an emotion, **not** an assertion. If we care to translate the entire argument into an expressivist analysis, we might get something like this:

1) **if** {Boo! Stealing!}, **then** [Hurray! Arresting a thief!]
2) {Boo! Stealing!}

C) **therefore**, [Hurray! Arresting a thief!]

Whilst at first glance this seems to be logically consistent, when we examine the semantics of this kind of propositional logic, it soon becomes apparent that it is not. When utilizing a conditional claim, the only official analysis we have of how this works in something like a modus ponens argument is a truth functional one, in which **if** and **then** should be understood as, "**if** *it is the case that* **x**, **then** *it is the case that* **y**". The same should be understood from **therefore**. This demands semantic content that is truth apt. So now let us compare the two

rival content analyses, truth apt and expressivist, in modus ponens, where each contains an unasserted context in the antecedent of premise (1) in greater detail:

- Truth Apt Content

1) **if** *it is the case that* {stealing is wrong}, **then** *it is the case that* [arresting a thief is right]
2) *it is the case that* {stealing is wrong}

C) **therefore**, *it is the case that* [arresting a thief is right]

Valid Argument

- Expressivist Content

1) **If** *it is the case that* {Boo! Stealing!}, **then** *it is the case that* [Hurray! Arresting a thief!]
2) {Boo! Stealing!}

C) therefore, *it is the case that* [Hurray! Arresting a thief!]

Invalid Argument*---Whilst we may infer observationally that *it is the case that* {Boo! Stealing!}, as the speaker has expressed that emotion, we cannot infer this as part of a deductive, propositional argument unless (2) is overtly **asserted**, otherwise premise (2) is incompatible with the antecedent of premise (1). It is also difficult to tell whether '*it is the case that* {Boo! Stealing!}', even makes sense as an assertion.

It seems clear that expressivists and quasi-realists cannot escape the Frege-Geach problem through propositional logic, and it is for this reason that Simon Blackburn attempts to escape by instead attempting to construct a 'logic of attitudes'. Blackburn conducts a thought experiment in *STW* (1984), which explains how he attempts to do this. It is not a great stretch to imagine having attitudes **to** one's own attitudes. So, imagine that a language, which we call E_{ex}, contains no evaluative predicates and contains the same 'hooray' (H!) and 'boo' (B!) operators as ordinary expressivism. It may then contain the following semantic structure in regard to moral attitudes: H! (|H! (Tottenham) | ; | H! (Arsenal) |), H! (|B! (lying) |;| B! (getting brother to lie)|). Here we see how in E_{ex} one may endorse attitudes which, if they endorse Tottenham, also endorse Arsenal, such that to express that Tottenham is a good team is to do so for Arsenal. This is similar to an ordinary conditional. However, the difference is that Blackburn treats the conditional itself as expressing an emotional attitude of approving the conjunction of approval of Tottenham and approval of Arsenal. The equal and opposite is

done in condemning lying and also condemning getting one's brother to lie, the example that Blackburn later uses (Blackburn, 1984, 193-195). Blackburn wants to compare E_{ex} to the language of expressivists. Imagine then that when I argue that 'it is because stealing is wrong that a thief should be arrested', I am utilizing the language of E_{ex}: H! (|B! (Stealing) | ; | H! (Arresting a Thief) |). As an expressivist, I am expressing my approval of a certain combination of attitudes, those being a disapproval of stealing and an approval of arresting a thief. (Blackburn, 1984, 193)

In order to further explain the somewhat convoluted semantics of E_{ex}, Blackburn provides a simple analogy. Say I project a negative attitude 'boo', (B!), onto the act of lying (L), so B!(L), and I also project this negative attitude onto the act of getting my brother to lie (BL), so B!(BL). These are **first** order attitudes; they are emotional reactions in response to an occurrence whether it be lying or getting one's brother to lie. Blackburn then asks us to examine **second** order attitudes, emotional reactions to first order attitudes (Blackburn, 1984, 195). For instance, I may approve of my own moral condemnation of stealing, lying or inflicting violence, or I may disapprove of my condoning those things. Whilst the latter may seem strange, it is important to recognise, for example, that I may follow a line of logical reasoning that leads me to conclude that violence against children is acceptable under certain circumstances, against my 'gut instinct'. I may, however, avoid this, owing to the realisation that I cannot possibly approve of any line of reasoning that leads me to conclude that violence against children is acceptable. Blackburn later calls this kind of attitude 'fundamental' (I will discuss this further in my next chapter), and it is a significant component of Blackburn's response to the Frege-Geach problem. For example, say I project a positive attitude 'hurray', (H!), onto my first order attitude towards lying, H!(B!(L)). Blackburn then asks us to examine the second order attitude towards a conjunction of first order attitudes, H!(B!(L) ^ B!(BL)). Here I approve of the conjunction of first order attitudes B!(L) my negative attitude towards lying and B!(BL) my negative attitude towards getting my brother to lie. Crucially, this kind of reasoning appears only to work under a truth functional analysis, particularly where the conjunction '^' is concerned. In propositional logic, a conjunction is true if and only if both of its conjuncts are <u>true</u>: 'A^B' is true if and only if both 'A' is true and 'B' is true. Expressivist semantics must deal with this, because otherwise 'A^B' cannot be functional in the same way, because 'A' & 'B' are not truth apt; they are expressions of emotions. However, according to Blackburn, if I were to commit myself to B!(L) and to H!(B!(L) ^ B!(BL)), then I ought to also accept B!(BL):

1) B!(L)
2) H!(B!(L) ^ B!(BL))
3) B!(BL)

According to Blackburn, someone who makes commitments to (1) and (2), but fails to accept (3) is exhibiting a kind of inconsistency:

> Disapproval of lying, and approval of making (disapproval of getting little brother to lie), follow upon (disapproval of lying). Anyone holding this pair must hold the consequential disapproval: he is committed to disapproving of getting little brother to lie, for if he does not his attitudes clash. He has a fractured sensibility which cannot itself be an object of approval. The 'cannot' here follows not (as a realist explanation would have it) because such sensibility must be out of line with the moral facts it is trying to describe, but because such a sensibility cannot fulfil the practical purposes for which we evaluate things. (Blackburn, 1984, 195)

How does this deal with the Frege-Geach problem? The original problem is that an expressivist analysis of moral language lacks a story about how moral claims function in unasserted contexts e.g. '**if** stealing is wrong, **then** arresting a thief is right'. Blackburn's response is to avoid having to 'correct' the language of propositional logic and instead provide an alternative language, specifically for moral discourse, which mimics the function of modus ponens. In the quasi-realist language of modus ponens, Blackburn treats the conditional as itself expressing an attitude: approving of the conjunction of two disapproving attitudes. In doing so, Blackburn seems to provide a plausible account of how expressive utterances function by providing a non-truth functional, expressivist way of understanding this particular unasserted context. So if we return to our stealing dilemma, to reiterate, a realist may read the following argument as follows:

- Realist Content

1) **if** *it is the case that* {stealing is wrong}, **then** *it is the case that* [arresting a thief is right]
2) *it is the case that* {stealing is wrong}

C) **therefore**, *it is the case that* [arresting a thief is right]

Valid Argument

If we listen to Blackburn, however, then a quasi-realist would read that same argument more like this:

- Quasi-Realist Content

1) B!(Stealing)
2) H!(B!(Stealing) ^ H!(Arresting a thief))

C) H!(Arresting a thief)

Valid Argument*---According to the context of Blackburn's 'logic of attitudes'

In the quasi-realist analysis, (2) effectively fulfils the equivalent role of a conditional. If we accept this analysis as functional, then the Frege-Geach problem ceases to apply to it. The semantic content of B!(Stealing) is the same in premise (1) as it is in (2), and there is no discontinuity occurring with unasserted contexts. All three utterances are expressed.

Chapter 7: Blackburn's Logic of Attitudes

As discussed in chapter six, one of the most famous arguments against expressivism is the Frege-Geach Problem, famously proposed by Peter Geach who again developed it from the works of Gottlob Frege (Blackburn, 1984, 189). To recap, it raises the following issue: does the meaning of the utterance "It is wrong to tell lies" differ when it is embedded in a conditional "If [*it is wrong to tell lies*], then it is wrong to get your brother to lie", because if it does, then expressivism is inadequate as an account of moral discourse. Through modus ponens one should be able to deduce in the following manner: 1) It is wrong to tell lies 2) If [*it is wrong to tell lies*], then it is wrong to get your brother to lie 3) Therefore, it is wrong to get your brother to tell lies. The utterance of the italicized words in premise two are the same as those of premise one, yet in premise two, they are not asserted. It is the antecedent of a conditional and thus the introduction of a supposition. However, they nevertheless must mean the same thing if the rule of modus ponens is to work and to avoid a fallacy of equivocation. For example, one might argue that "He is working at the bank; if he is working at the bank, then he must have his feet in the river; **so** he must have his feet in the river". If 'bank' in the first premise has a different meaning (e.g. a money bank) to the 'bank' used in the second premise, then the rule of modus ponens ceases to apply. According to Geach, the expressivist fails in the second statement in that they are asserting only the hypothetical premise and is expressing no moral position towards lying, condemnatory or otherwise. The expressivist cannot account for the meaning of moral language in an unasserted context. In the following chapter, I will go over how Blackburn develops quasi-realist semantics from dealing with these kinds of issues.

Before we continue, let us recap again on how Blackburn counters this by conducting the E_{ex} language thought experiment seen in chapter six. In the interest of clarity, I will abandon discussing Tottenham and Aresenal and focus the logic of attitudes directly on the lying brother thought experiment, using the following terminology: (H!) = Hurray!, (B!) = Boo!, (L) = Lying & (BL) = Getting little brother to lie. First, we project a positive attitude 'Hurray!', (H!), onto a first order attitude towards lying: H!(B!(L)). Then we examine the second order attitude towards a conjunction of first order attitudes: H!(B!(L) ^ B!(BL)). So I approve of the conjunction of first order attitudes, that is B!(L) (my negative attitude towards lying), and B!(BL) (my negative attitude towards getting my brother to lie). However,

according to Blackburn, if I were to commit myself to B!(L) and to H!(B!(L) ^ B!(BL)), then I ought to also accept B!(BL):

1) B!(L)
2) H!(B!(L) ^ B!(BL))
3) B!(BL)

According to Blackburn, someone who makes commitments to (1) and (2) but fails to accept (3) is exhibiting a kind of inconsistency (see chapter six).

Now let us take this line of thought further. Consider the following thought experiment[19]: Blackburn asks us to imagine we have something called a **moral sensibility** (M), (M) being the **set** of all our moral attitudes and sentiments. Now suppose I had a markedly different upbringing to the one that I had and that (M) is thus markedly different. For instance, (M) includes the sentiment (S), that is, for example, an approval of kicking cats. Now let us apply a kind of Blackburn[*esque*] interpretation of this: **H! (S) ∈ (M)**. In other words, approving of holding the sentiment (S) **is a member of** (M), appealing to set notation; I approve of having the moral sentiment (S) in which I approve of kicking cats, and this approval is a member (or element) of (M) my moral sensibility. Intuitively, we want to say that this is in some way **problematic**. One means of doing this, for example, might be to argue that this makes moral language too relativistic. As quasi-realists we would do well to be very careful of saying that all moral utterances are reduced to is blind attitudes alone or imply that all moral opinions are equally valid no matter what the **content** might be, otherwise any notion we might have about moral fallibility, moral improvement or moral truth conditions become very difficult to account for without simply appealing to error theory. Whilst this may not be a problem for some expressivists, for quasi-realists like Blackburn this is a serious problem if they want to gain entitlement to give the moral realist in the room everything they want to hear and more about the qualitative content of moral arguments.

Another issue is a more a general worry we may have about endorsing specific moral sentiments by virtue of endorsing more general moral principles. In other words, views about general moral principles may not necessarily lead to desirable conclusions or views about

[19] The following thought experiment is, I believe, a unique interpretation of Blackburn's position on moral improvement, as I see it, but which introduces notation of my own in an attempt to better understand it.

specific matters. I may approve of an overall moral principle and yet still render myself with a serious problem when I act in accordance with that maxim in every conceivable situation e.g. if I kissed my girlfriend in public in Dubai, because I believe in the general principle of liberty in regards to public displays of affection and utterly fail to consider the specific cultural context of the setting in which I do so. I may agree with a moral principle and yet disagree with certain conclusions of it when applied to certain situations. This seems to raise the question of whether I disapprove of kicking cats because I have the general moral principle about kicking living creatures that I have, or, whether I disapprove of the act in virtue of my perception of the fact itself. Do I hold specific views about specific situations? Or general views based on general principles?

Returning to the matter at hand, however, what we want is to provide **logical** reasons for rejecting this kind of coupling of first and second order attitudes and accepting their negation, (¬ H! (S)); that is to say, 'it is not the case that I approve of sentiment (S) in which I approve of kicking cats'. This brings us to Blackburn's interesting account of **moral improvement**, which I will attempt to explain in the following way[20]. In constructing an account in which we project second order attitudes onto first order attitudes, and so on and so forth, we can, in theory, account for the changing nature of moral sentiments. So once again, say I have a moral sentiment (S), but I admire an alternative sentiment, let us call this (S+) more than (S). As far as I can tell then, transitioning from (S) to (S+) would be a moral improvement. For instance:

(S) contains:
|H!(ϖ|,|B!(|H!(ϖ)|)| ^ |H!(|B!(ϖ)|)|

- I approve of ϖ
- I disapprove *of approving* of ϖ
- ...and...I approve *of disapproving* of ϖ

I then transition to (S+), which contains:
|B!(ϖ)|,|B!(|H!(ϖ)|)| ^ |H!(|B!(ϖ)|)|

- I disapprove of ϖ
- I disapprove *of approving* of ϖ

[20] Once again, I reiterate, the following analysis is not a direct quotation of Blackburn's words, but rather my interpretation of what **appears** to be going on.

- ...and...I approve *of disapproving* of ω

As we can see, in (S) we are left in a state of second order **disapproval** [towards first order approval of ω], and unfulfilled approval of the opposite, a disapproval of ω. By contrast, in (S+) that second order **approval** [towards first order disapproval of ω], is fulfilled.

Now, how does this help us with (M)? To recap:
H! (S) ∈ (M).
(S) = H! |H!(kicking cats) |

We now have a discontinuity between this second order disapproval of ω, and the state of approval over (S) expressed in (M). This facilitates a similar transition to what we will for the moment call (M+):

(M+) = H! (||we have sentiments S+||;|H!(B! (kicking cats))|)

In other words, I approve of having the sentiment S+ in which I **approve of** my disapproval of kicking cats. Now let us say that this results in the following attitude: [H! (ω > ¬ ω)], i.e. 'I approve of ω being greater than ¬ ω'. We may then gloss the two following propositions regarding its improvement... (I): (S+) is an improvement over (S), (I+): H! (S+ > S). So in (I) we have a preliminary recognition of (S+) being greater than (S), and then in (I+) we have a fulfilled approval towards this recognition.

So to translate in more laymen's terms what is going on here, as I understand it, we can account for the changes in people's moral sensibilities by accounting for how individuals continuously analyse their own attitudes and have second order attitudes towards them. When we admire moral sentiments that we perceive as somehow superior to our own, we inevitably seek to **assimilate** those sentiments into our own in order to adapt our overall moral sensibilities to additional ways of perceiving things. For example, I may at one point believe that it is wrong to kick people, but it is ok to kick animals. I may then begin to admire an alternative way of looking at things. For example, I may recognise that if the reason for my condemnation of kicking people is because they experience pain, then that should not

discount animals from a similar moral sentiment, as animals are also capable of experiencing pain, and so on and so forth.

Now that we have that understanding, let us return to Blackburn's words more directly. Suppose once again that I hold a moral sensibility (M) where (M) is the set of all my moral attitudes and sentiments. Now take (M*) to be the best possible set of attitudes, that is, "the limiting set which would result from taking all possible opportunities for improvement of attitude" (Blackburn, 1984, 198). Now let us say that, '*m* is a commitment expressing an attitude U' (Blackburn, 1984, 198). Blackburn makes the following claim that:

m is **true** = U is a member of M*

So according to Blackburn, a quasi-realist can even lay claim to a notion of **moral truth**. This of course raises a very big question. How do we know that there is such a thing as a unique best possible set of attitudes and what exactly does that mean? Naturally, Blackburn attempts to explain this by conducting another thought experiment. Of two literary critics, one prefers Ovid, and the other prefers Tacitus. Over time, the one who prefers Ovid improves his literary sensibility into M^*_o, and the one who prefers Tacitus improves his into M^*_t. Taking this into account, can we then accept the following three sentences simultaneously?[21]

(1) $M^*_o \rightarrow$ it is true that Ovid is better than Tacitus.
(2) $M^*_t \rightarrow$ it is true that Tacitus is better than Ovid.
(3) There is no possible improvement on either M^*_o or M^*_t.
(Blackburn, 1984, 200)

Blackburn insists that we cannot. If we begin to accept that a case, such as M^*_o, with this kind of structure applies, then we must also accept that the same is true of M^*_t, because it has the exact same structure (Blackburn, 1984, 200). To help Blackburn out here, let us consider the following. During my English Literature under graduate years in Aberdeen, I read far more, and in greater depth, of Shakespeare's plays than I did of Marlowe's. I thus have a

[21] It is not entirely clear why Blackburn uses (M*), the same notation he uses for **moral** improvement, for what seems to be a reference to **aesthetic** improvement/comparison between Ovid and Tacitus; nevertheless, I am using the notation that is given.

superior scope of Shakespeare's literary style than I do of Marlowe's, and that being the case might lazily conclude that Shakespeare is a better playwright than Marlowe. However, I would not approve of such a conclusion, because I believe that had I as great a knowledge of Marlowe's work as I do of Shakespeare's, then I may well be making the equal and opposite argument for Marlowe. The only way I could say with any real confidence that Shakespeare is a better playwright is if I had an idealised knowledge of both their literary styles, which I do not have. So, in the absence of idealised knowledge of this or any real life matter warranting artistic or moral judgement, how are we to determine a unique best possible set of attitudes?

One possible solution to this issue is what Blackburn's describes as **branching** (Blackburn, 1984, 199). Branching is the process by which a moral sensibility breaks down as it confronts a wide variety of situations and matters and alternative ways of perceiving them; the illustration Blackburn uses is literally of something resembling a tree that branches out in many different directions, yet all originate from the same trunk. This is analogous to how subjective attitudes develop, growing from a singular point yet ending up in many different places. However, what criteria do we have to deem one branch in any way superior to another? Blackburn's proposed solution is to attempt to look more closely at the concepts of obligation and permissibility (Blackburn, 1984, 201). Suppose we symbolise a moral sentiment that 'it is obligatory to A' by OA, and another that 'it is permissible to avoid A' by P ¬A. Logically, OA and P ¬A are in direct conflict. Now imagine the following:

1) $M^*_1 \rightarrow OA$
2) $M^*_2 \rightarrow P\ \neg A$
3) There is no possible improvement on either M^*_1 or M^*_2.

This is consistent with the aforementioned tree structure. However:

4) (3) implies that it is permissible to hold M^*_2; this implies it is permissible to hold that ¬ A; this implies that ¬ A is permissible.
5) If (4), then any view such as M^*_1 which implies the reverse is wrong, and *ipso facto* capable of improvement.
   (Blackburn, 1984, 202)

Here we have a means of transcending the tree structure, and Blackburn goes on to claim that if we have a prima facie case of (1), (2) and (3), then it is likely to be the obligation that is the more threatened:

(4') (3) implies that it is permissible to hold M^*_1; this implies that it is permissible to hold that OA; this implies that OA.
(Blackburn, 1984, 202)

The problem that (4') faces that (4) does not, is that we need an entirely different reduction principle to conclude that (A) is **obligatory** by virtue of being permissible. In other words, Blackburn claims that we can escape the inconsistency by seeing that M^*_1 at least, is capable of improvement (Blackburn, 1984, 202). So, to apply this to our Shakespeare vs. Marlowe case; if I was to hold that it was some kind of moral or aesthetic obligation or duty to prefer Shakespeare to Marlowe, rather than to simply claim that it is permissible to do so, then it would likely be the obligation that would require some form of improvement.

In conclusion, Blackburn not only attempts to circumvent the Frege-Geach problem by constructing a logic of attitudes, but he also develops this logic to an extent that he claims entitlement to an account of moral improvement and even moral truth, and this is achievable, such is the claim, through a combination of some ingenious semantics and set notation. By doing so, the aim is to gain entitlement to give moral realists all that they want to hear about the nature of moral language, but it is important to maintain for a quasi-realist that doing this is sufficiently different from simply being a particularly abstract form of moral realism. To do this, quasi-realism must make no compromises on what it claims about the ultimate nature of moral qualities and to maintain that they are accountable through projectivism, moral psychology and philosophy of language rather than by metaphysics or epistemology. Moreover, the quasi-realist must maintain a position of non-realism. This is an important distinction that must be made in order to separate the quasi-realist from the realist. This is also something which will come back to haunt the quasi-realist later on in section two.

Chapter 8: Andy Egan

Having now explained the basics of Simon Blackburn's quasi-realism, we can now start to look at the objections to quasi-realism that start to occur as the project continues. In Blackburn's 1998 work *Ruling Passions*, Blackburn discusses his account of moral fallibility, an integral component of the wider notion of moral improvement, which is vital to the quasi-realist project as discussed in my previous chapter. In 2007, Andy Egan challenged this account of moral fallibility in his paper 'Quasi-Realism & Fundamental Moral Error'. In the following chapter, I will discuss Blackburn's account of moral fallibility in *Ruling Passions*, Egan's challenge and finally Blackburn's response to Egan in 2009. I will then discuss the reasons why I have concluded that while Blackburn's response to Egan does 'in principle' refute Egan's argument, there nevertheless remains the problem of asymmetry between the way the subject views others and the way they view themselves. This line of thought will also bring us to a further, far more recent paper written by Nicholas Smyth called 'Resolute Expressivism', which does appear to engage with this dilemma, which I will discuss in chapter nine. The task then is to ascertain whether Smyth's account is in any way compatible with Blackburn's, or whether we need to look at the two as wholly different.

According to Blackburn in *Ruling Passions*, there are always particular values that the moral subject admires (e.g. information, sensitivity, maturity, imagination, coherence etc), and the subject knows that other people may have particular difficulties with these values (Blackburn, 1998, 318). For example, there is being immature, insensitive, incoherent, overly predictable, ignorant etc. The subject also knows that they themselves are not exempt from having these difficulties. This is what Blackburn calls 'Moral Fallibility'; it is the acknowledgement that the subject is imperfect.

> Of course there is no problem thinking that other people may be mistaken, or are indeed mistaken. Anyone thinking that kicking babies for fun is OK is mistaken. The real problem comes with thinking of myself (or of us or our tradition) that I may be mistaken. How can I make sense of fears of my own fallibility? (Blackburn, 1998, 318)

So to reiterate what Blackburn is saying here, it is not whether or not the subject is really mistaken about their moral belief that is of concern per se; to suppose such would surely be a

realist claim after all. Rather, it is whether or not the subject can make sense of having a moral belief whilst believing that that same belief is **possibly false**. Enter Andy Egan, who claims that this is indeed problematic. According to Egan, we ought to be able to logically say something like 'I believe that kicking babies is wrong, but I could be wrong about that' (Egan, 2007, 12). However, quasi-realism does not appear to deliver this. This is because certain strongly held beliefs, such as 'kicking babies is wrong', are what Egan calls '*stable*':

> Call a belief *stable* just in case no change that the believer would endorse as an improvement would lead them to abandon it. Call a belief *unstable* just in case it's not stable; that is, just in case it would be abandoned after some change that the believer would endorse as an improvement. (Egan, 2007, 12)

In other words, a stable moral belief is one that is entrenched in such as way that no change in any of the subject's beliefs that led to a change in that belief, and which the subject would endorse as an improvement, could ever be **seen** by the subject as a moral improvement. For example, even if I had a change of heart about the morality of causing unnecessary harm (which is possible but unlikely), I still could not ever see that as a moral improvement if it leads me to conclude that kicking a cat (or any other sentient being) is not **wrong**. I **fundamentally** believe that kicking cats is wrong; therefore, any change in my morality that leads me to a conclusion that contradicts this is something I will always see as a moral regression rather than a moral improvement. Interestingly, this raises an important question. Is the subject **prior** to the change in values the one endorsing (or refusing to endorse) that change, or is it the subject **after** the change doing so? The answer to this question would go a long way to explaining the meaning of Egan's notion of 'stability' here. Egan does not explicitly clarify which one this is, but I suspect that it is the subject prior to the change that refuses to accept an endorsement of any change which leads to the contradiction of a fundamental moral conviction.

While at first this may not seem like a difficult issue, it becomes so when we consider this in the context of a certain kind of quasi-realist subject. To explain why this is such a problem for the quasi-realist, suppose that the subject's moral sensibility can be demonstrated to work something like a 'web of attitudes', this being the best way to explain the process that I have at this stage. The web of attitudes, as I have chosen to call it, looks not unlike a food web in biology. Attitudes towards moral issues are linked. So for example:

- Suppose the subject has attitude P (P: Boo! (homosexuality)) i.e. the subject disapproves of homosexual behaviour and/or people.
- Now suppose that the subject later develops attitude P' (P': Boo! (P)) i.e. the subject disapproves of their previous negative attitude towards homosexuality.

These kinds of attitudes towards attitudes are how Blackburn proposes to solve the problem of moral fallibility[22]. To explain how this works it greater detail, we can picture the following:

- Suppose at t1 (time index 1), the subject has the following group within their web of attitudes:

  t1 {**P**, Q, R, S, T, U...} where P is (P: Boo! (Homosexuality)), and Q, R & S etc are all other attitudes that are linked to P, e.g. Q is (Q: Hurray! (Perceived natural order)) and R is (R: Hurray! (Traditional gender roles)).

- Now suppose at t2, the subject's attitudes have changed:

  t2 {P, Q, **~R**, S, T, U...} where R has changed to ~R, where ~R is (~R: Boo! (R)).

  So suppose here that the subject's perspective on gender roles and gender stereotypes has changed. This leads to a kind of quasi-realist inconsistency with P, not unlike disapproving of lying, yet approving of conjunctions involving getting one's brother to lie (see 'chapters five and six').

- As a result, we then progress to t3, where the subject views themselves as having made a moral improvement:
  t3 {**~P**, Q, ~R, S, T, U...} where ~P is equivalent to P' i.e. the realisation of moral fallibility and need for moral improvement, (~P: Boo! (P))

[22] This is very similar from the first and second order attitudes described in chapter seven.

There may even be a t4 {~P, **~Q**, ~R, S, **~T**, U...}, where a change in R has resulted in a similar change to several of the attitudes that are linked to R, including P, within the web of attitudes.

Egan's challenge is simply that there may be instances where P' is not present, and this creates a dangerous asymmetry between the way the subject views themselves and the way they view others. In order for there to be room for improvement, that is to be morally fallible, we must allow for the subject to believe that it is possible they are mistaken on some level. This is problematic, because for the subject to believe they may be mistaken, their belief must be **un**stable. Egan claims that for the subject to believe they are fundamentally in error, they need some moral view that is both stable and yet believe that are **possibly mistaken**, and that this is impossible under Blackburn's account of moral error (Egan, 2007, 212). According to Egan, for a moral belief to be stable, it must be impervious to 'improvable' change (i.e. change the subject would approve of). For a moral belief to be fallible, the belief must be subject to improving change that could lead the subject to abandon it. On Blackburn's account, Egan claims, a moral belief may be fallible if and only if it is not stable, so for the subject to believe they could be fundamentally in error, the subject must have some moral view that is stable yet not stable. In other words, a reflective quasi-realist cannot believe themselves to be fundamentally morally mistaken. This is clearly unacceptable (Egan, 2007, 212). The subject ought to be able to logically think, 'I believe that x, but I may be wrong about that', in any and/or all cases. However, where x is **stable**, i.e. 'I believe that x, because x is fundamental, and I could never view any change that leads me to ~x as an improvement', here, according to Egan, it cannot make sense to believe that x **may** be mistaken. What we need is to be able to logically say, 'I believe in x, and I could never endorse any change that leads me to ~x as an improvement, but I **may** be wrong about x.' According to Egan, quasi-realism fails to deliver this.

- So let us look once again at t1 within this context, where K is something far less controversial and fundamental like (K: Boo! (Kicking Cats)), and where Q, R & S etc are all the attitudes that are linked to K:

  t1 {**K**, Q, R, S, T, U...} Call K **stable**; there is no way to change K that the subject would endorse as an improvement i.e. (Boo! (Kicking Cats)) is <u>fundamental</u>.

- Now suppose that, as above, at t2 some attitude that is linked to K, e.g. T, has changed:

  t2 {K, Q, R, S, **~T**, U...} Where T is e.g. (T: Boo! (Causing Pain)), and ~T is (~T: Boo! (T)) [This is not inconceivable, there are instances where causing pain may be seen as morally acceptable e.g. if it is an inescapable result of a necessary medical treatment].

- This could, in theory, then lead the subject to t3:

  t3 {**~K**, Q, R, S, ~T, U...} where ~K is (~K: Boo! (K)) i.e. the subject disapproves of their previous negative attitude towards kicking cats.

The problem is that whilst this change in K uses the exact same mechanism as the previous change in P, it is unlikely that any sane individual would see any change in T as a moral improvement **if** it leads to a change in K, K being a fundamental belief. In the first instance, the subject comes to a conclusion of ~P i.e. they abandon their previous negative attitude towards homosexuality, and this is seen as a moral improvement by the subject. In the second instance, the exact same mechanism is used, but in this case it leads not to a moral improvement but rather a moral regression, where the subject abandons their fundamental attitude that kicking cats is immoral. This seems to demonstrate quite clearly that Blackburn's mechanism for moral fallibility does not work in all instances. So either we need an account where T is allowed to change without changing K, or an account where any change in T that results in a change in K is seen as regression rather than improvement. The problem is that Blackburn's mechanism for moral improvement ought to be universally applicable, but as we can see, this does not appear to be the case.

Egan's argument seems to be that a subject can believe that a given belief of theirs might be mistaken if and only if it is not stable. A belief is stable if the subject cannot bring themselves to accept that it could be false, whether it is in fact false or not. The dangerous asymmetry comes from the fact that it should at least be possible, according to Egan, to think that one may be mistaken in any and/or all case(s). If Egan is correct, however, then quasi-realism does not allow for this possibility in all cases, specifically where stable beliefs are concerned.

It is more likely that the subject will define their own beliefs as stable than those of other people, having a higher degree of confidence in their own attitudes. This is because their own attitudes inevitably influence their perceptions of both themselves and others. Put simply, it is easy to find room for improvement in the beliefs of others, because one has an outsider's perspective. It is much harder to form such evaluative judgements about one's own beliefs, because the apparatus of one's evaluative judgements rely on the very beliefs that one is trying to evaluate. Herein lays the dangerous asymmetry. In other words, Egan highlights the issue that a moral belief, whether truthful or mistaken, may be impervious to improvable change and thus be defined as stable under a quasi-realist context.

Blackburn attempts to refute Egan by clarifying the notion of stable & unstable moral beliefs. According to Blackburn, Egan targets the following; (M) if something is entrenched in my outlook, in such a way that nothing I could recognise as an improvement would undermine it, then it is true (Blackburn, 2009, 204). Egan targets (M) in the belief that (M) is something that Blackburn asserts. In fact, Blackburn agrees with Egan on this point, but (M) is not a claim Blackburn ever wanted to defend. Blackburn claims that (M) is clearly false. I can believe kicking babies for my amusement is morally permissible, & I can fail to recognise any improvements to my values that would undermine this belief. Instead Blackburn asserts: (I) if something is entrenched in my outlook, in such a way that nothing that is an improvement would undermine it, then it is true. According to Blackburn, Egan's attack on (M) leaves (I) unscathed. (I) does not have the same problematic asymmetry as (M), because (I) is a claim for which the impersonal holds the same logical structure: (I') if something is entrenched in anyone's outlook, in such a way that nothing that is an improvement would undermine it, then it is true. (Blackburn, 2009, 204)

While this does in principle refute Egan's argument, Blackburn later concedes that there nevertheless exists an asymmetry between the subject and others, but in a relatively benign way, and this seems to be largely to do with the notion of bias (Blackburn, 2009, 213). It is impossible for the subject to make an evaluative judgement on their own beliefs and opinions without employing those same personal beliefs and opinions to do so. However, it is in principle possible to make evaluative judgements about the views of others without employing the others' views, thereby giving an outsider's perspective. One must concede, however, that any such evaluative judgement is, like any other given by the subject, not free from the bias of their own personal opinions. In other words, in real life, there is no such

thing as an unbiased opinion. There is also another major issue with Blackburn's response. Egan writes of what the subject would **see** as an improvement. Blackburn in (I) writes of what **is** an improvement. Is Blackburn entitled to do this? It does seem as though Blackburn has a tendency to slide between 'quasi' and 'realist' at times. As I mentioned at the end of chapter seven, it is very important for the quasi-realist not to stray into becoming merely an abstract form of moral realism, but rather be a non-realist account of moral language that happens to deliver all that moral realism does with the exception of saying that moral terms refer to moral qualities which **are real** (ontologically/metaphysically). I also mentioned that this danger will return to haunt the quasi-realist in section two when we attempt to pin down the quasi-realist on what exactly the **quasi** of quasi-realism means, and how exactly it is **different** from moral realism.

Blackburn's response may successfully refute Egan's challenge in the sense that the dangerous asymmetry appears, at first glance, to have been dealt with, but only if he is entitled to make the move that we have just raised questions about. However, doing so depends on what exactly Blackburn means when he says 'nothing that **is** an improvement'. If we look at this response in the realm of moral realism, it is clearly successful; a belief that kicking cats is morally permissible is objectively wrong; therefore, the fact that a person may hold to it merely means that they are fundamentally in error. However, quasi-realists specifically want to avoid making those kinds of commitments. On the other hand, if we look at the response from a non-realist/expressivist perspective, then it becomes more difficult; if a person is not asserting anything about the metaphysical nature of morality but rather are expressing their approval of kicking cats, where others are expressing their disapproval, can we truly call this quasi-**realism**?

Chapter 9: A Defence by Nicholas Smyth

In the last chapter, we discussed how Egan's challenge of moral fallibility and Blackburn's subsequent response seems to put pressure of the successfulness of the quasi-realist project. The whole purpose of quasi-realism after all is to give moral realists everything that they are seeking without having to make any metaphysical commitments. Therefore, it is absolutely vital that the quasi-realist has a clear definition of what is meant by moral utterances like 'ω is wrong', and how the content of those utterances function semantically. Enter the resolute expressivist. In Nicholas Smyth's 2014 paper, 'Resolute Expressivism', we are introduced to a unique approach to this conundrum; Smyth attempts to defend quasi-realism by painting it as a pragmatist method rather than any kind of metaphysical one. This is particularly intriguing, because it brings us squarely back to the method of using pure philosophy of language as our primary tool in explaining how the quasi-realist project may be successful.

> Over the years, we have witnessed the rise of a metaethical cottage industry devoted to claiming that expressivist analyses cannot capture some allegedly important feature of moral language. In this paper, I show how Simon Blackburn's pragmatist method enables him to respond decisively to many of these objections. In doing so, I hope to call into question some prevailing assumptions about the linguistic phenomena that a metaethical theory should be expected to capture. (Smyth, 2014, Abstract)

After giving us first a recap on Blackburn's position and then Egan's challenge, Smyth begins by stating that it is actually rather difficult to imagine the kind of **stable** belief that is immune to 'improvable' change that Egan cites as a problem (Smyth, 2014, 7). One possible candidate is what Smyth refers to as (E). The content of (E) is that if an action would cause human suffering, then this is a consideration that counts against performing that action; this is hard to disagree with. This indeed seems a worthy candidate for the kind of stable belief that Egan needs for his objection to stand. It is difficult to imagine any change in our beliefs that would lead us to abandon the idea that human suffering is an important thing to consider in any applied ethical issue that we could then endorse as a moral improvement. However, Smyth confronts this by offering the following thought experiment. How would we actually respond, as moralizers committed to (E), to an encounter with someone who could not in any way be brought to accept it? (Smyth, 2014, 8) This does indeed occur in the case of

sociopathy. In such cases, we usually declare the views of any opponents of (E) false by fiat. However, this is surely a response that we would do well to preclude, as this is dangerous; this seems very much like the kind of metaphysical commitment that we want to avoid:

> The opponent of Quasi-Realism may try to strengthen **(E)** to preclude this kind of response, but this move will almost certainly render the attitude in question non-stable. The more content we pack into the relevant ethical attitude, the more likely it is that some other attitude will bear on its possible revision. The less content we pack into it, the more likely we are to judge that a person who could not accept it is fundamentally morally defective. (Smyth, 2014, 8)

It seems then to be the case that Smyth, like Blackburn himself, agrees with Egan in that a stable moral belief (such as 'kicking cats is wrong') may be immune to the kind of change that the subject could endorse as a moral improvement. Unlike Blackburn, however, Smyth's response to this is not to attempt to refute Egan's argument by claiming that Egan has missed his mark by targeting (M) instead of (I), but rather to pull the rug from under Egan's feet by claiming that the problem of having stable moral beliefs is not really any problem at all:

> I may quite naturally think that I *am* immune from error. Now, this account of an ordinary response to fundamental disagreement is undeniably speculative, and I do not claim that this would be every person's response to fundamental disagreement. However, it *would* be unpardonably smug of Egan to claim sole right to armchair sociology. He, like many other critics of expressivism, freely speculates about "our ordinary ways of thinking about morality," yet it is not at all obvious that his claims about ordinary practise hold up once we appreciate how profound fundamental disagreement actually is. (Smyth, 2014, 9)

Smyth's approach is to consider normative moral discourse and how we use moral language pragmatically. Egan's challenge depends on the premise that we ought to be able to logically say things like 'I believe that kicking cats is wrong, but I could be wrong about that'. Quasi-realism he says does not deliver this; you cannot be wrong about something that is an expression of your own disapproval. I disapprove of kicking cats; I cannot be wrong about that. Any change in my moral sensibility that leads me to approve of kicking cats is

something I would always consider a regression rather than an improvement, because my disapproval of kicking cats is fundamental. According to Smyth, none of this is problematic. His response is that the descriptive claim at the heart of Egan's challenge is **false**. That quasi-realism does not deliver a means of logically saying things like 'kicking cats is wrong, but I could be wrong about that' is irrelevant, because that is not something quasi-realism needs to deliver. This is where the resolute expressivist digs their heels in. They have, in a 'pragmatist spirit', explored what actually occurs in the alleged counter example and examined our moral practice from the perspective of a participant. In other words, Smyth attempts to overturn an underlying presupposition that ordinary discourse is robustly realist, and that 'the revisionist in the room is usually the expressivist' (Smyth, 2014, 10).

So let us look again at the original dilemma by using a more pragmatic approach. I claim that 'ϖ is wrong'. Whether or not ϖ is wrong is up for debate, though the quasi-realist may well claim that the 'assertion' that ϖ is wrong is really a projection of their disapproval of ϖ, which would be a member of the unique, best possible set of attitudes (i.e. is also **true** in a quasi-realist sense). This may be the case were the quasi-realist's belief that 'ϖ is wrong' **stable** i.e. no change in their moral sensibility that led them to abandon this belief could ever be endorsed as an improvement. However, not knowing the unique, best possible set of attitudes, the quasi-realist is not exempt from moral fallibility. In fact, I would even argue that it is necessary for any sensible notion of moral improvement. We then raise Egan's challenge. How can the quasi-realist make sense of the fact that they are morally fallible given that their belief that 'ϖ is wrong' is stable? Smyth's response is simply to say that they cannot, but that is fine, because they do not need to. In other words, we indeed conclude that the quasi-realist, in the case of stable moral beliefs at least, may be immune to the belief that they are fundamentally in error.

To test this, let us provide ϖ with some content e.g. ϖ [homosexuality]. Suppose then we have a homophobic quasi-realist who fundamentally believes that homosexuality is wrong. No change in their belief system that led them to approve of homosexuality could ever be endorsed as an improvement by them; their belief that it is wrong is so entrenched. Many examples of this kind of homophobia do indeed exist, and interestingly when we look at these cases in real life, rarely does such an individual consider that they **might be wrong** about homosexuality. The question we are interested in, however, is whether such an individual needs to consider this in order to have such a view. The answer would clearly appear to be

no. Many people recognise that they are morally fallible in a ‘we are all sinners’ kind of way, but that does not stop many people having beliefs that are so entrenched in a fundamental way that they would never consider that they are wrong about them. So both philosophically, and sociologically, we may well be inclined to believe Smyth in this case.

Chapter 10: Initial Conclusions

As mentioned in my introduction, the aim of the thesis is to eventually arrive at a non-realist theory of moral language which can account for realist sounding discourse without compromising its own non-realism, and which avoids the Frege-Geach problem. I began by examining the case against moral realism and for non-realism, some of the alternative non/anti-realist approaches, including error theory, and I have decided to look at expressivism in particular. In doing so, we have explored non-cognitivism and projectivism, the Frege-Geach problem and Blackburn's quasi-realism. Thus far in the thesis, it looks as though quasi-realism is the most promising of the non-realist approaches. Quasi-realism aims to give any realist in the room all that they want to hear about the nature of moral language, but it does this in a different way to what traditional moral realism offers. Rather than claiming that moral adjectives refer to moral qualities present in the world and in the moral acts themselves, quasi-realism claims that moral adjectives are projections which function very similarly to standard, fact oriented assertions, and this is accountable via a special logic of attitudes. Using this logic, Blackburn then goes some stages further, offering a quasi-realist account of moral fallibility, moral improvement and even moral truth.

We then looked at some of the problems with quasi-realism, in particular those brought up by Andy Egan, who targets Blackburn's claim of being able to account for moral fallibility. This is a particularly critical target, as without this account his ability to account for moral improvement and moral truth also fall apart by chain reaction. Here we see Nicholas Smyth enter the debate with his own take on quasi-realism, and who claims a form of resolute expressivism is a solution to Egan's problem. He does this by challenging the assumption that a quasi-realist must be able to recognise a seemingly stable moral attitude as fallible. I am inclined to agree with this approach, because it seems entirely plausible that someone with an attitude sufficiently entrenched may be disinclined, or even completely unable to seriously contemplate the fallibility of that attitude. However, my worry here is that this approach may or may not be consistent with Blackburn's own views on quasi-realism, and it is these kinds of uncertainty that will in section two prove detrimental to quasi-realism's survival.

Moreover, it is not entirely clear whether resolute expressivism is really quasi-realism. In the question of what exactly the quasi of quasi-realism means, it is clear that under resolute

expressivism, quasi-realism is not merely an abstract form of moral realism, but rather a form of hyper sophisticated and projectivist expressivism. However, in order for quasi-realism to escape the Frege-Geach problem, we need to look at Blackburn's account and his logic of attitudes, and this includes his account of attitude inconsistency and quasi-logical moral truth conditions. It is these elements in particular in which it is unclear whether or not they are consistent with resolute expressivism, because these elements seem to push quasi-realism beyond simple expressivism and towards something else, and it is frustratingly ambiguous what that thing ultimately is.

So in conclusion thus far, quasi-realism does look very promising on paper, and from what we have looked at thus far it certainly looks like a good account of non-realist discourse. However, we leave section one with some worries, namely as to what exactly **quasi**-realism means, how is it sufficiently different from realism, and in what ways does it offer a realist what they want without compromising its own non-realism. These are questions I will tackle in my next section, where we begin to look at more problems for quasi-realism and what ultimately compels us to leave quasi-realism in search of another semantic explanation of moral discourse: moral fictionalism. However, it is important that we understand as much as we possibly can about quasi-realism here in section one, because as we will begin to see by the end of my next section, moral fictionalism is lacking in some important areas (more on that in section two), and there are key elements which we have looked at from quasi-realism which might be able to augment it significantly. It is the task then of section three to marry the two accounts together and prove that they are not quite as mutually exclusive as one might come to expect.

Section 2: Quasi-Realism vs. Moral Fictionalism

Chapter 1: Quasi-Realism & Moral Fictionalism

In section one, 'Non-Realism, Expressivism and Quasi-Realism', I have outlined the aim of the thesis, some of the useful terminology used and have outlined some of the different non-realist approaches to moral language, the nuances of expressivism and projectivism and have gone into considerable detail about The Frege-Geach problem and Simon Blackburn's quasi-realism, as well as some objections to it. The aim of the thesis, as I have explained at the beginning of section one, is to eventually end up with a functional theory of non-realist moral language. Thus far, quasi-realism looks like the best possible candidate for this. Blackburn claims to be able to provide the realist in the room with all of the answers that they want to hear and to provide the non/anti-realist with reassurance that they can preserve a healthy scepticism of moral ontology (Blackburn, 1984, 183). Blackburn also claims to be able to handle the Frege-Geach problem. Quasi-realism also has an interesting story to tell us about not only how a non-realist uses moral language in the way that they do, but more importantly **why**, as quasi-realism portrays itself as a form of projectivism (speakers project attitudes of approval and disapproval onto the world and objects in it via moral sentences). In other words, this story tells us that we use moral language to communicate attitudes of approval and disapproval (expressivism), which on the face of it at least sounds perfectly plausible. Blackburn even claims entitlement to being able to talk about notions of moral improvement and moral truth, again without committing to any moral ontology in the process. He does this by creating a logic of attitudes which is meant to mimic the conventional logic of fact asserting propositions, but containing content not of assertions but of expressions. We then investigated the objections of Andy Egan and a valiant rebuttal and defence of Blackburn by Nicholas Smyth. In conclusion, we leave section one with a fairly positive outlook on quasi-realism as being the primary candidate for a functional non-realist account of moral language.

Now, in section two, the thesis begins to change direction as we begin to look into those objections to Blackburn that are **not** so easily thwarted. I shall begin with two chapters devoted to the objections of G.F. Schueler. Schueler argues that Blackburn's solution to the Frege-Geach problem is ultimately flawed and that there is a larger problem also with the

kind of embedding that Blackburn does in that solution. Whilst Blackburn appears to successfully counter argue many of Schueler's points in *Essays in Quasi-Realism*, he also exposes some potentially fatal flaws in quasi-realism that I have noticed, which I will outline in chapter three. The main issue in that regard, as I will explain therein, is that Blackburn's solution to the Frege-Geach problem hinges on his ability to demonstrate that attitudes themselves can be **inconsistent**. It is not at all clear that he has done this. The example I use is of a monk who takes an oath of celibacy. The monk makes a personal moral commitment to refrain from sexual activity, but this does nothing in regards to his attitudes or his intentions towards the existence of sexual activity in general. This also helps us to understand the difference between attitudes and intentions towards actions. Attitudes can be hypocritical easily enough, but I argue that it is only actions which can be inconsistent in the way that Blackburn describes. This can also be applied to Blackburn's favourite example of inconsistency which involves approving of getting one's brother to lie. It may not be inconsistent to compel one's brother to lie on one's behalf if the reason for the original disapproving attitude towards lying comes from a personal commitment to refrain from lying rather than a universal commitment towards a world free from liars. It is only inconsistent to hold an approving attitude towards the brother lying on one's behalf if one is committed to act only in accordance with that attitude and towards a world free from liars. A mere attitude, whether approving or disapproving, is insufficient to render this kind of strong inconsistency. All of this and more will be outlined in more detail in chapter three.

In chapter four, I will examine the objections of Van Roojen, who also attempts to undermine Blackburn's ability to demonstrate attitude inconsistency. Unlike Schueler, Van Roojen focuses his assault on Blackburn's account of moral fallibility. This is a particularly vulnerable pillar in the quasi-realist project, because without a functioning account of how quasi-realists can be morally fallible, there cannot be a functioning account of moral improvement, and without that there cannot be a functioning account of moral truth. Without a functioning account of moral truth, Blackburn cannot gain entitlement to give everything the moral realists want to hear. Therefore, any fault line in Blackburn's account of moral fallibility may prove calamitous by virtue of chain reaction throughout the rest of the project and its objectives. Luckily for Blackburn, much like Andy Egan, Van Roojen also inspires a defence by Nicholas Smyth. This defence does appear to pay off, but as I will discuss in that chapter, it is once again not clear that it is really **quasi-realism** that Smyth is defending so much as 'resolute expressivism'.

A similar theme arises in chapter five, where I will examine the battle between Blackburn and David Lewis. Much like Nicholas Smyth, David Lewis attempts to 'rescue' Blackburn. However, unlike Smyth's response to Van Roojen and Andy Egan, wherein Smyth champions what he calls the 'resolute expressivist', David Lewis's strategy involves shifting quasi-realism away from expressivism altogether and towards something called 'moral fictionalism', noting that many of the moves that Blackburn makes are remarkably similar to those that moral fictionalists make. Blackburn has a very interesting counter to this idea, however, which involves using Lewis's own argumentative strategy against him. Lewis claims that to separate the quasi from the realist, the quasi-realist must outline things which the realist claims which the quasi-realist does not. In response, Blackburn claims that to separate the quasi-realist from the fictionalist, there are some things which the fictionalist claims which the quasi-realist does not. I believe that Blackburn does successfully demonstrate that quasi-realism is not the same thing as moral fictionalism. However, therein lies another problem for Blackburn.

Returning to my example of the monk and the oath of celibacy, it is not at all clear that Blackburn can demonstrate attitude inconsistency in the strong way that he needs. This has two potentially calamitous results for the quasi-realist. First of all, it means that the solution to the Frege-Geach problem becomes much more suspect, as explained in chapter three in my examination of the objections of G.F. Schueler. Second, it makes Blackburn's account of moral fallibility, which we will discuss in chapter four in my examination of Van Roojen, similarly suspect. This is particularly worrying for the quasi-realist, as it may cause a chain reaction splintering the entire project, and as I will explain in chapter five, I believe it is this tendency to depend upon multiple layers of accounts which are so linked together that makes quasi-realism particularly vulnerable. The other problem that I believe quasi-realism suffers from, which I will also explain in chapter five is its refusal to accept any kind of comparison with moral fictionalism, which appears to naturally avoid many of the problems that quasi-realism faces and does not have any problem with 'biting the bullet' on certain difficult questions which quasi-realists incur so much cost to circumvent. In contrast to quasi-realism, moral fictionalism is not in the habit of pandering to moral realists or providing elaborate explanations about quasi moral truth(s); in that regard it is a non-realist account of moral language in the purest sense.

This then leads me to a dramatic shift in the thesis where I begin to heed the advice of David Lewis and abandon quasi-realist semantics in favour of moral fictionalism, wherein I outline most of the nuances of moral fictionalism as described and championed by Richard Joyce throughout chapters six and seven. In these chapters I will explain what fictionalism is, the advantages it has over quasi-realism, as well as useful terminology and concepts such as the tacit operator, pretence assertion as well as the difference between hermeneutic and revolutionary fictionalism. Then, throughout the remainder of section two, I will examine some of the notable objections to fictionalism and the problem of fictional discourse in chapter eight, and the objections of Jonas Olson and his alternative account of 'conservationism' in chapter nine.

In chapter nine, however, there arise some objections to Joycean moral fictionalism that lead me to defend it in a way similar to what Smyth does in regards to objections to quasi-realism. It is at this point that I begin to describe an alternative version of Joycean moral fictionalism which I later in chapter ten call '**expressive fictionalism**'. This is owing to another observation that what fictionalism excels at in its simple and effective approach to moral semantics is also its biggest weakness. Why, the reader may ask, if we are turning to fictionalism, were we so preoccupied with quasi-realism to begin with? The answer to this is simple. What fictionalism lacks, which quasi-realism excels at, is in telling a convincing story about why we use moral language in a realist **sounding** manner if we are not talking as realists do. Expressivism has a simple enough answer to that question, but as I have spent half a thesis discussing, in even its most sophisticated versions (i.e. quasi-realism), it is ultimately very vulnerable to numerous problems including Frege-Geach and others on a semantic level, which fictionalist semantics appears to naturally avoid. On a pragmatic level, however, most or none of these issues seem to apply. This then leads me to what I hope is a unique solution to all of these problems. Why not simply combine fictionalist semantics with a form of **pragmatic** expressivism? As I will explain in chapter nine, I believe this successfully takes care of the problems outlined by Olson. The task then is to actually build this kind of theory, which will be what I will be attempting to do in my third and final section.

Chapter 2: G.F. Schueler

In G.F. Schueler's article "Modus Ponens and Moral Realism", Schueler begins by presenting the debate between moral realism and anti-realism and explains that what he will focus on is the anti-realist positions of expressivism and projectivism (Schueler, 1988, 492). Thus, like Simon Blackburn, the debate is a linguistic one rather than a metaphysical or ontological one. In other words, we are less concerned here with the existence of objective 'moral facts' or 'moral properties' than we are with the semantic and syntactic meaning of moral sentences. However, Schueler holds reservations about Blackburn's account of 'indirect contexts', which include cases of embedding like those we saw in section one, claiming that Blackburn's account is incomplete, defective and that nothing similar could possibly succeed (Schueler, 1988, 492). To provide some context to these issues, he quotes Blackburn directly, "Nobody denies that the surface phenomena of language-the fact that we use moral predicates and apply truth or falsity to the judgments we make when we use them---pose a problem for projectivism" (Blackburn quoted in Schueler, 1988, 492). In other words, it seems strange to talk about morality in terms of what is right and wrong, in a seemingly realist manner, when really it is an **expression** of our disapproval of certain behaviours i.e. the expression of an attitude. As Blackburn explains in *Spreading the Word: Groundings in the Philosophy of Language*, this issue inspires the quasi-realist project, which aims to explain why we use seemingly truth conditional moral sentences in the absence of truth conditional moral content (Blackburn, 1984, 170-171). As demonstrated by the Frege-Geach problem (see section one chapter six), one of the major difficulties for quasi-realism is the embedding of sentences in indirect or unasserted contexts. For both Geach and Schueler, the modus ponens case reveals a particular problem for embedding in general.

Schueler's apparent difficulty with Blackburn's solution to the Frege-Geach problem is that it appears to treat modus ponens in a different way than what its logical form allows; in other words the problem lies in how we interpret modus ponens as a logical argument designed for a specific form of logic, or as a one that can be adapted into more unconventional uses. According to Schueler, if showing that an inference is modus ponens involves taking it 'as the realist picture' has it, then it cannot be explained in any other terms (Schueler, 1988, 495). Either the use of modus ponens is a valid one, or it is not, and that demands the use of certain truth conditions. If logical validity means that it is impossible for the conclusion to be false if

its premises are true, then those reluctant to apply truth and falsity to the premises of a syllogism will have to admit that the inference is invalid. Furthermore, Schueler seems to take issue with Blackburn's use of terms like 'commitments', in that it is not immediately clear what Blackburn is talking about. According to Schueler, it throws doubt on the **quasi** aspect of quasi-realism in that either Blackburn is committing to a moral truth, in which case he is a realist, or he is 'committing' to something different, in which case he is an anti-realist (Schueler, 1988, 495).

Blackburn has counter argued that none of these arguments are correct, but does concede that his work is far from complete and is certainly worthy of some degree of recalibration. *In Essays in Quasi-Realism*, Chapter 10: 'Attitudes and Contents', Blackburn addresses these concerns. For example, Blackburn claims that it is not as much of a given what it is for an argument to have the logical form of modus ponens as Schueler would have us believe (Blackburn, 1993, 182-183). Compare its form in propositional logic: "P, P→Q: Q" with a more applied form: "P→Q is the commitment of one who attributes a high probability to Q conditional upon P". Which of the two better reflects the rule of modus ponens? If we focus on one at the expense of the other, then it immediately becomes controversial whether language includes any logical inferences when applied in real life. However, if we accept them both, then modus ponens becomes compatible with numerous explanations of the semantics of its components (Blackburn, 1993, 182-183). Blackburn also addresses Schueler's concerns about commitments by explaining that the objective or subjective format of a commitment may vary. For example, the commitment that a Republican succeeded Carter is different from, say, the commitment to go for a jog once a week. I would add that the latter is wholly mind **dependent** on making the **decision** to go for a jog. In the context of meta-ethics, we are focused primarily on this latter form of commitment, and this is where the **quasi** aspect of quasi-realism can be found.

As explained in *STW*, the key to Blackburn's defence is to show that attitudes may be consistent or inconsistent with each other. For example, where one holds a negative attitude towards lying, as well as the positive attitude towards the conjunction of negative attitudes towards lying and getting one's brother to lie, then one must also, according to Blackburn, hold the negative attitude towards getting one's brother to lie (Blackburn, 1984, 189). If they do not, then their sensibility is 'fractured' in the sense that they cannot logically fulfil any practical aims from these conflicting attitudes. I would add that it is not logically impossible

to both want to win a competition and want to lose it; in fact, such conflicting attitudes are common. However, one cannot perform a deliberate act with the **intention** of both winning and losing; the agent must make a **choice**. This clarification is not recognised in *STW*, however, because Blackburn neglects to clarify what kind of practical aims are made unattainable by a conflict of attitudes. This encourages Schueler to challenge Blackburn on the notion that such a situation is an example of a logical mistake, rather than it being a moral failing on their part. If, however, there is only a moral error, not a **logical** one, then Blackburn's analysis of modus ponens has failed: he cannot explain why we logically ought to accept the conclusion if we accept the premises. I will return to this important point later.

Blackburn's continued defence in 'Attitudes and Contents', against Schueler's objections concern, as Blackburn puts it, Schueler "attacking the detail of my treatment" (Blackburn, 1993, 184). The 'treatment' is Blackburn's quasi-realist explanation of moral discourse. This involves differentiating between two methodological approaches within quasi-realism itself: what Blackburn calls fast and slow track (Blackburn, 1993, 183). Fast track quasi-realism involves earning the right for the truth predicate to be applied to moral sentences and then recognising that once this is established as legitimate then this frees the quasi-realist to legitimately talk about other seemingly realist things (e.g. moral truths), despite actually meaning **attitudes** towards certain things (e.g. behaviours or other attitudes that we either approve or disapprove of). This is similar to Blackburn's methodology in *STW*. Slow track quasi-realism is far more meticulous. Rather than seeking to earn the right to a truth predicate in a single move, slow track quasi-realism involves demonstrating the acceptability of various kinds of apparent realist discourse in stages. This involves devoting energy to showing how sentences taking the form of expressions of attitudes, like 'Stealing: Boo!', may function in a way similar to antecedents of conditionals, and therefore as premises of valid, logical arguments. The purpose of this distinction seems to be to portray Schueler as attacking fast track quasi-realism rather than the more meticulous slow track version, as slow track quasi-realism involves steps which Schueler does not touch upon in his arguments. This sees Blackburn begin the convoluted task of constructing a specific logic of attitudes, the aim of which is to mirror the logic of proposition-asserting sentences.

The first step of this involves assimilating attitudes into the kind of logical relations between imperatives and norms. This involves imagining various scenarios within possible worlds (not entirely unlike modal logic semantics). Norms may be satisfied, and imperatives may be

obeyed. The question is whether two different norms may be satisfied in the same possible world. For example, if we use Blackburn's 'Hurray!' and 'Boo!' operators once again, I may hold H!(p) and B!(p), in other words I hold both positive and negative attitudes towards p, and thereby express that 'p' is a goal, and '~p' is also a goal. If B!(p), then an ideal possible world is one in which '~p' is the case, where p does not exist or occur. If H!(p), then an ideal possible world is one in which 'p' is the case, where p does exist or occur. Thus, the H!(p) and B!(p) may be inconsistent, as there is no possible ideal world in which p both exists and/or occurs, and does not exist and/or occur. In other words, the attitudes 'contravolit' each other, as is the term coined by R.M. Hare. Where Schueler challenges the notion of attitudes being inconsistent in this way, Blackburn does not. He also makes a distinction between desires and wishes (Blackburn, 1993, 203). Wishes are not directly related to action, but rather they are mental states, in which a person may hold a positive attitude towards something but has no intention of making that thing a reality. Desires, on the other hand, are linked to intention, where the agent has the goal of making their desire a reality. To explain this, I would put forward the following example: imagine you visit a friend's house, and they let you play one of their computer games, which you instantly fall in love with. It is an extremely rare game which is no longer sold in any stores, and you do not know of anyone who sells it on eBay. Because they are your friend, they let you borrow that game, with the understanding that you will return it in one week. After a week and a half of overzealous gaming, you realise that your friend has completely forgotten that you borrowed it, as they have not asked for it back yet. You wish to keep the game, but you also do not wish for your friend to remember that you have it, to get annoyed, fall out with you over it, and to not be your friend anymore. These two conflicting wishes may coexist in your mind with no inconsistency whatsoever. However, they will stress you out a great deal, because you also realise that ultimately a decision must be made. This game has a security key as part of its program which prevents you from copying it to your hard drive, and your friend is not the sharing type, and as a result you cannot both keep it and return it. There is no action you can take in which you can keep the game and see it returned to your friend. Therefore, you must **choose** whether your desire to keep the game is worth the risk of losing your friend.

According to Blackburn, this kind of logic also applies to beliefs. You may be open to the possibility of inconsistent things, but you cannot hold active beliefs about them, otherwise this will result in an inconsistent world view (Blackburn, 1993, 215). You may be open to the possibility that Elvis is still alive, and you may be open to the possibility to he is not, but you

cannot believe that Elvis is both alive and dead at the same time. The purpose of both beliefs and desires is to guide one's actions; I will return to this later on in this section. If you believe that Elvis is alive, and you want to get his autograph, then you may well choose to seek him out. However, you will discover that finding him will be extremely difficult, given that he has been dead for over three decades. What is worse is if you want to get Elvis's autograph, despite knowing that he is already dead. In such a situation, you can only wish that you could get it; you cannot desire/intend to get it.

If this logic holds, then Blackburn has demonstrated that attitudes may be consistent or inconsistent with one another, and if this is true, then it goes part of the way towards providing another layer of defence against the Frege-Geach problem. It means that Blackburn can counter the problem of modus ponens in unasserted contexts by demonstrating that conflicting desires may result in inconsistent aims. Whilst it may be logically consistent, albeit morally suspect, to hold a negative attitude towards lying and hold a positive attitude towards getting one's brother to lie, it is more problematic to disapprove of lying and approve of sensibilities that conjoin disapproval of lying with disapproval of getting one's brother to lie, as well as approving of getting one's brother to lie (Blackburn, 1984, 189). Here we see a potential inconsistency within all the attitudes involved in the modus ponens case. A directly asserted inconsistency based on propositional modus ponens is not necessary to demonstrate how this is the case, nor would it work as the Frege-Geach problem so eloquently shows. Furthermore, Blackburn shows how holding both such desires is incompatible with any logical action one could take to satisfy both desires simultaneously. He does this by asking us to imagine a possible world in which there is a best possible action taken to fulfil both desires. If there is a desire that there be no lies, 'B!(Lying)', then the best possible world is one in which lying does not occur '~Lying', but if we approve of getting our brother to lie 'H!(Brother lying)', then the best possible world is one in which lying does occur. There is no possible world in which lying both occurs and does not occur. Therefore, the desire to get one's brother to lie, whilst desiring that there be no lies, is practically inconsistent and results in a fractured sensibility.

Chapter 3: Desires, Wishes and Intentions

Blackburn's defence hinges on a sense of 'must' in the context of desires being 'inconsistent'. If desires merely conflict, but are not logically **inconsistent**, then there is no logical force which Blackburn can use to infer that in a case of a) disapproving of lying and b) approving of sensibilities that conjoin disapproval of lying with disapproval getting one's brother to lie, one 'must' accept c) disapproving of getting one's brother to lie. The individual who gets their brother to lie, according to Schueler, may be guilty of a moral failing, but is not guilty of a logical one. Thus far, I have attempted to explain Blackburn's defence in the sense that I understand it, that being that whilst 'wishes' may conflict with no inconsistency, 'desires' necessarily 'must' conflict for there to be a fractured sensibility like Blackburn describes. I have also explained Blackburn's concept of desires being related to action, in a similar way to intentions. However, Blackburn does not talk about intentions, but desires, and one of **my** objections is that when Blackburn talks about desires, it seems suspiciously like he is really talking about intentions. If we examine Blackburn's argument within the context of intention to act, then as far as I can tell it holds up. It does not seem logically consistent to **intend** to win a competition whilst simultaneously intending to lose it. It also does not seem logically consistent to intend to seek out Elvis to get his autograph whilst believing that he is already dead. It appears that Blackburn is attempting to find a link between positive and/or negative mental states and motivation for action and using 'desires' as this link. However, I argue that it is debatable whether desires really are sufficiently different from wishes.

Whilst I can have a deep emotional longing for Elvis's autograph, which if he was alive would give me more than enough motivation to seek him out, that kind of motivation is not necessarily enough to generate any action on my part. In order for it to make sense for me to seek out Elvis, I also need to believe that Elvis is alive. Otherwise, I would be deliberately embarking on a hopeless quest. However, my belief that Elvis is dead does not change the fact that I have a positive attitude towards the notion of getting Elvis's autograph. Similarly, I can have a positive attitude towards keeping my friend's computer game and have a positive attitude towards keeping my friend. This is not problematic. It is only problematic when I attempt to **act** in accordance with either of those attitudes. If keeping my friend is dependent on my returning the game, then I cannot keep both the game and my friend. This raises the question of whether it is truly the **attitudes** that are incompatible or whether it is really the

actions that are acted in accordance with those attitudes that are incompatible. Blackburn clearly sees that such actions are incompatible and attempts to link them to attitudes in order to show how attitudes may themselves be incompatible. However, it is not clear that separating desires from wishes proves that desires are necessarily linked to action. The notion of incompatible beliefs, intentions and attitudes, however, may prove useful to Blackburn in this case. Referring back to my Elvis thought experiment, merely wanting Elvis's autograph is not enough to produce action; belief is an essential component of motivation for action. As such, belief may be the important link that Blackburn is looking for. Unfortunately, this may not be enough to save the notion that desires can be inconsistent in the quasi-logical sense that he needs in order to successfully provide a quasi-realist account of embedding and related linguistic phenomena.

The central question throughout this chapter is, "what kind of mental state do we need for there to be a **logical** inconsistency here?" It is not controversial that attitudes can in many ways be hypocritical, but are they inconsistent? If we evaluate this in the context of wishes, these being mental states surrounding particular actions with no intention of making those actions transpire, then no, there does not appear to be a logical inconsistency so much as a moral failing. According to Blackburn, it becomes inconsistent when they are desires with the intention of action. However, he never specifies this, but merely separates desires from wishes and tells us that there is a link between desires and actions. He does this by using his logic of attitudes. There is no possible world in which lying both occurs and does not occur; therefore, it is illogical to desire and intend to act towards an ideal world that cannot logically exist. However, I would like to examine this somewhat deeper.

It is not explicitly addressed, in Blackburn or Schueler, what the reasons are for the individual having a negative attitude towards telling lies or specifically what form that takes. If the individual holds a negative attitude towards **all** lying, no matter what form it takes, then of course such a desire is incompatible with the intention of getting one's little brother to tell lies; B!(all lying) and H!(no lying) is not compatible with H! (lying), that is, if there is intention to act upon both these attitudes. If you have a negative attitude towards the existence of lying and a positive attitude towards a possible world in which lying does not exist, then it is surely inconsistent to act in accordance with a world in which lying takes place. Once again, however, the motivation of the individual who gets their little brother to tell lies is not dependent on this. Such an individual may not have a negative attitude towards

the **existence** of lying, but rather to **being a liar**. For example, a monk or a nun may take an oath of celibacy for the rest of their natural lives. To this end, they have a strong negative attitude to the notion of having sex. For that matter, they may also have a strong positive attitude toward that notion, but for the sake of their oath, they make a 'moral' choice to act only in accordance with the first desire, it being incompatible with actions taken towards the second. However, this is not to say that monks or nuns necessarily have a negative attitude towards the existence of sex, nor are they in the habit of acting with the intention of creating a world in which sex does not occur. They are happy to allow others to have sex but do not intend to take part themselves. Similarly, an individual might have reasons for not wanting to tell lies, but is not against others lying or lying taking place. Such an individual would have no problem getting one's little brother to tell the lies for them, and that would not be inconsistent with any desires on their part. I believe that this might prove problematic for Blackburn's defence; it is unclear what the precise content of the attitude is. Is it disapproving of lying in general? Or is it disapproving of being a liar? There appears to be a significant logical difference between mental states along the lines of 'would that there were no lying', which appears to have a wish like quality, vs. 'I will not lie', which appears to be a much stronger commitment to desire forming the basis for intended action. This could be represented via a Venn diagram scenario in which the subject does not object to the circle in which lying occurs but strongly objects to being placed within that circle.

My larger objection, however, concerns whether various different attitudes can be practically inconsistent with one another in the relatively strong way that Blackburn needs. I suggest that only attitudes closely related to producing action can. In an action related spectrum, with wishes on one end and intentions on the other, the questions become, 'Where do desires go?' 'Where does moral approval go?' 'What sort of approval is moral approval?' It seems that Blackburn wants to paint desires as significantly towards the intentions end of this spectrum, with moral approval being strongly linked to action. To Blackburn the purpose of moral approval is to guide action; we use moral sensibilities to outline particular rules or axioms by which we 'ought' to behave (see section one chapter seven). However, I would clarify this point by differentiating between personal moral commitments and general moral attitudes and suggest that only personal moral commitments can be inconsistent in the way that Blackburn wants. For example, I may express my moral disapproval of lying in general, but that does not logically force me prevent myself from getting my brother to lie for me. Unless I have an **intention** to act in accordance with an idealised, non-lying world, then there is nothing

stopping me from getting my brother to lie regardless of any **moral** qualms I may have about it. Indeed, it is surely a common occurrence to hold moral attitudes towards behaviours where we have no intention to act upon them. I may not be proud of getting my brother to lie, as I know that I have committed the kind of moral failing that Schueler is talking about, but that does not necessarily mean that I have committed a logical inconsistency. It is only if I have made a personal **commitment** to act only in accordance with a world in which lying does not take place that I 'must' avoid getting my brother to lie. Merely disapproving of lying in general is not sufficient to incur this kind of logical force. This also goes both ways, as shown by my monk example. A monk who takes an oath of celibacy makes a personal commitment not to engage in sexual activity, and so it is logically inconsistent to act in such a way that would satisfy both their sexual desire and their oath of celibacy. The sexual desire in and of itself, however, may not be incompatible with their oath, as they have no intention of acting upon such desire. For there to be a logical inconsistency between (a- disapproval of lying and b- approving of conjoining disapproval of lying with disapproval of getting one's brother to lie) and (c- approving of getting one's brother to lie), as Blackburn wants, Blackburn needs to show that disapproval and approval carry with them the intention to act in accordance with these kinds of moral attitudes, and as yet he has not done this.

Chapter 4: Van Roojen

In Nicholas Smyth's 2014 paper 'Resolute Expressivism', Smyth cites the argument of Mark Van Roojen in his 1996 paper "Expressivism and Irrationality", in which he claims that Blackburn's quasi-realist account of moral fallibility is **unacceptably** revisionary (so different from the original idea that it becomes lost). Like G.F. Schueler, who claims that Blackburn's solution conflates logical and pragmatic inconsistency (Schueler, 1988, 492), Van Roojen draws on what he calls our "pretheoretical judgments about logical implication" (Van Roojen, 1996, 331). In this chapter, I will briefly explain the stages of Van Roojen's argument and then discuss Smyth's response in 'Resolute Expressivism'. Just as is case with Andy Egan, Van Roojen strikes at the heart of Blackburn's account of moral improvement according to which in order to arrive at an account of moral truth, we need an account of moral fallibility (which in turn depends upon the possibility that attitudes can be inconsistent). Moral improvement depends upon the subject's ability to recognise inconsistencies within the 'web of attitudes' (see section one chapter seven) that form their moral sensibility.

Van Roojen begins by distinguishing between genuine contradictions and mere pragmatic ones, likening the problem to a Moorean paradox, in which sentences like "It is raining but I don't believe that it's raining" are greatly puzzling. Van Roojen asks us to consider the following two sentences:

(VR1) "Stealing is wrong, and stealing is not wrong"
(VR2) "I don't disapprove of stealing, and stealing is wrong"

Blackburn does not make a sharp distinction between the two, and it is fairly straightforward why. In a quasi-realist sense, sentences like "stealing is wrong" and "I disapprove of stealing" are not too different; the only real difference is that as the quasi-realist sees it, the former involves a **projection** of a disapproving attitude rather than merely asserting that they have that attitude. Therefore, a quasi-realist will no doubt see little significant difference between (VR1) and (VR2); both are examples of attitudes which contradict each other. However, Van Roojen claims that the two sentences are intuitively different to the extent that any account which maintains that they are inconsistent in the same way is unacceptably revisionist:

> To explain implications between expressions of attitude, Blackburn must invoke a notion of inconsistency strong enough to rule non-cognitive attitudes inconsistent, hence extending the notion of inconsistency beyond cases where the truth of one content rules out the truth of another. But the stronger notion invoked rules logically consistent claims inconsistent, precisely because it is stronger. (Van Roojen, 1996, 322)

The simpler explanation of this is that Van Roojen is **unconvinced** by Blackburn's attempt to create a logic of attitudes similar enough to the conventional logic of fact asserting propositions. Ultimately, these discussions always seem to boil down to the issue of what exactly constitutes 'moral inconsistency', or what **kind** of inconsistency is it that Blackburn requires for his account of moral fallibility to work. Where (VR1) clearly appears as a semantically **logical** contradiction, the implication being that stealing cannot be both wrong and not wrong at the same time, (VR2) appears as wholly different, because it is entirely consistent for the subject to approve of stealing but be mistaken in their moral attitude. Even the quasi-realist must concede that this is problematic, because moral fallibility depends upon the subject's ability to say, "I don't think stealing is wrong, but I may be wrong about that". This is similar to Egan's challenge in that it attempts to pin Blackburn down on what exactly he means by "**is** wrong", but it is different in that rather than appealing to the notion of stability of the subject's beliefs, it simply asks the question of whether an inconsistency within a logic of attitudes indeed caries the same weight as a **normal**, logical contradiction. Van Roojen claims that it does not.

Van Roojen frames the problem as a 'dilemma' (Van Roojen, 1996, 322). Either the 'logic' of attitudes is too weak to deliver logical inconsistency between conflicting attitudes, or it is strong enough to do this, but then it is too strong, making things that should not be logically inconsistent count as logically inconsistent. In fact, Van Roojen frames this dilemma in such a way that it is intended to affect non-cognitivism more broadly; he asks us to consider the following argument:

1. *If* I don't disapprove of Y, *then* X is wrong (Premise)
2. *If* X is wrong, *then* Y is wrong (Premise)
3. I don't disapprove of Y (Premise)
4. **Therefore**, X is wrong (From 1 and 3)

5. **Therefore**, Y is wrong (From 2 and 4)

This argument is valid, and it can be invalidated by the negation of its fifth statement and conclusion, ("Y is not wrong"). Now Van Roojen asks us to look at the translation of this argument offered by the non-cognitivist and by Blackburn's account that conditionals are disguised claims that one attitude involves another:

1'. If I don't disapprove of Y, then B! (X)
2'. The disapproval of X involves a disapproval of Y
3'. I don't disapprove of Y
4'. B! (X)
5'. B! (Y)

So to translate, Blackburn's explanation, according to Van Roojen, seems to be that to have (2'), (4') and ~(5') is inconsistent in some sense. However, surely having (3') and (5') is just as inconsistent in that this also seems to display a contradiction of attitudes. This would in turn mean that (1'), (2') and (3') must also be inconsistent since it implies that (3') and (5'). So on one hand, according to Van Roojen, Blackburn has given us an explanation of the validity of the argument only if he has given one in which the **premises** are inconsistent. However, the premises of the original argument both seem consistent and support a valid argument. So to reiterate, in order for the logic of attitudes to be strong enough to deliver logical inconsistency between conflicting attitudes, then this involves making things that should not be inconsistent count as inconsistent. Van Roojen claims that Blackburn has failed to deliver an account that escapes this problem, because as it is if the premises of the original argument are consistent, then the premises of Blackburn's translation of that argument should be also, and according to Van Roojen they necessarily are not (Van Roojen, 1996, 320).

Next, Van Roojen asks us to consider one more argument:

P1. It would be wrong for me to believe ill of my friends. (Premise)
P2. My parents, father and mother alike, are my friends. (Premise)
P3. It would be believing ill of a friend to believe he would be duplicitous with another one of his friends. (Premise)

P4. If the coded valentine is not a joke, my father is being unfaithful to my mother, hence duplicitous. (Premise)
P5. The coded valentine is not a joke. (Premise)
C1. It is wrong for me to believe that my father is unfaithful to my mother. (From P1, P2, and P3)
C2. My father is unfaithful to my mother. (From P4 and P5).

Here the argument for the two conclusions is valid (whether or not it is sound is another matter entirely); however, this creates another problem. The logic of higher-order attitudes which Blackburn uses to explain the validity of evaluative arguments is supposed to render the two conclusions inconsistent; hence it must also regard the premises as inconsistent. Where C1 would translate to {B! (Believing that my father is unfaithful)}, C2 expresses exactly that belief. Ideally, the quasi-realist wants to convey the idea that having higher-order negative attitudes towards other attitudes that the subject has is inconsistent. Despite this, whilst the two arguments may be sound or unsound, depending on whether the subject genuinely believes in the premise that it is wrong to believe ill of one's friends/family (which on reflection is not unusual for some), the important point is that the argument is valid/consistent and so are its premises:

> If a person who thinks it is wrong to believe something and yet believes it anyway commits an error, it is not an error of logic. Since the explanation used to explain logical inconsistency is equally applicable here, that explanation commits Blackburn to finding logical inconsistency [contradiction] where there is none. (Van Roojen, 1996, 321)

This raises important questions about what constitutes a logical contradiction in the first place. The standard Aristotelian account states that a contradiction is a basic form of error (Translation of Aristotle's *Metaphysics* by Ross, 1928). Language functions to describe observable reality, and because reality cannot be two contrary ways at the same time and in the same respect, to say that it does creates this error. Modern adaptations augment this concept by appealing to possible world semantics. Two sentences contradict if and only if there is no possible world in which both sentences are true e.g. I may be married or unmarried but not both at the same time within that same possible world. According to Van Roojen this does not occur within a logic of attitudes.

> Perhaps both sorts of inconsistency are logical in some good sense, and perhaps 'pragmatic' is a bad term for the contrast class. The point is just that we can draw a contrast and then recognize instances of each sort in both the normative and the nonnormative domains (Van Roojen, 1996, 331-332).

Even the quasi-realist must concede that it is not logically impossible for a moralist to have inconsistent attitudes; hypocrisy is common after all, just as it is possible to have contradictory beliefs; people are prone to inconsistency. To say something like, "I disapprove of eating meat, but I approve of eating chicken", appears to be an absurd, hypocritical thing to say, and according to the quasi-realist, worthy of moral improvement. However, it is at the very least plausible that someone might have such a combination of attitudes. It is **possible** to disapprove of eating meat yet approve of eating chicken, however strange as that may be. It does not carry the same logical force as saying something like, "I am a married bachelor". The latter is not so much hypocritical as plainly **false**, in all possible worlds. It is **not possible** to be a married bachelor[23]. A point of note is that the sentence, "I am a married bachelor", says nothing about attitudes but rather is a non-expressive assertion about some matter of fact. Once again this brings us back to the old Frege-Geach discussion about whether quasi-realist projections of attitudes can indeed contradict. Van Roojen, like Egan, like Geach, seems to maintain that they cannot.

Smyth's response to this is very similar to his reaction to Egan in that, unlike Blackburn, who attempts to wriggle out of the problem by talking about things like stability and attitudes, Smyth simply attempts to pull the rug out from under Van Roojen completely:

> In the case of (VR1) and (VR2), I think that there is a very important difference between considering them abstractly, as sentences spoken by no-one, and considering what our actual responses would be to persons who affirmed these propositions in the course of ordinary moralizing. This is a distinction that makes an enormous difference, not in the least because, *pace* Van Roojen, there cannot be any such thing as a *pretheoretical* distinction between logical consistency and mere pragmatic inconsistency. (Smyth, 2014, 15)

[23] Note that this is **excluding** any and **all** arguments involving time travel; I will leave **that** can of worms to the metaphysicists.

An interesting move here is that Smyth claims that making a distinction between logical and pragmatic inconsistency is not something that Blackburn needs to do in the first place for the quasi-realist account to maintain its basic structure. For example, Smyth asks us to imagine two interlocutors deciding what policies to adopt about stealing. One asks the other about the wrongness of stealing, and they say that it is both wrong and not wrong. This at first sounds exactly like a (VR1) situation, but Smyth then asks the question of what if they had said that stealing is wrong and that they have no negative attitude towards stealing (i.e. Van Roojen's version of VR2). Surely this would be at least as confusing as the (VR1) taken at face value. According to Smyth, the resolute expressivist's response is to point out that Van Roojen presupposes everyday speech is necessarily realist, because to the quasi-realist at least, this is not the case. More importantly, both (VR1) and (VR2) are equally useless for the purposes of real life planning and coordination, so whilst (VR2) is more realistically plausible than (VR1), because it is at least possible to approve of stealing and yet it be wrong, but not for it to be both wrong and not wrong, it is equally unhelpful in the context of conventional, real life moral discourse. Most importantly, if we are operating on the Aristotelian account of language being used to describe observable reality, then (VR2) has clearly made an error in doing so, as they have failed to convey in any meaningful way, or at least any way the listener will understand, the morality of stealing:

> I claim that it is not at all obvious that there are, in this context, categorically different norms governing the use of these two sentences. Each will leave you very confused about your interlocutor's moral stance, each will provoke doubts about their sincerity, and both are equally useless for the purposes of shared planning, co-ordination or policymaking. In other words, each sentence will seem to violate the same conversational norms in basically the same way, and for a pragmatist, this entails that the two sentences cannot fail in *categorically* distinct ways. (Smyth, 2014, 16)

I too would argue that Van Roojen's account presupposes realist moral reference, where for the quasi-realist it does not exist, at least not in the same sense as the moral realist. As we will discover later in chapter five, the main difference between the moral realist and the quasi-realist seems to be that the moral realist makes assertions with the intension of referring to moral truths which are accounted for by **metaphysics**. The quasi-realist, by contrast, **projects** moral truths, which are accounted for by **psychological** states in the speaker and by

a logic of **attitudes**. Smyth appears to defend this account fairly well. However, what I believe is more open to question is whether or not this is an account which Blackburn **shares**. In other words, what precisely is it that makes a 'quasi-realist' rather than merely an 'expressivist'? Is quasi-realism merely a highly sophisticated form of expressivism, or does it claim to be something more? If so, what exactly **is** that in the sense of what exactly do these moral utterances refer to? **Do** they refer to anything? If we claim, as a resolute expressivist might, that they do not, because what we are saying is not fact asserting but rather expressing, [Boo! (ϕ)], then where exactly does this leave Blackburn's notion of moral truth? If, on the other hand, quasi-realist discourse is somehow referential, in the sense that it does somehow refer to something **true** about (ϕ), can we really continue to call this expressive? Surely, if this is the case, then quasi-realist discourse begins to take on aspects which are more descriptive than expressive. So in conclusion, it seems as though either Blackburn must defend quasi-realism from the position of a resolute expressivst, as Smyth does, thereby casting real doubt upon what makes the quasi- **realist**, or, he must argue to preserve an account of moral language which somehow **refers** to quasi- moral truth(s), in which case Van Roojen's arguments may become more of a problem for Blackburn. The problem lies in quasi-realism's aim at being what appear to be two conflicting things, between being a) pragmatic and expressivist and not quasi-**realist**, and b) aiming to be quasi-**realist** but thereby becoming something more than (a), and what that is being somewhat nebulous. The other problem lies in quasi-realism's inability to clearly demonstrate attitude inconsistency through a logic of attitudes. Both of these problems remain serious obstacles for the quasi-realist.

Chapter 5: Quasi Realism no Fictionalism

Having now explored many of the avenues of Blackburn's quasi-realist project, from its origins in *STW*, to its approach to the Frege-Geach problem, to Andy Egan's objections, as well as those of G.F. Schuler and Van Roojen, and a valiant defence by Nicholas Smyth, the time must now come to decide, once and for all, if quasi-realism is a worthy enough candidate upon which to base our non-realist theory of moral language. I will therefore look at what I believe is a particularly important moment in quasi-realism's story, that of David Lewis's paper 'Quasi-Realism is Fictionalism' vs. Blackburn's response, 'Quasi-Realism no Fictionalism'. In this chapter I will examine how David Lewis actually attempts to save quasi-realism by proposing a **fictionalist** solution to what he calls the 'error' of moral realism. I will then examine how, in response, Blackburn successfully severs any potential ties with moral fictionalism, but in doing so, exposes what I believe to be **fatal** weaknesses in quasi-realism. In discussing fictionalism, this chapter will also serve as an important preview of what is to come.

David Lewis begins with an endorsement of quasi-realism, saying that it is not unlikely that it succeeds perfectly on its own terms (Lewis, 2006, 314). This means that it has 'offered a special semantics for sentential expressions of moral attitudes', and has therefore earned the right to say everything that the moral realist says. The challenge is then laid out by Lewis for Blackburn to explain to us what, precisely, makes a **quasi**-realist. In other words, what is it about quasi-realism in particular that makes it distinct from simply being a rather abstract form of realism? One immediate issue which Lewis cites is that it is not enough to simply say that the truth conditional criteria of quasi-realism differ from conventional realist truth conditions. Moreover, 'once the quasi-realist semantics is perfected, it becomes inapplicable' (Lewis, 2006, 314). It is not entirely clear what Lewis means by this, but it seems to be a problem linked to what semantic content fits a particular person's linguistic dispositions rather than their philosophical predilections:

> Presumably the realist semantics is right for the realist. It best fits his linguistic dispositions. But once quasi-realism has been perfected, the quasi-realist's linguistic dispositions are exactly the same, so the realist semantics best fits them, so it's right for the quasi-realist too. (Lewis, 2006, 314)

So what seems to be the problem is that if quasi-realism succeeds in its mission to gain entitlement to everything the realist says, it thereby becomes indistinguishable from what it is supposed to be mirroring. Lewis does, however, have an ingenious solution to this. It is fruitless to search for something which the quasi-realist says which the realist does not, because the quasi-realist's task involves mirroring what the realist says as closely as it can without compromising its own projectivism. However, if there is a distinction between realism and **quasi**-realism, then there should be things which the realist says, which the quasi-realist does not. This alleged distinction is what Lewis calls the 'error of moral realism' (Lewis, 2006, 315). Lewis continues to discuss what this 'error' might be, in terms of what Blackburn says when discussing moral realism, eventually settling on the following, clarifying for the sake of argument, that we are taking 'moral realism' to be a theory unwittingly committed to that error:

> The distinctive error of 'moral realism' says that there are properties, perhaps non-natural properties, such that we can somehow detect them; and such that when we do detect them, that inevitably evokes in us pro- or con-attitudes toward the things that we have detected to have these properties...Projectivism is the view that this is indeed an error; our pro- and con-attitudes actually originate within us as a result of contingent aspects of our psychology and upbringing. (Lewis, 2006, 315)

So it seems that Lewis's interpretation of what Blackburn is saying is that what is different about quasi-realism is the alleged origin of approving and/or disapproving attitudes. Rather than being caused by moral properties in the world, they are caused by psychological phenomena constituted by observations and experience, which are then projected through moral utterance.

Something else which Lewis asks is if the error of moral realism is theory based (an error in the methodological approach of moral realism) or conclusion based (an error with the conclusion that moral adjectives refer to natural properties), this being eventually what leads him to try to 'save' the situation by suggesting moral fictionalism as an interpretation of what Blackburn is saying (Lewis, 2006, 318). Lewis describes the difference between a theory based error and a conclusion based error, and its significance, by alluding to the archaic phlogiston theory. If we 'incorrectly' follow through on a reasonable theory, then we may

still reach an error laden conclusion despite there being nothing inherently error laden about the theory we are applying. If, however, the theory is itself laden with error, then no matter how diligently we apply it, we may still be left with conclusions which are laden with error. In other words, a theory based error is more serious than a conclusion based error. Lewis then applies this question to morality. If we take the error theorist to be correct in their assessment, then morality is victim to a theory based error in the sense that the theory of morality presupposes some account of things which is laden with error. Should we then abandon morality as we have the phlogiston theory? Alternatively, should we attempt to 'correct' morality and fix this supposedly error theoretic account? Lewis's suggestion, however, is that a far easier solution would be to subscribe to a theory of **fictionalist** moral discourse, were we can 'retain morality, but treat it as a fiction.' (Lewis, 2006, 318).

I assume from this that Lewis is claiming that Blackburn's quasi-realism does indeed subscribe to this conclusion of a theory based error of realism and thus is in need of such a move. It becomes less obvious that this is the case in Blackburn's response to Lewis, which I will discuss further down. For now, let us assume, for the sake of argument, that Lewis is correct in this assessment. Based on what Blackburn says of his projectivism, the quasi-realist appears to be doing exactly what the fictionalist does in regards to moral discourse. According to Lewis this involves providing a disowning preface, 'let's make believe that moral realism is true, though it isn't' (Lewis, 2006, 319). Thereafter, the fictionalist/quasi-realist gains entitlement to say anything and everything that the realist says, including discussion of moral discourse, 'but the fictionalist is not asserting what he says, rather he is quasi-asserting it because of his disowning preface' (Lewis, 2006, 319). Lewis goes on to claim that Blackburn's quasi-realism is indeed a brand of fictionalism not only for this reason but also because of what the aim of his project seems to be. Blackburn aims to gain entitlement to whatever the moral realist says, and this, 'means either being, or make-believably being a realist' (Lewis, 2006, 319). So Lewis seems to lay down a dichotomy here. Either the quasi-realist is a kind of realist, in which case he is not so much a quasi-realist so much as a queasy-realist, as Blackburn himself puts it (Blackburn quoted in Lewis, 2006, 319), or, it is not any kind of realism, in which case it is **pretending** to be so for the sake of preserving moral discourse. In this case, it seems indistinguishable from fictionalism, which, as I will talk about in later chapters, is doing precisely that.

Blackburn, of course, does not agree. Somewhat ingeniously, he actually uses Lewis's own argumentative strategy against him. Just as Lewis claims that there are things which moral realism claims which quasi-realism does not, so too does Blackburn claim that there are things fictionalism claims, which quasi-realism does not. Therefore, quasi-realism is distinct from moral fictionalism. To explain why he thinks this, Blackburn asks some important questions, most notably, what exactly does it mean to talk 'as if' something is the case? (Blackburn, 2006, 323). Blackburn initially uses colour fictionalism as an example of what it is that he is asking: "on the face of it, for fictionalism to gain a foothold, we apparently know what it is to talk as if there are colours, although there are none, in which case we need an explanation of what the content of this saying might be" (Blackburn, 2006, 324). In other words, we have an idea of colour even though, according to colour fictionalism, there are really no colours. Fictionalism requires us to understand the fiction of something (or the idea of something), being told as known fact. In other words, we need to understand what it would be for the thing to be a reality in order to understand it as a fiction.

Where quasi-realism differs from fictionalism, according to Blackburn, is that where fictionalism claims to talk about things 'as if' they are real when they are merely fiction, quasi-realism claims to talk about things 'as if' they are real **and that they are**:

> What then is the mistake of describing such a philosophy [quasi-realism] as holding that 'we talk as if there are necessities when really there are none'? It is the failure to notice that the quasi-realist need allow no sense to what follows the 'as if' except one in which it is true, and conversely he need allow no sense than (say) one holding Locke's theory of colour need accept the view that we talk as if there are colours, when there are actually none. This is doubly incorrect, because nothing in the Lockean view forces us to allow any sense to 'there are colours' except one in which it is true; conversely neither need it permit a sense to 'there are actually none' in which that is true. (Blackburn, 2006, 323)

> I went on to say that if the words retain an uncorrupted, English, sense then the Lockean and similarly the quasi-realist, holds not just that we talk and think as if there are...but that there are. (Blackburn, 2006, 323)

Admittedly, it is difficult to describe exactly what Blackburn means by all of this without delving into Lockean metaphysics, but to summarize, Locke describes in his second book of the *Essay*, a kind of second order property that comes from standing in relation to an object, originally an idea proposed by the Greek Atomists (Locke's *Essay* as described by William Uzgalis, 2017). The important point, however, is that there is nothing in that view which commits the quasi-realist to either the view that 'there is colour' or that 'there is no colour' per se. To help Blackburn out here, there appears to be a kind of uniformity in language in which projected qualities act very similar to natural qualities that are referred to. I can talk about something 'as if' it exists when its actual existence is subject to intense philosophical debate, for example when I talk about imaginary numbers like the square root of minus one[24]. We can even formulate whole equations around these imaginary quantities. I can also talk about something 'as if' it exists because, for the sake of argument at least, it definitely does, such as this laptop on which I am typing this. Semantically, there is very little difference between the two, whereas metaphysically there seems to be a great deal of difference. The important point, however, is that to talk about something 'as if' it exists does not commit one to a conclusion one way or the other metaphysically. However, linguistically the truth conditional content of the quasi-realist semantics is another matter entirely, as Blackburn describes in *STW* when he talks about his accounts of moral improvement and moral truth (see section one chapter seven).

Moving onto the important matter, however, that of moral fictionalism vs. moral quasi-realism, Blackburn rightly challenges the fictionalist to answer what the content of a fictional moral claim is when the fictionalist is claiming that 'there are no morals'. In other words, what is it for something to be true in fiction but not to be true in the world? To help Blackburn out here, let us think of an example. We know what it is to live in a world in which there are no unicorns. We also understand what it would be to live in a world in which there are unicorns. We therefore understand what it is to form a fiction around the idea of unicorns. It is much more difficult, however, to understand what it is to live in a world in which there are no moral facts, and for that we need to know what it would be to live in a world in which there are moral facts:

[24] There is a great deal of debate about whether and/or in what sense imaginary numbers like the square root of minus one 'exist'.

> I say that it is bad to neglect children...what would it be for it to be fact and to be known and told as such, that it is bad to neglect the needs of children? It is not so in this world, evidently, so what is different about worlds in which it is? Do children suffer more? But why would that cast doubt on it being bad to neglect ours? (Blackburn, 2006, 325)

To say that it is fiction does not solve this problem, according to Blackburn. This is also the case if, rather than saying semantically 'within the fiction x is true', we instead only **pretend** to make moral assertions. Even in this case, 'we still need to know what it is that we are pretending' (Blackburn, 2006, 326). Blackburn continues to outline some of the other problems with moral fictionalism throughout the rest of his text, and for the sake of time I will not delve into them, for many of the same and similar objections I will deal with in later chapters when I discuss moral fictionalism more directly. The important conclusion we can draw, however, is that it seems that quasi-realism is not exactly the same creature as moral fictionalism. However, in saying that, I believe Blackburn has exposed some **fatal** flaws in the quasi-realist project.

By his own admission, Blackburn's aim is to give everything that the moral realist wants without actually being a moral realist. This raises an important question about exactly what the 'quasi' in quasi-realism actually means. On the one hand, David Lewis appears to want to help Blackburn out by providing a fictionalist solution that would allow him to preserve quasi-realism as an anti-realist account. What seems clear in Blackburn's response, however, is that he has no intention of preserving quasi-realism as **anti**-realist any more than he wants to commit it to being realist. This at first seems to help us in particular, because as we discussed way back in section one, we are not in fact looking for a form of anti-realism so much as a form of non-realism, which quasi-realism seems to be if we choose to believe Blackburn's story. Unfortunately, should we choose to do that then this inevitably drags us towards two, I think, **insurmountable** problems.

The first problem stems from stacking quasi-realism against fictionalism and certain realisations that arise from making that comparison. As I will discuss much later in section three, fictionalism, like quasi-realism, is not necessarily bound to anti-realism either; it can also be depicted as **non-realist**. The difference, however, is that in doing so it retains the ability to 'bite the bullet' on some seemingly 'difficult' questions. It does this, as I will

explain throughout the rest of section two, by appealing to varying levels of something Richard Joyce calls 'critical contexts' (Joyce, 2011, 289). When we are in a philosophy classroom asking a very critical metaethical question about something relatively non-controversial in applied ethics like, "is it **really** wrong to murder children?" this throws a proverbial gauntlet down to the fictionalist and quasi-realist alike. As Blackburn puts it for the quasi-realist, "we talk and think as if there are...[and] that there are." (Blackburn, 2006, 323). In other words, the quasi-realist **commits** to saying, in some quasi-logical sense, that it really **is** wrong to murder children, but that what constitutes that is not the metaphysical state of the real world but rather psychological states within people. This seems almost like saying something akin to, 'it is real because people **think & speak** 'as if' it is it is real'. The problem is that there needs to be some account of how exactly that works which makes sense. This leads us to the second problem, which the first problem necessarily relates to.

Recall back to what I was talking about in chapter three of this section about the difference between attitudes towards doing something and attitudes towards something taking place. For Blackburn to gain entitlement to say what the realist does about things in applied ethics discourse which **are** or **are not** the case, just like the realist, he relies on an account of moral truth, which necessarily **depends** on another account of moral improvement, which necessarily depends on an account of moral fallibility. This tendency towards multi layer **dependency** already creates a dangerous risk factor for the quasi-realist, because should Blackburn's account of moral fallibility fail, for example, so too fails everything else by chain reaction. For his account of moral fallibility to work, Blackburn needs to demonstrate that attitudes can be **inconsistent**. In chapter three, I discussed how, in response to the objections of G.F. Schueler, Blackburn seems to successfully demonstrate a form of inconsistency that arises between a conjunction of a disapproving attitude towards lying and an approving attitude towards not lying, and an approving attitude towards lying. He does this by appealing to an argument in which he demonstrates that there is no possible world in which one can be committed to a world in which lying does and does not take place. As I discussed, however, what is lacking in that argument is a consideration of the varying **forms** of attitudes that exist surrounding, for example, lying. If I were to say that I disapprove of lying **in general**, or that I disapprove of **the existence of lying**, then Blackburn's argument may work; there is no possible ideal world in which lying both does and does not occur that I can be committed to. If, however, I disapprove of **me being a liar**, but I have no qualms about other people lying, then there is absolutely no inconsistency, logical or otherwise.

Similarly, I may disapprove of lying in a particular situation but approve of lying in another. Furthermore, Blackburn's argument only seems to function where there is a commitment to **act** in accordance with a desire (as opposed to a mere 'wish'). There is no inconsistency between wanting to both lie and not to lie, but there is an inconsistency between actually lying and not lying or committing to worlds in which lying both does and does not take place. This does not, however, prove that attitudes can be inconsistent, only that actions and commitments can (which is not particularly illuminating). If attitudes are fundamentally attached to actions or commitments, then there may be grounds for saying that they can be inconsistent. However, Blackburn has yet to demonstrate that this is the case, and indeed there are many examples in which they are not. In chapter three I use the analogy of the monk who has taken an oath of celibacy. The monk's disapproving attitude towards his own sexual behaviour is not inconsistent with the sexual behaviour of other people who have not taken an oath of celibacy. There is also no real inconsistency between a disapproving attitude towards the monk's own sexual behaviour and an approving attitude towards it[25]. An inconsistency only arises where the monk intends to **act** in accordance with both attitudes (this is of course impossible), in which case I argue that it is the **actions** which are inconsistent, not the attitudes. If Blackburn cannot demonstrate significant inconsistencies between attitudes, then he cannot demonstrate moral fallibility in the sense that he describes it. If he cannot demonstrate moral fallibility, then he cannot demonstrate moral improvement, and if he cannot demonstrate moral improvement, then he cannot demonstrate moral truth, and he cannot gain entitlement to say, as the realist does, that it **is** wrong to murder children, only that he disapproves of it. This is a serious problem...for the quasi-realist.

This is not, as I will reveal later on, a serious problem for the moral fictionalist. By contrast, the fictionalist is not looking to pander to realists, but merely to account for why realist **sounding** discourse is uttered. This is much simpler and easier to accomplish. Referring back to our first problem of answering that most difficult of questions about whether or not it is really wrong to murder children, a non-realist fictionalist can bite the bullet in the way that a quasi-realist cannot, by simply and unapologetically answering that, 'there is no reason to think that it is **really** wrong to murder children when we examine that claim in the most critical context possible'. Whilst this sounds nihilistic, and potentially worrying, it is important to remember that the fictionalist may also answer the equal and opposite response

[25] The monk may be going through a moral crisis where he is torn between his sexual desire and his oath of celibacy.

as well, 'there is no reason to think that it is **not** really wrong to murder children when we examine that claim in the most critical context possible'. A non-realist fictionalist is no more bound to error theory than the quasi-realist is. It is only if the fictionalist is an **anti**-realist that the second claim becomes inapplicable. In neither version, however, is the fictionalist vulnerable to the same pitfalls of multi layer dependency or the challenge of proving quasi-logical inconsistencies between attitudes. Within the realm of moral fictionalism, there is no inconsistency between disapproval of lying and approval of getting one's brother to lie, nor does there need to be for fictionalism to function effectively, nor is fictionalism vulnerable in the same way as quasi-realism to the Frege-Geach problem, because it is not a form of semantic expressivism. Either assertions are made in the form of 'within the fiction, x is true', or assertions are pretended, in which case there is no need to form arguments in asserted contexts so much as pretence arguments based on pretence assertions (all this and more I will discuss this later on in this section).

By rejecting fictionalism, Blackburn effectively condemns quasi-realism to all of the problems that fictionalism escapes from with relative ease. The main issues, as I see them, seem to be that Blackburn does not demonstrate that attitudes can be **practically** inconsistent in the relatively strong way that he needs to effectively demonstrate his account of moral fallibility, and so by chain reaction, this negatively affects moral improvement and moral truth also, and this in turns casts real doubt on whether Blackburn can indeed gain entitlement to all that he wants to say in appeasement to the moral realist. In addition, quasi-realism, when stacked against moral fictionalism, suffers from an inability to bite the bullet on difficult questions. Referring back to that dilemma posed by Lewis about whether or not Blackburn peddles quasi- versus queasy- realism, by saying that we speak as if it is wrong and that it is, Blackburn seems to push away from the 'quasi' and towards the 'realism', at the expense of his expressivism. Expressivism, at its heart, is the theory of moral language based on expressions of approval and disapproval, but Blackburn's quasi-realism seems to be pushing for something more, something which, ironically enough, seems more and more nebulous the more Blackburn squirms and wriggles towards it.

Chapter 6: Fictionalism

Over much of the course of the previous two sections, I have explained my reasons for pursuing a non-realist account of moral discourse after acknowledging grounds for doubting moral realism. I then turned my attention to Simon Blackburn and quasi-realism, which appeared to fulfil many of the criteria I was looking for, something which can give any realist in the room everything that they want to hear, which accounts for how and why non-realist language is used in a realist sounding manner, and without the expense of additional inquiry into any moral ontology. The answer is found not in the metaphysics of moral language but in the language itself and in the psychology of the language user. Unfortunately, I have also explained how there are unexpected and undesirable expenses attached to this approach of quasi-realism. As an expressivist account, it renders itself vulnerable to the Frege-Geach problem, something which Blackburn has attempted to deal with, the success of which is debatable (see section one chapters six and seven). However, I think that the larger problem for quasi-realism lies in its reliance on multi layer **dependency** between elements like moral truth, moral improvement and moral fallibility, and in how one problem may cause problems throughout the project by chain reaction. Specifically, quasi-realism struggles to satisfy the need to demonstrate how attitudes alone can be **inconsistent** in such a way that it can demonstrate logical, non-realist moral fallibility, and so by chain reaction, this casts doubt on quasi-realist claims which necessarily depend upon on this idea, namely moral improvement and moral truth, without which, the quasi-realist cannot achieve their aim of gaining entitlement to appease the moral realist in the room. A third problem lies in Blackburn's inability to bite the bullet on difficult questions, which rather than being confronted head on and diffusing, as an anti-realist may be inclined to do, push the quasi-realist away from the 'quasi' towards the 'realism' at the expense of their expressivism, a move dramatically different to something like the 'resolute expressivism' of Nicholas Smyth. This means that rather than maintaining their position as a non-realist peddling a theory of moral language based on expressions of approval and disapproval, quasi-realism tries to become something more, and what precisely that may be is frustratingly difficult to put a definitive finger on.

This brings us to another theory of moral language: moral fictionalism. Before I continue throughout the remainder of the thesis, however, I would do well to explain myself a little here. After all, the reader may be wondering at this point: if quasi-realism is not all that it is

cracked up to be, then why have I subjected you to half a thesis talking about it? The answer to that question is very simple indeed. As I will explain throughout the rest of section two, moral fictionalism is not all that it is cracked up to be either. **However**, I believe that there are **key insights in both accounts which are <u>not</u> mutually exclusive**. In other words, both are required for what I will be doing later on in section three as you will discover. In essence, and as I will explain again at the end of this section, what I will be doing will involve taking some of the **semantic** content of moral fictionalism, reworking the expressivist components of quasi-realism to take on a more **pragmatic** role and splicing the two together with modifications of my own. The final result should be an account which is non-realist, as opposed to anti-realist, which accounts for how and why realist sounding moral language is used by non-realists, which is not dependent on error theory, which can and does bite the bullet on difficult moral questions without wriggling away from them, which survives the Frege-Geach problem and which does not depend on any moral ontology to explain any of this. I do not believe that either quasi-realism or moral fictionalism achieves all of this independently, but together, I believe that all of this and more is achievable. Without further ado then, let us delve into moral fictionalism.

Let us first talk about fictionalism more broadly. In order to get a sense of why fictionalism is so interesting, let us first consider the strangeness of a particular phenomenon. When talking about things that are clearly fictional, we nevertheless seem to talk and feel about them as if they are real. For example, when talking to a Marvel comics fan, it would not seem out of the ordinary to say something like, "we don't actually know Spiderman's identity do we?" and for the Marvel fan to respond, "of course we do! Spiderman's real name is Peter Parker!" No doubt that Marvel fan might actually be annoyed that someone has not only not read the comic but has completely failed to pick up on one of the key plot points. We see similar levels of annoyance about many subjects of fiction from continuity errors to arguments about who does and does not deserve to be tortured or killed in the next episode of Game of Thrones. What is interesting is that we seem to talk about fictional characters as if they are people that we know in real life to the extent that we feel elated when good things happen to them and heartbroken when bad things happen to them. What is strange is that we know that they are not real, so why do we talk about them as if they are, and why do we get so worked up about people who fundamentally do not really exist? Even more bizarre is that actual sentences like "Spiderman's real name is Peter Parker", **appear** to be true, to the extent that we are more often than not inclined to 'correct' people who say otherwise; this is despite us

knowing that the sentence surely cannot be true in the same way as normal assertions are, as there is no referent to whom the sentence refers. What on earth is going on here?

At its most basic level, fictional**ism** is this practice of likening types of language (e.g. language about colours, mathematics or even ethics), to talking about things which do not exist outside the context of **fiction** (e.g. language about Spiderman or characters in Game of Thrones). This of course raises a very important question about the discourse of fictions. What exactly is going on semantically when a speaker utters a sentence about some fictitious entity, which for all intents and purposes **sounds** like an assertion about something real? [26] There are at least two rival accounts that we should pay special attention to. One explanation is that the speaker of a sentence like, "Spiderman's real name is Peter Parker", **is** making an assertion and that that assertion, out of context, is ultimately false. Alternatively, some people think that when we speak about something fictional, we are tacitly using some kind of fiction operator, as if to say "according to the Spiderman fiction, 'Spiderman's real name is Peter Parker'". The operator ensures that the sentence is true, because 'according to the fiction', it is, and without this operator it is not. This is not to be confused with lying, as there is no intentional deception on behalf of the speaker. To the contrary, the speaker relies on the listener knowing that what they are saying is not true in the absence of this fiction operator. Otherwise, we may get a conversation like this:

> Spiderman's real name is Peter Parker / Who is Spiderman? / He's a person who lives in New York who dresses like a Spider, shoots webs from his hands and fights criminals. / Wow! When can I meet this guy?! / He's not real you realise. / Oh! Then why did you talk about him like he is then?

It is indeed part of the strangeness of fiction that such conversations do not happen more often. Indeed, while this explanation seems sufficient to explain semantically what is going on, it does little to explain the strangeness of people's reactions. Yet another explanation is that rather than making false or tacit assertions, sentences like "Spiderman's real name is Peter Parker" do not actually assert anything at all but rather are examples of **pretending** to assert something. All explanations, especially the latter explanation, are highly context

[26] This more than anything else does seem to be a central question particularly in Joyce's main body of work *The Myth of Morality*. That being said, many of the other questions in that work I will address by references to other works of Richard Joyce, and as such I have not felt any pressing need to include undue references to *The Myth of Morality* on top of those.

dependent. If we say the sentence out of context, then we are likely to engage in a conversation like the one above. The main reason that we do not usually have these sorts of conversations is that we usually only utter such sentences in very specific contexts. For example,

> Who is your favourite super hero? / Spiderman, because he's someone I can identify with, having to hide his identity to live a normal life etc. Spiderman's real name is Peter Parker. Peter Parker has loved ones who can be harmed…

The idea here is that the sentence is prefixed by a prior established context, as if to say, "Let's pretend that for the duration of this conversation, there exists a guy called Peter Parker who is also Spiderman".

All of this and more is discussed by Richard Joyce in his paper 'Moral Fictionalism' as one of several accounts of fictionalism, and one that is of particular interest and relevance to any discussion on comparison with quasi-realism. What then is **moral** fictionalism? If we are to be consistent with our above definition of fictionalism, then moral fictionalism is the practice of likening moral discourse to discourse about fictional entities (e.g. Spiderman). Joyce begins by citing John Mackie's *Ethics: Inventing Right and Wrong*, in which Mackie suggests that moral discourse "can continue with the status of a 'useful fiction'." (Joyce, 2011, 287). Joyce then asks how useful a system of morality can be when we cease to believe it is real, proceeding on the assumption that "Mackie's arguments for moral error theory are cogent", (Joyce, 2011, 287), which amounts to assuming firstly that moral sentences express belief states and second that moral assertions are untrue. Mackie's response to error theory is to 'carry on with morality as a fiction', and Joyce seeks to understand precisely what this answer means. According to Joyce, moral fictionalism is not so much a description of all moral discourse so much as a feasible description of the discourse of moral sceptics (i.e. those who do not believe in the existence of real moral properties). In other words, Joyce claims to be peddling a 'revolutionary' rather than a 'hermeneutic' form of moral fictionalism (as found in Burgess, 1983). This means that rather than trying to account for simply how mainstream speakers usually talk about morality, Joyce is instead describing how non-realist speakers may talk about morality.

So why might someone be attracted to fictionalism in first place? In Joyce's section on what he calls 'critical contexts', Joyce outlines "how it might be that a person might carry on using a discourse that she has come to see as flawed" (Joyce, 2011, 289). To this end, Joyce makes an analogy with colour fictionalism. Suppose that David, for whatever philosophical or scientific reasons, has come to the conclusion that despite the appearance of objects to have the surface properties of red, blue or yellow, such properties do not **really** exist. Rather, they are an invention of his consciousness, a means by which his brain has evolved to detect the physical surface properties of these objects, their chemical composition and/or their spatial location in relation to a given light source. Believing that these 'colours' do not really exist, does David go about the rest of his life correcting everyone who refers to the grass as green? Alternatively, does David, 99% of the time, when referring to grass **pretend** and refer to it as green? The latter is more likely according to Joyce (Joyce, 2011, 290). In fact, it is only in such places as the philosophy classroom, and only when properly pressed on the issue that David admits what he really thinks about the so called 'colour' of grass, or the so called 'colour' of his mother's eyes. Only then does David begrudgingly say, "My mother's eyes are not really green; nothing is **really** green". Joyce claims this to be an example of varying levels of critical contexts, contexts of critical analysis:

> ...it would be too bizarre to hold that an individual, who has never given the issue any careful thought whatsoever, but thinks and acts in accordance with theory *T*, does not really believe *T* simply because if he *were* to think carefully about it, he would deny it. But if we add that at some point he has adopted a critical perspective and therein sincerely denied *T*, and remains disposed to deny *T*, were he again to adopt that perspective, then he disbelieves *T*, regardless of how he may think, act, and speak in less critical perspectives. (Joyce, 2011, 290)

This brings us back to the issue of those competing accounts of fictionalism. First, let us look at the tacit operator version, which says that the sentence 'Spiderman's real name is Peter Parker' really means, 'according to the Spiderman fiction, Spiderman's real name is Peter Parker'. The alternative is a 'pretence assertion' account, according to which when I utter the sentence about Spiderman's real name I merely pretend to assert and the listeners join in the pretence, as if to say, 'for the duration of the conversation, let us pretend that what I am saying is a true assertion'. Joyce argues against the first account. In David's case, his most critical context is when he reflects on the matter of colour's existence critically and

philosophically, but it is not what occupies his thoughts most of the time. For the majority of his time he still has this sceptical belief that colour does not really exist, but he is not attending to it. However, David remains disposed to deny *T* (T in this case being the theory that colours exists) when placed in his most undistracted, reflective and critical context. There is a difference then between describing the story as opposed to telling it. David always believes that the world is not coloured, even when he is pretending not to, which is most of the time (Joyce, 2011, 291). Similarly, when I say the sentence, "There once was a goblin that liked jam", what I am not doing is saying, "According to Hans Christian Andersen's story, there once was a goblin who liked jam." According to Joyce, this is inadequate as a general claim (Joyce, 2011, 291). This is because if every sentence of the story contained an unpronounced fiction operator, then the above sentence ceases to make sense. According to Hans Christian Andersen's story, there once was a goblin who liked jam; this is not pretence but fact. Joyce prefers to say that the speaker is not asserting anything at all but rather is pretending to assert this (Joyce, 2011, 291). Even when pressed on whether he is asserting that the grass is green, David will likely respond with an affirmative answer. However, Joyce argues that this affirmative answer is itself part of his semantic pretence (Joyce, 2011, 292). This is largely due to certain problems with the alternative explanation that David is making an assertion and that assertion is deliberately false. For example, consider the following:

P1 Fresh grass is green.
P2 My lawn is made of fresh grass.
C Therefore, my lawn is green.

If P1 is elliptical (according to the fiction of a coloured world, fresh grass is green), then the argument is not valid. Even if we suppose that all three claims are so prefixed, the problem then arises that we suppose that, "according to the fiction of a coloured world, my lawn is made of fresh grass". Joyce argues that this is clearly false (Joyce, 2011, 292).

It might seem obvious then that the concise explanation of Joyce's argument against a tacit operator is that it is much simpler to assume a pretence assertion. For example, suppose that David says, "Fresh grass is a type of vegetation", in order to get the conclusion that "my lawn is made up of a type of vegetation". In order for this argument to be valid we need to interpret this new premise and this new conclusion as also bearing the prefix of 'according to the fiction'. In fact, any assertion that David makes may utilise colour claims as a premise of an

apparently valid argument. Things of course get even more complicated if we suppose that colour is not the only fiction that David partakes in:

> Eighteenth-century philosophy may also lead him to endorse an error theory for sound and smell, for causation, for virtue and vice, and thus in order for all his apparently unremarkable, apparently valid argumentative moves to be genuinely valid, we will have to interpret every claim issuing from his mouth as brimming with unspoken prefixes…All such unpleasantness is avoided if we do away with tacit operators, and simply interpret David's utterance 'Fresh grass is green' as a kind of make-believe assertion. (Joyce, 2011, 293)

If we therefore suppose that fictional claims are pretence assertions rather than tacit operators, then the semantics become much simpler. The content of the proposition 'There once was a goblin who liked jam' does not change whether it is said as part of a story or when it is said as part of an attempt to genuinely assert something false. What changes is the force with which the content is uttered. However, this presents us with another dilemma; what are we to make of arguments that feature both fictional and real content? For example, consider the following:

P1 It is cold tonight.
P2 It is the height of summer.
P3 A cold night in the height of summer is unusual weather.
C Tonight is unusual weather.
(Joyce, 2011, 294)

Suppose I present this as an example of validity. Because it is merely an example, I am not at that moment asserting either the premises or the conclusion of this argument, and yet it remains valid. Suppose, however, that I utter P1 during my performance in a play given on a hot summer's night. When pressed on the current climactic circumstances I genuinely assert that P2 and P3. Clearly, I have not committed myself to C, which I believe is false, because I also believe that P1 is false, despite pretending that I do not. However, there is nothing stopping me from combining P2 and P3 with the fictional P1 and endorsing C as part of the fictional act. What this does is force me to accept the following: "If it **were** cold tonight, that

would be unusual weather", but as far as what I am genuinely asserting is concerned, it is precisely because I have **not** asserted P1 that I do not have to accept C.

So let us recap what we have learned about fictionalism thus far. Fictionalism is a way of likening discourse about something, whether it is colour, mathematics or ethics, to discourse about fictional objects such as Spiderman. In other words, it is a way of explaining the semantics of things which do not really exist, or at the very least things we have **reason** to doubt the existence of. We have also learned about two opposing interpretations of fictionalism in particular, those being the tacit operator account and the pretence assertion account. In Joyce's case, we have reason to believe that in the case of moral fictionalism at least, the pretence assertion account is better than the tacit operator account, as the former avoids many logical problems that the latter is particularly vulnerable to. I argue that this is indeed the case. I agree with Joyce on this point, and in my next chapter, I will discuss Joycean moral fictionalism in greater detail and provide some additional insight into why the pretence assertion account works particularly well for the moral fictionalist, as well as Joyce's account of **why** moral fictionalists engage in this kind of pretence.

Chapter 7: Moral Fictionalism

Now let us talk more specifically about **moral** fictionalism. In its most basic sense, moral fictionalism refers to a brand of fictionalism aimed at describing moral discourse (i.e. how people talk about 'moral things'). According to our preferred brand of fictionalism (pretence assertion rather than the tacit operator), the sentence, 'ω is wrong', holds exactly the semantic structure that it appears to. The speaker is pretending that they are making an assertion that ω has a property of wrongness (according to some prior established context). It is important that within a given context, the sentence **appears** true, even though really it is not. Once again, let us compare this to literary fiction. When we talk about Spiderman's real name being Peter Parker, we are essentially saying something that is not **really** true. There is no man called Peter Parker who shoots webs from his hands and swings around New York City fighting crime dressed as a spider. Nevertheless, I am not lying when I say the sentence; there is no deception involved.

Why, you may ask, would anyone compare moral discourse to fiction discourse in the first place? In Joyce's 'Moral Fictionalism', the key words that we ought to remember are 'motivation' and 'emotion'. Joyce is quick to admit that the emotions that arise from fiction do not necessarily affect human behaviour in the same way that emotions that arise from genuinely held beliefs do (Joyce, 2011, 302). That is not to say that they are necessarily less profound or less motivational, but they are certainly different. For example, a tragic scene in a film portraying the death of a leading character that you have spent the last two hours getting to know may well affect you in a far more powerful way than a news report about a person you've never met dying in exactly the same way in real life. On the other hand, as Joyce points out, the fear of fictional vampires is consistent with someone sitting and eating popcorn, whereas the fear of actual vampires is more consistent with them buying wooden stakes and garlic (Joyce, 2011, 302). What are we to make of this? What is interesting about the fictional examples is that despite not really being believed in, they are nonetheless provoking emotion and appear motivational on some level. For example, suppose I watch an episode of Star Trek: The Next Generation in which Patrick Stewart's character is captured by aliens and tortured for information. Whilst the setting and the characters are entirely fictional (and scientifically implausible), because we have spent several seasons getting to know Patrick Stewart's character Jean Luc Picard, the audience's attitude towards what is happening to him is real. Somewhere within the episode, Picard utters the line, "torture has

never been an effective means of obtaining information; one wonders why it is still practiced in the 24th century" (*Star Trek the Next Generation*, series six episode eleven ("Chain of Command")). Again, this has elements of fiction, as torture may or may not still exist in the 24th century, we cannot know for sure, but there are also strong elements of truth within this quote. Torture is indeed notoriously unreliable (often the victim will say anything just to make the pain stop, whether they actually know anything useful or not), and yet it is still practiced in the 21st century, and this is exceedingly controversial. It is unlikely that a majority of the audience watching this Star Trek episode have had any firsthand experience with torture, nor do they likely know anyone who has had firsthand experience of it. Nevertheless, they are easily able to comprehend the horrifying effects of torture and to deem it undesirable simply by watching someone **pretend** that they are being tortured. The pretence alone may even be enough to inspire action against it. On a side note, it seems that the most powerful and motivational fictions tend to be those that make references to true occurrences. For example, sticking with our Star Trek theme, one of the most famous episodes of Deep Space Nine depicts Avery Brook's character Benjamin Sisko, who is a successful commander of a space station in the 24th century, dreaming about living life as a black story writer in 1950s America. Again, whilst the character is fictional, the setting and the events that transpire are uncomfortably realistic (*Star Trek: Deep Space Nine*, series six episode 13 ("Far Beyond the Stars")).

So how does this apply to what we think we know about actual moral codes and moral systems? To that Joyce offers us the following:

> Suppose I am determined to exercise regularly, after a lifetime of lethargy, but find myself succumbing to temptation. An effective strategy will be for me to lay down a strong and authoritative rule: *I must do fifty sit-ups every day, no less.* (Joyce, 2011, 303).

Joyce continues to remind us that it is not strictly speaking a true thing to say that if he fails to perform these fifty sit-ups every day then he will indeed slip back into lethargy. Indeed, maybe forty or so every other day would be perfectly sufficient. However, to pay much attention to the flexibility of this 'rule' would be to dangerously bend and perhaps even break any progress that he might make. The attitude is more successful when viewed as inflexible. He may even convince himself that bad things will happen if he fails to perform fifty every

day, despite knowing in a more critical context that really bad things merely **could** happen. Within this critical context, however, does he really need to believe? Not at all, he need only convince himself to the extent that progress is made. This does indeed sound familiar.

Motivation is often characterised as deeply personal. For example, Hume alludes to a drunkard who is repeatedly told about the statistics of alcohol related deaths, which has little effect on him; until he sees his own friend, who is also a drunkard, die from alcohol poisoning. It is not that the drunkard failed to believe the statistics prior to his friend's death, but the "tangibility of the one death has, in Hume's words, 'a superior influence on the judgement, as well as on the passions' (Hume, 1739/1978: 143-4.)" (Joyce, 2011, 304) According to Joyce, following the death of his friend, the drunkard's perception of alcoholism may shift from one of 'if I carry on, this terrible thing **might** happen to me', to, 'if I carry on, this terrible thing **will** happen to me'. The latter is far more dubious. There are many cases in which people carry on drinking or smoking heavily into old age and die from something completely unrelated. However, it is true that the odds are not in the drunkard's favour should they continue drinking. In order to motivate themselves out of alcoholism, it is surely more effective to think in terms of how this terrible thing **will** happen, as opposed to the more realistic **might.**

Now how do we apply this to **moral** discourse? Going back to a conversation in a previous chapter, we can look again at kicking cats. The fictionalist may claim that it is not **really** wrong to kick cats any more than it is really right to kick them. However, is it at all useful, practical or ultimately desirable to kick cats? To some of our more sadistic fellow human beings, perhaps it is, but to most ordinary people it is not in the slightest. Typically, one does not need to invent a rule to persuade oneself to refrain from kicking cats; such behaviour would not normally occur to the individual in the first place. However, there are far more generalised social rules that the individual may care to invent, e.g. "People must not cause unnecessary suffering to other living things", which kicking cats would clearly fall under the category of, thereby provoking no small degree of outrage from beholding one who blatantly flouts this rule. The fictionalist may well believe that the culprit does not **really** have anything to lose by kicking the cat. If they can get away with kicking the cat, and kicking the cat brings them sadistic pleasure, why **not** kick the cat? However, an observer may not be comfortable living in a world in which cats are kicked and will do everything possible to dissuade people from that kind of behaviour, and the most effective way of motivating others

to refrain from kicking cats is to give people a sense of 'wrongness' (that some behaviours are intrinsically unacceptable). Another, perhaps more practical approach is to create some concept of cosmic justice, whereby if you do kick cats, the universe will find a way to make you suffer as a result. To the non-religious, this is not true in the slightest; the universe cares little for the plight of cats. However, we do know that if you can but convince enough people, then cats cease to be kicked. [27]

How can we relate this to actual moral discourse? According to a moral fictionalist, what does a person mean when they say that, "ω is wrong!"? Based on what we have learned so far it seems that the answer is that saying, "ω is wrong!", in a pre-established context, is like saying that "Spiderman's real name is Peter Parker" without a tacit operator. In other words, what we are doing is **pretending** to assert a true fact, when in reality we are not, in order to convey some other, more pragmatic information to the listener. This ensures that the semantic structure of the sentence remains consistent (i.e. "ω is wrong!" is treated as truth apt; it is **taken** as true; it is **really** untrue[28]. The pre-established context of such an utterance also ensures that it is not taken as true in the same way as a non-fictional utterance (i.e. the listener pretends that it is literally true), but when either of us are pressed on the subject in our most critical context, we admit that it is really untrue. What pragmatic information is being conveyed to the listener surely depends on the context of the utterance. If, for example, I say to someone kicking a cat, "kicking cats is wrong!" then it would seem a Joycean fictionalist would argue that I am pretending to assert this and that I know that the listener knows that I do not really believe it, but that I still do not want them to kick the cat. When I say, for example, that "cats ought not to be kicked", this then **seems** to be making some kind of universal claim. In reality, it says a lot more about the kind of world that I want to live in rather than the reality of the world I actually live in.

To conclude this chapter, I would agree with most of what Joyce is saying here. However, I believe that this is only the tip of the iceberg. Much of the first half of this thesis now becomes very important, because when Joyce talks about motivations and emotions, I believe that there are some striking parallels with expressivism. However, unlike expressivism, which deals with emotions and attitudes like approval and disapproval, and the communication of

[27] These two approaches are very similar to the two models of translating emotions into moral language I describe much later in section three, and which form the backbone of my own expressive fictionalism later, these being the empathy model, and the pride, shame and responsibility model (see section 3 chapter 5).
[28] 'Untrue', however, should not be confused with 'false'. 'False' is still truth apt.

those attitudes through projectivism, quasi-realism and the semantics of moral utterances, this kind of thinking by Joyce seems more targeted at the pragmatics of moral discourse. This is particularly interesting, and something which I will come back to at the end of this section, and something which will be a key theme throughout my third and final section of the thesis. What is also interesting, however, is that these motivations and discussions of the psychology of moral discourse stand related to but ultimately separate from the semantics of moral discourse involving pretence assertion and critical contexts. To my knowledge, Joyce does not peddle himself as a pragmatic expressivist, and frustratingly what I have said here alludes to most of what Joyce has to say on the subject of **why** moral fictionalists talk about morality in the form of pretence. It is for that reason that I believe that this kind of fictionalism could be developed a great deal further, as it has some very interesting potential, and I will be doing exactly that in section three. However, first let us see what could be wrong with this kind of fictionalism and see if this idea has any real merit by examining the case **against** fictionalism, which I will discuss in my next chapter.

Chapter 8: The Case against Fictionalism

Let us now look at what might be wrong with fictionalism and speaking in fictional terms about something. Moral fictionalism claims that we can use and make sense of moral language even in the absence of any actual moral facts, in exactly the same way that we can use and make sense of language involving references to Spiderman in the absence of any actual Spiderman. This claim is both strong and fairly justifiable according to the previous two chapters, and it is a worthy candidate on which to base a non-realist account of moral language, **at least on a semantic level**. However, this does not come without a price. First of all, in rebelling against expressivism and quasi-realism, we abandon many of those pragmatic aspects of both that were so attractive to begin with. Quasi-realism, and expressivism more broadly, in comparison to moral fictionalism, has a comparatively detailed story to tell us about not only how we talk about ethics but more importantly **why**, the idea being that we use moral language to communicate attitudes of approval and disapproval, but as we have seen, trying to explain this through the semantics of moral discourse comes with certain objections and complications. Moral fictionalism on the other hand has just as interesting a story to tell us about **how** we can use moral language in a realist seeming, but ultimately non-realist manner, but does not have as detailed an account of the pragmatic reasons **why** we talk about ethics fictitiously. As discussed in chapter seven, we are given at least a strong attempt at this by Richard Joyce, but as I will address throughout section three, I am convinced that there is a great deal more to this than even Joyce supposes and that attitudes of approval and disapproval like those proposed by quasi-realists, and our capacity to communicate them, must play a central role as well. The second issue that must be resolved before we go any further is that just like quasi-realism; fictionalism is not free from controversy either. In the next two chapters, I will address some of the concerns addressing fictionalism, and the use of fiction more broadly, and then focus on one set of objections in particular in chapter nine by Jonas Olson.

Attempting to undermine the fictionalist account on one side is at least straight forward, and there is a significant amount of material to use against hermeneutic fictionalism (using fictionalism to describe mainstream moral discourse). For example, among many objections to hermeneutic fictionalism made in 2001 by Jason Stanley, Stanley argues that while some autistic people struggle to comprehend concepts of fiction or were unable to distinguish between fiction and reality (a similar well known pop culture example would be the

Thermian species from the film *Galaxy Quest*, who lack a concept of 'pretence'), they nevertheless did not struggle to comprehend what some hermeneutic brands of fictionalism portray as being fictional e.g. colour, mathematics, ethics etc. These cases offer some empirical evidence against hermeneutic fictionalism (Stanley, 2001, 36-71). Like Joyce, however, the form I am using is more revolutionary than hermeneutic in the sense that I am not asserting that the normal, standard form of moral discourse necessarily utilises fictionalist semantics. To the contrary, I believe that a vast majority of moral conversations are decidedly realist in the sense that moral assertions are made on the basis that they are at least intended to refer to actual moral ontology, whether such ontology is accurate or not, but I would argue that a majority of moral speakers either believe that it is or are morally agnostic on that point. However, not all moral speakers are realists, and this is what I seek to address. Semantically, what is going on in the case of the non-realist who does not want to make any commitments whatsoever to any moral ontology? How then does the fictionalist account for their language, which certainly appears referential, even when they ultimately do not **believe** there is anything really being referred to? So it is this kind of revolutionary form of fictionalism that we must attempt, and hopefully fail to undermine if we are to strengthen it.

The most obvious way of doing this is to attempt to counter the idea that we can, quite reasonably, make truth apt assertions based on referring to things that do not exist, in which case this would also apply to our Spiderman example and literary fictionalism, and perhaps all fictional discourse. First, let us remind ourselves of what we are attempting to argue against. Under a literary fictionalist analysis, for example, the sentence, "Spiderman's real name is Peter Parker", should be taken as a true statement, because in context, it is understood as being such even though both Spiderman and Peter Parker constitute identities of a fictional character. One advantage in making this claim is that it is metaphysically economical. We do not necessarily need to account for Spiderman/Peter Parker's ontological nature/existence **if** we accept that he does not **need** to exist in order to be referred to in a meaningful sentence, if we accept that ontological existence is not a **necessary** condition of referential language, or the truth aptitude of a given sentence. Rather, what is necessary is a mutually understood **context** (e.g. in this case the canon Marvel universe) in which, for the duration of a given conversation, the following references, in this case references to Spiderman, will be treated as truth apt. This is not unlike a Gricean concept of a cooperative principle (Grice, 1975, 41-58). Outside of this or any other context the sentence loses its meaning. For example, if I were to randomly say something like, "Jerry is a 'Sparpleplat'

[gibberish utterance]", then we have no means of recognising whether or not this is a true sentence, as we have not assigned any context in which this can be treated as true or false; we have not assigned who this Jerry character is or what on earth a 'Sparpleplat' is supposed to be. It is, therefore, **meaningless.** In other words, truth apt fictionalist sentences are supposedly made true or false in virtue of their **contextual dependency**. However, all of this is not without some serious philosophical problems.

The point in saying all of this is specifically to avoid serious philosophical problems regarding the ontology of Spiderman. However, fiction of this kind can, some argue, fail to deliver on this. Therefore, the supposed advantages in, for example, mathematical, modal or moral fictionalism being that these entities characteristic of the relevant discourse have the same ontological properties of fictional entities, can be problematic specifically because the ontology of fictional entities is not without controversy. Specifically, I refer to something called the problem of fictional discourse. Consider the following sentence:

(1) Pegasus is a flying horse.

As with many other sentences used in fictional discourse, this sentence seems to fulfil three specific conditions:

1. It has a predicative grammatical structure, i.e. a structure that is rendered in logical notation ("Fb"---where "F" is the predicate expression "is a flying horse", and "b" is the singular term "Pegasus"---in other words "Pegasus is a flying horse")
2. The singular term ("b"), is a name for a fictitious object
3. The sentence is normally taken to be a true sentence in the context in which it is normally uttered.

The problem of fictional discourse arises from two closely connected logical principles. The first is the principle of existential generalisation:

Existential Generalisation

$Fb \rightarrow \exists x(Fx)$, i.e.,

If *b* is *F*, then there is something that is *F*

The second is the often tacitly assumed predication principle:

Predication Principle

$Fb \rightarrow \exists x(x = b)$.

(PP) may be read in two ways:

(Ppa) If *b* is *F*, then there is something that is identical with *b*.

(PPb) If *b* is *F*, then *b* exists.

Both of these principles are prima facie plausible in that if it is true of an individual that the predicate "F" applies to it, then the predicate "F" applies to something. However, this does not seem to be the case when we apply "F" to fictional objects like Spiderman, and if we attempt to apply both of these principles to a fictional discourse like "Spiderman's real name is Peter Parker", they lead to further issues that seem to contradict empirical facts on the one hand and the ontological status of fictional objects on the other. According to the principle of existential generalisation, the following sentence:

(1) Pegasus is a flying horse

implies

(2) There are flying horses

...and yet, as far as we know empirically at least, this is false. Meanwhile, according to the principle of predication:

(1) Pegasus is a flying horse

implies

C) Pegasus exists

...and yet Pegasus is a fictional object, and this seems to call into question whether or not we simply call an object 'fictitious' (i.e. to assign it a quality of "fictitiousness") as a kind of synonym for saying that it does not exist. So to summarize, if (1) is true and (2) is false, then (3) cannot be true for fear of logical contradiction with (2); therefore, we cannot logically apply **both** the principle of existential generalisation **and** the principle of predication to fictional discourse (as seen in Reicher, 2014, &, Eklund, 2015).

In other words, seemingly true fictional sentences seem to lead to logical contradictions if we insist on binding fictional discourse to these two principles. However, we may alternatively reject the premise that either the principle of existential generalisation or the principle of predication are universally applicable, proposing instead that some sentences, notably ones about fictional objects, do not **necessarily** operate under this exact kind of logical notation. For example, logicians such as proponents of Free Logics take exactly this position, such as Hintikka in 1959, Lambert 1983 and in 1991 and also Leonard in 1956. Similarly, while Simon Blackburn, who we have discussed extensively in previous sections, makes no claims towards fictionalism, his quasi-realist semantic form of expressivism and his logic of attitudes does resemble an attempt to steer moral language away from these two logical principles (Blackburn, 1984, 200).

My approach, however, is to simply look at these principles again within the context of either the pretence assertion account of fictionalism vs. the tacit operator system discussed by Joyce. Under the tacit operator system, what is literally being said is that "[according to the fiction of Greek mythology], Pegasus is a flying horse". This kind of sentence does not seem to conform to the principle of existential generalisation, as it does not seem to imply that there are flying horses. It is also clear that it does not conform to the principle of predication either, because, in this case, we are (tacitly) saying that Pegasus is fictional. Frustratingly, however, the tacit operator system whilst seemingly more equipped to handle these difficulties is problematic in the manner described earlier in chapter six. As a result, we preferred the pretence assertion account, which **appears** to assert that "Pegasus is a flying horse", but in reality this is a pretence, thus it retains a more straight forward semantic content. However, it is important to make this distinction else it exposes this brand of fictionalism to the aforementioned problems that come with existential generalisation and predicative grammatical structure. The simplest solution then seems to be to deny the claim that the utterance is **really** asserted in the same way that a realist sentence would be. Instead,

as with keeping with the pretence assertion account, the speaker only **pretends** to assert the information. In other words, "Pegasus is a flying horse" is not the same kind of assertion, if indeed it is even a real assertion at all, as the one I would be making if, for example, I said that "water molecules consist of 2 hydrogen & 1 oxygen atom(s)", unless of course you also happen to be a fictionalist about scientific discourse about molecules, which in this case I am not. One is a pretence assertion, an utterance resembling an assertion made within a cooperative dialog context in which such utterances are treated exactly **like** real assertions; the other **is** a real assertion in the sense that it conveys information about the real world. So if we treat "Pegasus is a flying horse", as not really being asserted, it then becomes less clear that it is as vulnerable to problems with predication and existential generalisation, because it is less clear that, for example, there is a **real** implication that there are flying horses, or that there is a **real** implication that Pegasus actually exists. I argue that there is not and that the speaker who utters the sentence is in no way, deliberately or inadvertently, attempting to deceive the listener into really believing that flying horses exist. This does, however, seem like a very thin defence leaving the overall meaning of the sentence in question. If the semantic content of the sentence does not convey any real information to the listener, then what is the point in even uttering it? This is exactly why the pragmatic dimension of fictional discourse is so vitally important, as I argue that that is where the real **meaning** of the sentence is conveyed. This is, of course, also heavily dependent on the listener understanding the rules of the fictionalist dialog in question, because as we see in the film *Galaxy Quest*, where the Thermians interpreted a science fiction TV show as "historical documents", not understanding that what was really being conveyed meaning wise lay behind an act of pretence led the listener(s) to take the fictionalist utterance as literal information about the real world, which in the film led to disastrous consequences.

To conclude then, the threat of problems like predicative principles and existential generalisation can, I argue, be circumvented by standing as a 'resolute' fictionalist in a similar way to how Smyth defended objections to quasi-realism by standing as a 'resolute' expressivist. This seems to be the case both for the pretence assertion account and for the tacit operator account alike. In both cases, the solution seems to be to simply avoid any predication that implies that the object in question is real and not fictional. In the tacit operator account, we (tacitly) say that Pegasus is a fictional object. In the pretence assertion account, we do not assert in the first place that Pegasus is anything other than a fictional object; we merely pretend to do so. However, in taking this approach, it is more important

than ever to decipher what the ultimate purpose behind fictional utterances, and particularly important for our moral fictionalist utterances, is meant to be. In other words, it may be enough to avoid many semantic problems by simply saying that the moral discourse of non-realists is fictionalist, that we do not make assertions, we simply engage in an act of pretence, but this is not enough to explain the communicative **act** of moral discourse and what is really being **communicated**. In my next chapter, where I talk about Olson's case against fictionalism, I will unveil a brand of fictionalism that I have specifically designed to tackle this issue in particular.

Chapter 9: Olson & the case against Fictionalism

If we were to attempt to truly undermine the fictionalist account that we are currently working with, then we would need to look specifically at a non-hermeneutic account. One notable example of such an assault on revolutionary fictionalism in particular is in Jonas Olson's paper 'Getting Real about Moral Fictionalism', in which Olson talks specifically about Joycean fictionalism. In this chapter I will examine Olson's arguments against revolutionary, or what he calls 'revisionary' fictionalism, mentioning first that Olson seems to assign a necessary connection with error theory to the revisionary fictionalist account. Whilst this is entirely appropriate in, for example, a Joycean account of moral fictionalism, it does not necessarily carry the same force if directed at a brand of moral fictionalism that is only revisionary in the sense that it is not hermeneutic. Earlier in the thesis we have looked at standard moral discourse, quasi-realism and the Frege-Geach problem. Now, for the purposes of countering Olson, and as a springboard into section three of the thesis, I will start to look at the discourse of a particular kind of non-realist; this is what I will ultimately refer to as 'Expressive Fictionalism'. This is also where my account begins to differ from Joyce's, because I claim that by constructing an account by which a non-realist **can** consistently speak of morality in a fictionalist manner, I do not say that they **ought** to speak in this way for the sake of ontological correctness. With this, I believe, unique approach to fictionalism, I avoid having to commit to error theory in order that my account should function. There are, however, other concerns which Olson does raise that I answer as well.

One objection that Olson raises concerns about is whether moral fictionalism can adequately deal with the phenomenon of moral disagreement. In his second section, Olson explains the route between moral error theory to revisionary fictionalism, in his third section he explains why moral disagreement must be accounted for, in both revisionary and hermeneutic fictionalism, in his fourth section he leaves hermeneutic fictionalism behind and specifically confronts Joycean fictionalism, and finally in his fifth section Olson introduces an alternative which he calls 'conservationism', which he explains as the "preservation of ordinary (faulty) moral discourse" (Olson, 2011, 183). I will now go through each of Olson's sections individually. I will argue that much of Olson's criticism of Joyce is successful but that my alternative form of non-hermeneutic moral fictionalism is not as vulnerable to his objections.

In his second section, Olson claims that the attraction of revisionary fictionalism is that it can very easily be adopted by anyone convinced of error theory but who does not wish to accept an abolitionist approach (i.e. that we 'ought' not to use moral language in the absence of any accurate moral ontology) (Olson, 2011, 183). I certainly agree with this in the sense that if, for example, I were to commit fully to an error theoretic account of moral discourse, that revisionary fictionalism is preferable to abolitionism, which I argue is significantly more expensive, because it incurs greater change, meaning that we 'ought' to change how we actually speak and the words we actually say rather than simply to change how we think about what we are already saying. Certainly I argue that abolitionism is an extremely costly position to accept. However, I am convinced that this is not the only path to a revisionary fictionalism of some kind. Suppose, for example, that an individual wishes to take an **agnostic** position on moral ontology, where then does this leave them if they then want to account for moral semantics? Normal moral language to listen to certainly appears to 'resemble' the account favoured by the realist, but to commit to moral realism is also to commit to moral ontology, which the agnostic is disinclined to do, as the former is dependent on the latter. In other words, this is also a rather expensive option. Alternatively, they may embrace an anti-realist semantic account leading for instance to something resembling expressivism (which has merits, but is vulnerable to difficult problems e.g. the Frege-Geach problem, and is thus expensive in other ways)[29]. Alternatively, they could embrace fictionalism, which then raises the issue of whether they embrace hermeneutic or revisionary fictionalism. Hermeneutic fictionalism, however, is also vulnerable to expensive problems, particularly when we look at moral discourse empirically (as discussed in my previous chapter). So by process of elimination, the moral agnostic may also easily arrive at a kind of revisionary, or at least non-hermeneutic fictionalism just as the individual who is convinced of error theory might do.

In his third section, Olson claims that moral fictionalism should be able to accommodate an explanation of moral disagreement. This is because, as Olson claims, it is important to explain how moral discourse is useful in having the function of "coordinating and regulating interpersonal relationships" (Olson, 2011, 184). I will address this lower down. In doing so, he makes a well known distinction between what he calls 'content' moral fictionalism (what we have in previous chapters described as the 'tacit operator' account), and 'force' moral

[29] See Section one chapter six and my analysis of Frege-Geach.

fictionalism (what we have preferred to call the 'pretence assertion' account). Olson tackles content moral fictionalism first. However, whilst his analysis of content moral fictionalism is very interesting, it is not strictly speaking relevant to my account given that we are not constructing a theory of moral semantics which uses a tacit operator. Therefore, I shall proceed directly to his analysis of force moral fictionalism. Under the force moral fictionalist account, two speakers engaged in a moral argument cannot have a genuine moral **disagreement**, because neither of them is making any genuine moral assertions. Rather, they have a pretence disagreement subject to the context of that dialog. This is problematic for Olson, because it has "no dialectical advantage over a view that allows only cases of apparent moral disagreement (or 'disagreement in attitude') and apparent logical fallacies." (Olson, 2011, 191) Here Olson is directly comparing force moral fictionalism to quasi-realism, which makes a similar move. In short, this move runs a serious risk of placing fictionalism into the same firing line of the Frege-Geach problem that quasi-realism incurs so much cost to circumvent, because it calls into question whether or not we can logically make moral inferences in the absence of genuine moral truth aptness. This is additionally problematic, because if we cannot make real moral inferences, then we cannot accommodate an account in which conflicting moral beliefs are inconsistent and that genuine moral disagreements can occur. As a result, the moral agnostic, in searching for an economic theory of moral language, may find that revisionary fictionalism is a more expensive option than they previously thought. Blackburn attempts to circumvent this issue through construction of a quasi-realist logic of attitudes, but fictionalism seems not to have a similar story in response to this issue. However, I argue that this is because it does not necessarily **require** one, and thus it is still an economic position to take.

To make one thing clear, it is Olson that raises this connection of the Frege-Geach problem:

> ...another advantage of moral fictionalism is its immunity to Frege-Geach problems, i.e. problems having to do with inferences involving moral terms. These problems are notorious for quasi-realists, but NRW (Daniel Nolan, Greg Restall and Caroline West) claim that they don't even arise for moral fictionalists. But given that we have opted for a version of moral fictionalism according to which utterances of moral sentences are pretence assertions, this is far from clear. (Olson, 2011, 190)

In other words, force moral fictionalism faces a similar threat that expressivists face, but for slightly different reasons. For expressivism the difficulty is in making an inference where at least one of the premises is, "Boo! (ϖ)" (an expression rather than an asserted proposition), whilst another similar utterance is embedded in a conditional. Force moral fictionalism may face a similar difficulty in making inferences where at least one of the premises is a pretence assertion, while another similar utterance is also embedded in a conditional:

- Truth Apt Content

1) **if** *it is the case that* {stealing is wrong}, **then** *it is the case that* [arresting a thief is right]
2) *it is the case that* {stealing is wrong}

C) **therefore**, *it is the case that* [arresting a thief is right]
Valid Argument

- Force Moral Fictionalist Content

1) **if** *it is the case that* {stealing is wrong}, **then** *it is the case that* [arresting a thief is right]
2) {stealing is wrong}---*Not <u>actually</u> asserted*

C) **therefore**, *it is the case that* [arresting a thief is right]
Invalid Argument*--- (2) is not an asserted proposition, it is merely <u>treated</u> as one.

So the problem seems to come from the conditional '**if** *it is the case that* (x)', where (x) is not asserted in premise (2) and thus is not the same as the antecedent of the conditional in (1). However, where force moral fictionalism differs from expressivism is that the semantic structure is still there, albeit conveyed in a fictitious way. My point is that this does seem to be a problem if we want to accommodate genuine inconsistency of moral inferences, but this is not something which force moral fictionalism claims to be able to do. It is enough, I argue, to place the entirety of the inference in a fictional context:

Force Moral Fictionalist Content
Let us pretend that...

1) **if** *it is the case that* {stealing is wrong}, **then** *it is the case that* [arresting a thief is right]
2) *it is the case that* {stealing is wrong}

C) **therefore**, *it is the case that* [arresting a thief is right]

Valid Argument*--- Within the context of '*let us pretend that...*'

In other words, what we get is pretence consistency but consistency nonetheless. The only problem I see with this is that this looks a lot more like a tacit operator than pretence assertion. However, I argue that this "*let us pretend that...*" condition does not need to be overtly stated provided that the listener understands that it is there. For example, if I pretend to assert something about Spiderman, then the listener's knowledge of that context should provide the "*let us pretend that...*" context/condition required and for any supposed inferences made within that context to be treated as sound and valid. So my argument here is that force moral fictionalists cannot account for genuine moral consistency or inconsistency in the case of actual moral propositions simply because they cannot infer on what they do not assert, but this is not something which force moral fictionalism ever set out to do. It is not a means by which speakers describe or infer on what is **actually** going on, rather what **would** be going on were morality real (at least as far as the non-realist/fictionalist is concerned), and so it is enough to **suppose** valid moral inferences, but it is not required that we actually make them.

However, surely this still leaves us with the problem of lacking a convincing account of moral disagreement. Again, however, I maintain that this is not problematic in so far as it is required on a propositional level. Assuming, for the sake of argument, that we cannot accommodate moral disagreement under fictionalist semantics, the problem then is, as Olson puts it, that we lose an explanation of the useful function that moral discourse has at coordinating and regulating interpersonal relationships (Olson, 2011, 184). I suggest that this is only true if we focus on moral semantics to the exclusion of moral pragmatics, which I suggest is more than capable of doing this. It may be enough to suggest that we can explain the use of pretence moral disagreement in terms of communicating conflicts of interest or attitude, just as expressivists do. So in terms of the old cat thought experiment, we can explain the function of a pretence disagreement about cat kicking, where I say, "kicking the cat is wrong", and another objects by saying that one communicates a motivating attitude towards cat kicking and another communicates an attitude towards the opposite. Even if the two utterances are not semantically inconsistent, as no genuine assertions are made about whether or not it is really wrong to kick the cat, there nevertheless remains a conflict of interest, because both attitudes are incompatible when acted upon; one intends the cat to be

kicked, and the other intends the cat not to be kicked, but the cat cannot be both kicked and not kicked in the same universe. If we want to maintain a fictionalist account of moral disagreement, then we should argue that moral disagreements are similar to disagreements about fictions. For example, I may claim that Hamlet is mad, and another may disagree and say that Hamlet is only pretending to be mad. However, to have such a disagreement relies on both of us pretending that Hamlet is a real person and that the events of Shakespeare's play really took place, hence we have a disagreement within a pretence, but a disagreement nonetheless. For a moral fictionalist, ethics work in a very similar way, the only real difference is that the consequences of disagreements within the pretence of right and wrong are far more profound.

We may choose to counter all of this by saying that not all moral utterances communicate any attitudes or intentions. For example, I may say that, "kicking cats is wrong, but I don't care if it happens". My response to this is very simple, and that is to say that this is actually an inconsistent sentence (on a pragmatic level), if **sincerely** uttered by a fictionalist of the kind I am describing i.e. one who uses fictionalism as a semantic basis for pragmatic expression. In other words, if the motivation towards moral language is ultimately to express emotions towards certain behaviours in an authoritative manner, thereby instilling a sense of 'must' in the listener, then it makes little to no sense to use fictionalist semantics to describe the opposite of what the speaker is actually feeling, unless they are deliberately being **insincere**. In other words it seems intrinsic to moral language and its usage to be inherently motivating, whether the content is believed by the speaker or not, especially in an environment awash with moral realists, who do believe the content. Were I to assert, directly, that I am merely 'uncomfortable' with the kicking of cats, but that I do not care if it happens, then I would be directly contradicting myself, because the whole point of me phrasing it like that is to communicate the fact that I **do** care and that I want it to stop.

In his fourth section, Olson targets Joyce more directly. This involves making the case that Joyce fails to 'sufficiently' motivate his target audience towards revisionary fictionalism, and that many of the analogies he uses are highly vulnerable to counter example(s). His most salient example of this is on Joyce's claim that true beliefs are inherently valuable and that false beliefs are inherently disvaluable, the idea being that fictionalism is attractive in virtue of its ability to preserve normal sounding moral discourse without consciously embracing false ontological beliefs (Olson, 2011, 193). Olson objects to this premise. For example, he

agrees that someone who is about to eat a poisonous apple would do well to believe that the apple is poisonous, and so true environmental beliefs can be inherently valuable. This can also be true of accurate mathematical beliefs; if someone is being pursued by two wolves, and one breaks off, the prey would do well to believe that there is one still pursuing (Olson, 2011, 193)[30]. However, as Olson points out, there are very significant counter examples. Olson talks about very controversial philosophical positions, such as the value of the 'illusion' of libertarian free will or the disvalue of beliefs on matters like personal identity, bias, responsibility and desert. However, I think there are less controversial examples that could lend weight to Olson's argument. In terms of beneficial false (or at the very least empirically unjustified) belief, one need only look at any pharmaceutical clinical trial and the effect(s) of placebo to confirm that there are cases of this, e.g. giving a subject a sugar pill and telling them that it will help them to quit smoking. Similarly, practices like meditation, yoga or hypnosis are widely regarded as producing similar results if embraced. Similarly, whether any particular world religion is true or not, it is hard to deny that in many, for example in Judeo-Christianity and Islam, Hinduism or Zen-Buddhism, the belief in the efficacy of prayer and meditation appears to have psychologically beneficial effects on believers. In terms of detrimental true beliefs, the discovery of any negative yet completely unalterable and/or irrelevant fact could, I argue, in some situations, be better left unknown[31].

Regardless, Olson's point is that Joyce should not infer that fictionalism has any advantage based on the intrinsic value of true belief, and I am inclined to agree here. What Olson suggests in response is what he calls conservationism, which he explains in his fifth section. Conservationism, as Olson describes it, is the theory that we should preserve normal moral discourse but that we should simply embrace false beliefs and false assertions. So in the cat case, this means that "kicking cats is wrong" is an assertion, but it is false, and it is

[30] I would also argue this is largely true of accurate scientific beliefs; if it is true that the altitude of a jump is insufficient for a parachute to sufficiently reduce the velocity of the fall, and the impact force upon landing, or that nuclear fission for useful energy production is mechanically achievable under certain very specific physical conditions, then someone to whom this kind of information applies would do well to believe it. Similarly, in terms of false beliefs, if embraced, a belief that spending an inordinate quantity of money on a race horse with very extreme odds against it is 'likely' to go well is statistically unlikely to pay off; a belief that one is immune to bullets also has the potential to cause intense, and brief, disillusionment.

[31] It is entirely possible, scientifically speaking, that a nearby star may supernova, or has already done so in the last few decades or centuries, and for a catastrophic amount of solar radiation to travel through our solar system, travelling at the speed of light and thereby giving us absolutely no warning of an impending extinction level event (which we can do literally nothing to prevent) comparable to a similar event that may have happened 440 million years ago [during the Ordovician period] in which 70% of marine and all terrestrial life on earth was wiped out. However, whilst it is true that this is a genuine scientific possibility, it is questionable how useful believing this is, given that there is nothing possible that can be done about it. To the contrary, this information is likely to further the believer's anxiety about the possibility of bad things happening if focussed on, so it is arguably better not to focus on it, or even to completely ignore it.

deliberately false i.e. the speaker does not believe that kicking cats is wrong. The advantage of this, Olson claims, is that conservationism does not need any self surveillance to be maintained when using moral sentences (Olson, 2011, 198). Olson seems to base this idea on a key passage from Hume's *The Sceptic*:

> If we can depend upon any principle, which we learn from philosophy, this, I think, may be considered as certain and undoubted, that there is nothing, in itself, valuable, or despicable, desirable or hateful, beautiful or deformed; but that these attributes arise from the particular constitution and fabric of human sentiment and affection. [...] We may push the same observation further, and may conclude, that, even when the mind operates alone, and feeling the sentiment of blame or approbation, pronounces one object deformed and odious, another beautiful and amiable, I say, that, even in this case, those qualities are not really in the objects, but belong entirely to the sentiment of that mind that blames or praises. (Hume, 1993, 97)

Hume advocates habit as a 'powerful means of reforming the mind and implanting in it good dispositions and inclinations' (Hume, 1993, 97). Olson advocates conserving the language of moral realism, and the semantics of assertion and inference, but with the belief that these assertions are ultimately false. This does seem to preserve semantic simplicity and, via error theory, has a clear position on moral ontology as well. However, I have my worries about this solution.

The first is one of the same worries I myself have about fictionalism on its own; if the speaker deliberately speaks in a way that does not accurately reflect what the listener believes, then there is a case to be made that the speaker is deliberately being deceptive when talking to a moral realist and that this is an unattractive conclusion. Both fictionalism, particularly of the hermeneutic variety, and conservationism depend on some kind of cooperative principle that for the duration of the following conversation, we will **treat** x as true, but this is not obviously the case where a fictionalist, or a conservationist, is talking to a realist, who specifically believes that moral language is true and that it refers to real moral values and properties. Whilst I can address this problem pragmatically under an expressive analysis in regards to the fictionalist, as I will do later on in chapter ten and later again in section three, I am not convinced this is possible if the moral discourse used relies on **deliberately <u>false</u>**, semantic assertions.

My second worry follows from an alternative perspective on fictionalism, and I believe this is an idea that expressivism shares. This is the idea that we should be able to accommodate the intuition that saying something like, "it is wrong to kick cats, but I don't care if it happens", reveals some sort of **logical inconsistency**. Expressivism explains this by stating that the "...it is wrong to kick cats..." clause of the sentence is equivalent to "...Boo! Kicking cats..." from which we should be able to infer observationally that the speaker does care if it happens, which directly contradicts the second clause of the sentence. I argue that we can use a similar move in regards to fictionalism, albeit slightly differently. Rather than the first clause being an expression, we can instead see it as a pretence assertion and one from which we can also reasonably use as a communication that the speaker does care what happens to cats, which much like the expressivist case directly contradicts the second clause. The difference is, however, that this is not a semantic inconsistency, but rather a pragmatic one:

- Fictionalist/Pragmatic Expressivist Content

1) *Let us pretend that...it is the case that* {kicking cats is wrong}[32]
2) (1) = {I care if cats are kicked}[33]
3) *it is the case that* ~{I care if cats are kicked}

C) therefore, *it is the case that* [{I care if cats are kicked} ^ ~{I care if cats are kicked}]

Inconsistent Argument* --- At least in terms of its **communicative function** i.e. it is inconsistent to communicate that I both do and do not care if cats are kicked.

Conservationism, however, cannot make this move:

- Conservationist Content

1) *it is the case that* {kicking cats is wrong}---**false**
2) *it is the case that* ~{I care if cats are kicked}---**true**

C) therefore, *it is the case that* [{kicking cats is wrong} ^ ~{I care if cats are kicked}]---**false**

[32] "*Let us pretend that...it is the case that* {x}", is a pretence assertion, which, if left without additional, explanatory premises, we cannot make inferences from.
[33] Whilst this is not a stated assertion, it is, however, a truth condition which we can pragmatically infer from as I will outline in section three chapters four and five.

Valid (Unsound) Argument* --- (1) is false, rendering (C) also false, so this is still a consistent, albeit unsound, argument.

So under a conservationist analysis, the sentence, "kicking cats is wrong, but I don't care if care if it happens", comes from a perfectly valid and consistent argument; it is simply interpreted as false. The fictionalist, with a little pragmatic expressivism, on the other hand can tackle this sentence with significantly more force by saying that the sentence is not **only** false, but inconsistent on logical grounds (albeit in terms of its pragmatic, communicative function rather than under a very strict, exclusively semantic analysis of its spoken content).

My third and final worry is that conservationism seems to heavily rely on error theory, which is an ontological commitment in and of itself, and one which I am inclined to avoid. If the speaker actively believes that when they say a particular moral sentence that the sentence is false, then this surely seems to suggest that they are communicating a commitment of some kind towards moral ontology on some level as well. I believe that the kind of fictionalism (with subtle hints of expressivism) that I am peddling, which Joyce is not, has a significant advantage here over Olson's conservationism, because whilst conservationism seems to rely on error theory, this newer brand of fictionalism does not. This is because the claim it makes is not one about how morals really are, nor is it a claim about how moral agents 'ought' to be speaking if they are to be correct; I believe that both kinds of claim are unnecessarily limiting and committal. My claim is one about how a non-realist can consistently speak, whether error theory is true or not. Suppose, for the sake of argument, that error theory is false. Suppose next that morals do, ontologically, **exist** on some level[34]. In such a case, the conservationist might intend to speak falsely about ethics not knowing that what they are saying is actually referentially true. By contrast, this possibility should not faze the non-realist that is **agnostic** when it comes to moral ontology. That it is ontologically wrong to kick cats does not alter the truth value of what they are saying, because no propositions are really being made, nor does it alter the information they are attempting to convey, because whether kicking cats is ontologically wrong or not, the point in saying so is to communicate attitudes with an **air** of authority motivating action, which stands regardless of what the moral ontology is or is not.

[34] That morals have some kind of objective, existential qualities is certainly not without a convincing case; however, my concern has always been in regards to the suggestion that we can **access** them epistemologically (see section one chapter two). My position argues that the question of whether or not morality objectively 'exists' and in what sense is less serious a question than the one about how on earth we could access or have knowledge of it if even if it does.

In other words, moral ontology need not be seen as necessarily error laden, rather it is simply **irrelevant** as far as the agnostic non-realist is concerned. So the advantage is one of a versatility and flexibility which conservationism lacks. Nevertheless, Olson's arguments are extremely useful in that they help to highlight the key differences between the Joycean fictionalist account and my own.

This, I believe, unique approach to the case against fictionalism, in which I take the stance of an agnostic in regards to moral ontology, and infuse moral fictionalism with pragmatic elements of expressivism, is what I 'Expressive Fictionalism'. In my final chapter of this section, I will talk more about expressive fictionalism and discuss a little of what is to come in section three.

Chapter 10: Expressive Fictionalism

So what exactly is my own account, and how exactly does it differ from the Joycean one? Unlike the argumentative strategy used by Lewis and Blackburn in chapter five to separate the realist from the quasi-realist and the quasi-realist from the fictionalist, by pointing out what the other says which I do not, I will instead do the opposite and point out what I say, which I believe Joyce does not. In other words, I will make a similar move in regards to moral fictionalism that Nicholas Smyth makes in regards to quasi-realism, where I attempt to help it out by modifying it slightly. I do not believe that the semantic components of moral fictionalism are as problematic as Jonas Olson supposes. However, Olson does make an excellent point that when we follow through on moral fictionalism, we begin to lack a convincing account about the true **function** of non-realist moral discourse. In other words, if non-realists speak about morality in a fictionalist manner, then fictionalism accounts for how that works, but it struggles to account for **why** fictionalists do this. Why not simply adhere to error theory, talk about intentions and attitudes but leave 'morality' out of it? Why the pretence? Why make your moral discourse **sound** realist when in reality it is not? In chapter seven I describe Joyce's account of moral fictionalism, which concerns motivations and intentions. According to Joyce, if I convince myself (at least when I am not in a 'critical context'), that I 'must' do something, then I am more likely to actually do it, and this is somehow beneficial. I think that Joyce is right in saying this, but I am convinced that there is a great deal more to say about this, and I believe that there is room for some pragmatic expressivism here. In other words, I believe that we can utilize fictionalist semantics and pragmatic expressivism, and effectively splice them together. This is the approach I call, '**Expressive Fictionalism**'.

Something I will discuss at great length in section three is the relationship between moral language and the psychology of emotions and attitudes. We often use words like 'good' and 'bad' both to describe moral properties and to describe positive and negative **feelings**. We also have a tendency to immerse moral language in a language of feeling, "I feel like this is the right thing to do" or, "This feels wrong". We are also inclined to describe 'moral' acts in emotive terms, for example, 'What a joy it is to see you behave so responsibly', or, 'I am disgusted by your behaviour today'. We also appear to struggle to separate emotions from moral discourse. For example, in my undergraduate days when I studied English literature,

we spent a week studying *Lolita* by Vladimir Nabokov, the subject matter of which is famously controversial, and so unsurprisingly nearly all of our discussions were emotionally charged. Much as we do with works of fiction, we have a tendency to become very much **invested** in the subject matter particularly where issues associated with morality are concerned. It is at least partially owing to these observations, among many others I will explore in section three, that I am convinced that expressivism is actually onto something when it describes moral language in terms of approving and disapproving **attitudes**[35]. What I think is interesting, however, is how compatible this seems to be, at least on a pragmatic level, with the kind of fictionalist semantics and their usage described by Joyce, albeit with some nuanced differences. In previous chapters we have both talked about the relationship between pretence in the form of fiction and motivation towards encouraging or discouraging action. If I convince myself that I 'must' do (x) amount of sit ups a day, then it is more likely that I will actually do it. To compel others, we would do well to also instil a sense of 'must' in them. What I think Joyce misses, however, is the relationship between inclinations towards affirmative or discouraging action or intention, and emotional, attitudinal dispositions towards these things and the communication of those attitudes that I believe are taking place in and around the pretence. In order to motivate a sense of 'must' in myself or others, I 'must' surely appeal to some **force** which is inherently motivating. One potent means of doing this is in appealing to human emotion. For example, if I want to 'instil' a sense of 'must not' in a child, then I am more likely to describe it in more decisive terms like, 'kicking cats **is wrong**', or at least, 'kicking cats **is unacceptable behaviour**', than a more passive 'please don't kick the cat; it upsets me'.

I argue that what is being communicated between all of these sentences is **almost** identical, but that the former, more decisive utterances, in maintaining a **pretence** about 'must' and 'must not', is subtly different, because rather than being phrased as a passive request, like the latter sentence, it is portrayed as coming from a position of **authority**; this is particularly important, for example, in child development. In other words we appeal to the child on an emotional level not only in terms of their capacity for empathy (which in the case of children is often still developing, as any parent or teacher will attest to), but also their **fear** of looking

[35] 'Attitude' as a word is not without variations in meaning, however, and so for the purposes of this, let us take 'attitude' to mean a person's emotional disposition towards something; to have an approving attitude towards kindness means that I feel positive and affirm kindness as a 'good' thing, and to have a disapproving attitude towards cruelty means that I feel negative and repulsed by cruelty as a 'bad' thing. Let us then suppose that I am disposed or inclined to pursue and/or encourage 'good' things and to avoid and/or discourage 'bad' things.

'bad' in the eyes of others and being punished or chastised for it. As I will investigate further in section three, fear (of reprimand) even carries through into adulthood, not only in our justice system but on a social level as well, where much of how we speak and act is determined by concerns we have about how we **appear** to others. Concerns we have about how we appear to **ourselves** also play a crucial role, for example I argue that much of what constitutes a stable and positive sense of self-esteem often involves **convincing** ourselves that we are a 'good person'. To a non-realist, this convincing involves some level of pretence, but it is nevertheless important, because as I will discuss later on in section three, the absence of this pretence is problematic not only socially, but also psychologically.

Before continuing to build upon this idea in section three, let us first recap where we are right now. In our search for a functioning non-realist account of moral language, we first looked at the case against moral realism to give us some foundation to start with. We then discussed some of the useful terminology including expressivism, non-cognitivism and projectivism, and then looked at some of the other less common forms of non-realism, that we will not be focussing on and why that is the case. Next, we focussed our attention on quasi-realism, a seemingly promising form of non-cognitivist projectivism, which seemed like a worthy candidate on which to base a functioning non-realist account of moral language. This then became the primary focus of section one, in which we discussed the origins of quasi-realism in *Spread the Word: Groundings in the Philosophy of Language* by Blackburn, the Frege-Geach problem and Blackburn's proposed solution to it, and then we looked at Andy Egan's objections and Nicholas Smyth's defence coming from the position of the 'resolute expressivist'. We then ended section one feeling reasonably positive about quasi-realism and its prospects.

This then brought us to the beginning of section two, where we began to evaluate those objections to quasi-realism which were not so easily thwarted, in particular those of G.F. Schueler, in which I cited a new problem for quasi-realism that I have noticed, that being its apparent inability to accommodate the subtle differences between general vs. personal moral commitments and the difficulties that poses in demonstrating attitude inconsistency. We then looked at Van Roojen's objections and a second defence by Smyth, in which we began to suspect that Smyth's expressivism was not the same thing as quasi-realism. This then became very much apparent in chapter five when we looked at David Lewes's paper, and Blackburn's reply/response 'Quasi-Realism no Fictionalism', where we concluded that Blackburn had no

intention of preserving quasi-realism as resolutely expressivist in the same way that Smyth does, and we noticed further that this inability to accept any aid from the fictionalist also helped to condemn quasi-realism to the same problems that I noticed in my chapters on Schueler. Finally, we concluded that Blackburn seems to move away from the quasi- towards the realism, at the expense of his expressivism. As such, it became apparent that fictionalism was the far purer and far less costly form of non-realism of the two.

We then focussed our attention on fictionalism, and in particular the account peddled by Richard Joyce, in which we discussed the difference between the tacit operator and the pretence assertion, as well as the difference between hermeneutic and revolutionary fictionalism, concluding that, like Joyce, we favoured a revolutionary account based on pretence assertion. We then looked at how this applied to moral discourse semantically and began to look at **some** of the pragmatic elements as well. We then looked at some of the objections to fictionalism and the problem of fictional discourse. This then brought us to the previous chapter detailing the objections of Jonas Olson to Joycean fictionalism. In that chapter, we concluded that fictionalist semantics are not necessarily as vulnerable as Olson supposes, but that there are major concerns which the Joycean fictionalist struggles to deal with, most notably the ability to account for the **function** of non-realist moral discourse.

Joyce does provide an account of emotions and motivation in regards to moral fictionalism in which he likens moral language to a language of personal commitments, (e.g. 'I **must** do x number of sit ups a day in order to maintain my fitness' or 'I **mustn't** smoke or else I will die of lung cancer'). Such pretences motivate reinforcing or preventative action, and interestingly these kinds of sentiment are reminiscent of expressivism. However, I argue that this line of thinking, whilst provocative, is thus far underdeveloped and that Joyce could say a great deal more on this, namely about how these pretences are applied in varying levels of critical contexts, whether or not these are constituted by attitudes, how these attitudes may be communicated to others, whether the Frege-Geach problem is a concern here or not and whether or not there is an issue of deception involved in these kinds of pretence assertions. All of this and more I delve into in section three to provide some insight into what a more developed Joycean account of motivation may look like, and what happens when we combine revisionary fictionalist semantics with pragmatic expressivism.

My solution to this issue, therefore, which I have discussed in this chapter, is to use what we have learned in the first half of the thesis from the expressivists and quasi-realists, and to work this **around** the semantics proposed by fictionalists. My hope is that this will finally be what we have been looking for: a non-realist account which successfully accounts for the semantics of non-realist moral discourse **and** gives us a convincing account of the pragmatic discourse involved as well. My hope is also that this will be compatible with both error theory and moral realism alike, as it stems from neither, but rather from a position of agnosticism regarding moral ontology. Much like quasi-realism, I claim that what we treat as morally true or false stems from a psychological source rather than a metaphysical one. Unlike quasi-realism, whilst communication of attitudes plays a central role, these attitudes are not communicated **semantically** but rather through an alternative means of **pragmatic** discourse, an account of which I will build upon throughout section three. However, there is still a semantic **content** uttered, which exists in a form reflecting a Joycean fictionalist account in which we speak about 'morality' as a form of **pretence** in which we instil a sense of 'must' in both the listener(s) and in ourselves. It falls then to section three to explain exactly how these two different components of expressive fictionalism ultimately fit together.

Section 3: Expressive Fictionalism

Chapter 1: (Introduction)

In this chapter I will be doing a little sign posting about what to expect in the third, final and most important section of my thesis. Having now explored non-realist moral language and having, I believe, established that a Joycean fictionalist account is a worthy candidate for the semantics of non-realists, the task is then, as I have mentioned, to marry that with a pragmatic component thereby creating a more complete analysis of the moral communication of non-realists. In this section, I will explore that component and in so doing go through various stages in that process.

The first stage, which I will discuss in chapter two, will be to briefly talk about the differences between semantic and pragmatic discourse in order to give us some insight into what exactly we are looking for when we talk about providing a pragmatic component to our theories about the moral communication of non-realists. For instance, what effects do phenomena such as irony, intonation, emphasis or conversational implicature have on the meaning of an utterance? How is meaning conveyed beyond or around the semantic content of what words literally translate to within a given sentence? What role does context play in determining the meaning behind what people are saying, or moreover where, when and how they are saying it? In exploring these well known questions, but (less commonly) applying them specifically to moral communication, I hope to demonstrate the importance of going beyond semantic theories, and to place previously discussed semantics firmly within a broader context of a more complete analysis of moral communication.

In order to make any of what comes later more than mere speculation, however, it is important that I provide some justification for thinking that emotions and attitudes play a central role in moral cognition and moral communication. To that end, I have devoted chapter three entirely to providing some kind of empirical, scientific basis for this link between emotional sensitivity and moral thought. In that chapter, I look at correlations between attitudes and behaviour with neurobiological phenomena in the ventromedial prefrontal cortex, observations of behavioural anomalies in humpback whales and case studies and testimonies of severe depression and anti-depressant medications, specifically of selective

serotonin reuptake inhibitors. By the end of that chapter, I conclude that whilst it would be difficult to demonstrate a necessary or logically deductible causal link between emotion sensitivity and moral cognition, there does seem to be more than enough empirical evidence to demonstrate at the very least a correlation and in all probability some kind of causal link to that effect.

Once that link has been established, I will then set about applying this emotional sensitivity to moral language, developing a kind of pragmatic expressivism to fictionalist semantics. In that chapter, I will introduce two separate models by which we may translate emotions into moral language, these are the pride, shame and responsibility model and the empathy model. I will then construct a ten stage inferential process by which emotion sensitivity in the form of empathy may be psychologically translated into the moral language of a fictionalist. In the following chapter, I expand further on this empathy model, and begin to break down this ten stage process, looking at each stage individually to determine how well it stands up to scrutiny. In the process, I also attempt to combine and integrate the empathy model with the previous pride, shame and responsibility model.

Chapter six then revisits the previously explored Joycean concept of critical contexts, which I will endeavour to expand upon myself by applying my new expressive fictionalism to it. In this chapter, I introduce concepts which further elaborate on how expressive fictionalism works in critical contexts, such as that of emotional override and the problem of deception. Chapters seven through to nine then focus on applying expressive fictionalism to an argument **for** a concept of moral progress. Unlike Blackburn, I do not argue for a notion of moral truth within expressive fictionalism. However, I do argue that there are nevertheless still grounds for differentiation between moral actions and attitudes on grounds of quality and consistency within the morality fiction and that we can apply that to historical analyses to determine moral progress or regress. However, in keeping with the fictionalist content of the moral language involved, constituted not by assertion but by pretence assertion, this too is part of the pretence when analysed within a maximised critical context. This becomes evident particularly in chapter eight when I examine a very similar approach by the pragmatic expressivist Catherine Wilson and her comparison between the historical narratives of slavery and the phlogiston theory of combustion. Finally in chapter nine, I don the pretence mask to demonstrate how we may apply an expressive fictionalist style of language when talking

about moral progress throughout human history. In chapter ten, I will then put all that we have discussed here together and draw my final conclusions.

By the end of my thesis, I aim to finally have what we have been searching for; that is a complete theory of non-realist moral language, which explains how and why we use seemingly realist moral language in the absence of reference to real moral content, which is compatible with an agnostic perspective on the existence of moral properties within metaphysics, that does not rely or supervene on any additional epistemology, that does not constitute a reduction to error theory, that does not compromise its own non-realism, that avoids the Frege-Geach problem and which is ultimately consistent and valid.

Chapter 2: Semantic and Pragmatic Discourse

Communication goes far beyond the code of what words literally mean within a sentence. Nearly all animal species for instance rely heavily on non-verbal communication (e.g. body language). This is not to say that in human communication semantic meanings of words are not important; they certainly are. It is, however, only one component in a vast system of how one communicates. In this chapter I will begin to discuss the concept of 'pragmatics'. To avoid any unnecessary confusion, I will be treating 'pragmatics' as an aspect of communication about which we might theorize about how **context** contributes to the meaning of an utterance. In particular, I will be focussing on phenomena like conversational implicature (Grice, 1989, 26), emphasis and intonation as examples (though it should be noted that many other fields are associated with the term 'pragmatics' e.g. speech act theory, conversation analysis etc). To put things even more simply, I will be looking into systems of how people communicate beyond or outside of the literal meanings of the words they use.

For example, to use a very dramatic case of pragmatic communication, **sarcasm**, we can clearly see how meaning is often **context dependent**. For instance, suppose you already know that I dislike mayonnaise. If upon receiving a sandwich drenched in mayonnaise, I suddenly roll my eyes and say, "oh yeah, I really *really* like the look of *this* sandwich", then my meaning should be apparent---I do not like the look of this sandwich. Sarcasm is so dramatic an example, because the ultimate meaning behind the utterance is the complete opposite of the literal, semantic meaning of the words and sentence that I used. So whilst sarcasm will not prominently feature in any of the following theories of moral pragmatics, it is nonetheless a useful example we can use to demonstrate how dramatically the meaning of an utterance can differ from its semantic meaning within a given context. It is, in effect, a useful springboard into the world of context dependency.

Pragmatics is not merely about altering the meaning behind an utterance from its semantic components; it can also be about augmenting semantic meaning(s) with additional meaning(s). For example, to use a famous Gricean concept, conversational implicature describes how context implies additional information beyond a sentence's semantic content: "Mary had a baby and got married". This sentence **suggests** that Mary had her baby before the wedding. However, the sentence does not say this, and it would still hold true as a sentence with the exact same semantic content were the opposite case true (i.e. if Mary had

the baby after the wedding). Similarly, if I were to say the words, "I ate some pie", this **suggests** that I did not eat all of the pie. However, the sentence's semantic content would still produce a true sentence if I had eaten all of the pie. In audible communication, emphasis can also be an important means of conveying meaning. If I were to say, with emphasis, that, "I ate **some** pie", then this very clearly suggests that I did not eat **all** of the pie (Grice, 1989, 22). Again, however, if we look merely at the semantic content of the sentence, this is not necessarily the case, as the sentence would still be literally true if I had eaten all of the pie, though the listener would no doubt question why I emphasised the word 'some' and may even accuse me of dishonesty despite the fact that, semantically speaking, I am not actually lying.

So what does any of this have to do with moral sentences? Moral sentences constitute communicative acts just like any other sentence. The difference is that unlike many sentences, it is not entirely agreed upon what function or form of communication moral sentences involve. If we are to have any hope of resolving this issue, then we must surely look upon the entirety of the communicative act when speaking about ethical issues. For example, if we take a kind of realist stance on the issue and say that moral sentences **assert** moral facts, then the semantic content of the sentence then becomes very clear and straight forward (at least in a linguistic sense albeit not in a metaphysical sense). What is more open to discussion, however, is the pragmatic motivation and contextual meaning behind the assertion. Suppose we say that the primary motivation behind any assertion is to enlighten the listener. In other words it is to make the listener **aware** of a particular fact and/or set of facts. However, there may be any number of reasons **why** the speaker may do this. Perhaps the speaker feels the listener might benefit from additional moral knowledge, or perhaps the speaker feels they need to adhere to some kind of 'responsibility' and ensure as many people are as aware as possible of certain moral facts. Regardless of the specific motivation behind their utterance, there is surely at least some kind of contextual meaning behind it.

Things become even more interesting if we adopt a non-realist or a non-cognitivist stance. If we assume, for the sake of argument, that moral sentences do not assert moral facts, they nevertheless **appear** as though they do. This raises very interesting questions about what the speaker who speaks in a realist way (but who is not a realist) wants to accomplish by speaking like this. One solution is expressivism, which we have explored in previous chapters and which claims that the semantic content of these sentences convey expressions of

emotional approval/disapproval. This is interesting, because it introduces us to the idea that there is some kind of connection between emotion and ideas and language about ethics; this is something I will come back to. This emotional connection is not so emphasized in moral fictionalism, however, which is the other non-realist solution I have investigated in previous sections.

What motivates the revisionary fictionalist speaker who knows that they are a fictionalist? This is, perhaps, the most important question that should drive the remainder of the thesis. According to the pretence assertion account of moral fictionalism, when we utter moral sentences we are **pretending** to assert moral facts. This is already highly context dependent, the rules of the conversation need to be understood by both speaker and listener, as if to say, "Let us pretend for the duration of the conversation that moral facts exist and that I am referring to them". The important question is simply, **why do this?** In section two, we concluded that according to fictionalism, the semantic content of the sentence remains as it appears. Therefore, any alternative meaning behind the utterance or from the communicative act in general must come from the **context or way** in which it is uttered; this brings us squarely back to the pragmatics of moral discourse.

So what we have here is thus the unenviable task of linking semantic fictionalism with some kind of compatible pragmatic account of moral discourse as a communicative act, as opposed to merely focussing on what moral sentences literally translate as. I have mentioned earlier that when it comes to non-realist accounts of motivations behind moral discourse, as opposed to simply how they work, that expressivism is particularly interesting for several reasons. The first reason is that expressivism has a convincing story to tell us about moral attitudes as opposed to moral beliefs, and why exactly it is so convincing in my opinion I will delve into in the chapter which follows this one. This very much helps us explain the contextual nature of moral discourse. According to expressivism, moral discourse takes place where there are events (or theoretical events) which provoke **attitudes** or feelings of approval and disapproval, and the aim of moral language is to communicate these attitudes. The second reason is that this does not necessarily contradict semantic fictionalism so long as we style ourselves as pragmatic expressivists rather than semantic expressivists. For example, as we saw half way through section two, if we were to try to marry fictionalism with another **semantic** theory within expressivism, such as quasi-realism, then problems would occur. As Simon Blackburn (correctly) points out, quasi-realism is not same animal as fictionalism. If,

however, we focus on those elements of expressivism associated with communicative acts and contexts in which moral discourse occurs, then we may have something which does tie in with and augment the semantic components provided by Joycean fictionalism quite nicely.

Before we do this, however, there is one more important matter to attend to and that is the matter of what basis we have for supposing that moral discourse necessarily has anything whatsoever to do with attitudes or emotions. Unfortunately, it is not enough for me to simply say that the expressivist account of why we use moral language in the way that we do 'sounds' convincing. In other words, we suppose that there is some kind of connection between emotions and attitudes and morals and sensibilities. We suppose that morality, and moral language has something to do with processing and communicating feelings of approval and disapproval. It 'sounds' convincing, but it would most definitely help if we had some **evidence** to suggest that this is indeed the case. This brings us to my next chapter in which I will attempt to provide this.

Chapter 3: Emotions and Ethics---A Scientific Perspective

I would like to begin this chapter by talking about the apparent relationship between moral attitudes (i.e. attitudes of right and wrong), and feelings of strong emotion. I imagine that for most readers, moral decisions usually seem to be coupled with a strong emotional opinion on whatever the issue in question is. For example, I for one tend to feel some strong emotion (e.g. sympathy or in some cases worry for the wellbeing of others) prior to wanting to do something altruistic, such as giving up a seat on a train, lending out a coat or in extreme cases putting my own safety at risk to ensure the wellbeing of another, and that action tends to be followed by different emotions (e.g. relief, satisfaction or pride). I would imagine that others hold similar sentiments as well. It seems then that there is some kind of connection or correlation between emotions and what we would normally recognise as 'moral' decision making. However, this raises some important questions. Just how evident is this correlation, and what **kind** of connection are we even looking for? I argue that it would be difficult at best to find a connection between emotion and moral decision making that is necessary or self-evident. To do so would surely require us to demonstrate that it is somehow logically contradictory to posit that a moral decision can be both sincerely and **dispassionately** made (without any emotion being present or being conveyed). In other words, if there is even **one** example where this occurs, then the conclusion is by definition false. I argue that there is, however, more than enough **empirical** evidence to suggest at the very least a **correlation** between emotion and moral judgement. The question then becomes, as it always does with any empirical research, whether or not this correlation is **causal** or not. Do our feelings cause our moral judgements? Do our moral judgements cause our feelings? Nevertheless, throughout this chapter I will examine some of this evidence (in truth there is far too much to examine in its entirety), and in so doing it should gradually become apparent that the conclusions that we reach clearly provoke questions for any potential discussion about what a moral fictionalist is truly hoping to achieve by communicating in the way that he/she does.

First, let us look at what the evidence for this correlation is. Liane Young and Michael Koenigs suggest that neuroscience provides evidence that emotional processes underlie individuals' moral decision-making. According to their abstract, "Neuroscience offers a unique perspective on this question by addressing whether brain regions associated with emotional processing are involved in moral cognition." (Young & Koenigs, 2007, 69) The study involved conducting a narrative review of neuroscientific studies which focus on the

role of emotion in morality. Specifically, the study describes evidence implicating the ventromedial prefrontal cortex (VMPC), a region of the brain known to be associated with emotional processing (Young & Koenigs, 2007, 69). MRI scans demonstrated VMPC activation during tasks probing moral cognition. In addition, studies of clinical populations, specifically patients with damage to their VMPC, revealed an association between impairments in emotional processing and impairments in moral judgement and behaviour. The study concludes that these studies indicate "that not only are emotions engaged during moral cognition, but that emotions, particularly those mediated by VMPC, are in fact critical for human morality." (Young & Koenigs, 2007, 69).

Assuming that the evidence provided by Young and Koenigs is correct, then there is a clear and evident correlation between emotion states and moral cognition and moral decision making. However, we would do well not to get too carried away here. What the initial part of this study shows us is that there is an evident correlation rather than a **necessary** causal relation. The second part of the study is, however, more promising in that regard, as it seems to reveal that if the ventromedial prefrontal cortex is damaged, then the patient's moral cognition seems to suffer as well. However, whilst this does suggest a causal relation of sorts, this is not to say that emotion is necessarily the exclusive source of moral judgement. For example, one possible explanation is that emotion is an important component of moral decision making, because morality applies to how we treat others in a social context and this is also an evolutionary function of emotion. This at least sounds plausible. In an evolutionary context, much of our survival has been down to how well we interacted with each other in a group. In this context, it stands to reason then that the application of moral judgements is something we ought to care deeply about. One issue with this explanation, however, is that it does not tell us anything specific about the application of moral **language**, which from a philosophical point of view is what we really want to focus on; this is something I will address further down. Nevertheless, I argue that Young and Koenigs work is a great stepping stone into the world of relationships between emotion and what we might recognise as 'morality'.

This recognition is also very important, because the matter of what constitutes morality is a very controversial issue. For the sake of argument, let us take altruism as an example. We can look for and often find correlations between seemingly altruistic behaviour and strong emotions. In fact, we can see this not only in people but in certain animals as well,

predominantly those that live in social groups. For example, in 2012 a BBC/National Geographic film crew for the series Planet Earth took incredible footage of a previously unseen phenomenon. Whilst filming an orca attack on a grey whale mother and her calf around Monterey Bay, two humpback whales were seen 'intervening' on the hunt. Humpback whales are known for making a high pitched trumpeting sound when agitated. The crew observed the humpbacks trumpeting, diving and violently slapping their pectoral fins against the water. According to observer Victoria Bromley "It didn't seem at all like they were confused… they were definitely there with a purpose," (as seen in Davis, 2012). Nevertheless, the attack resulted in the killer whales successfully killing the grey whale calf, after which all of the larger whales, including the mother grey whale fled the scene. After the attack, however, the humpback whales were seen moving back into the area where the calf was last seen, trumpeting and rolling in the water and aggressively swiping their tails at any orcas that approached. According to observers, "the whales [we watched] should have been off feeding: instead, they deliberately stayed in our area, loudly announcing their presence." (Davis, 2012) "This was not a curious approach by these humpback whales: they seemed truly distressed." (Davis, 2012) The observations have since been backed up by more academic studies into the behaviour of humpback whales, for example by zoologists like Robert L. Pitman, Volker B. Deecke, Christine M. Gabriele and others (Pitman et al, 2016, 7). The research in this study strongly suggests a conclusion that these instances of humpbacks interfering in orca attacks are examples of interspecies altruism. Other possible explanations include kin selection theory and reciprocity, both of which are lacking in direct evidence in observations of the behaviour of humpback whales. It is clear that humpback whales can be highly emotional before, during and after performing apparent altruism, but the still unanswered question is this: was this apparent altruism **caused** by the whales' emotions?

The only way we could truly answer the question of whether or not there is a causal relation between emotion and moral decision making and to what extent it plays a role in moral language is if we were to look at a case or cases where a subject is entirely devoid of emotion, and then see whether or not they are even capable of making what they or anyone would comprehend as a moral decision. This is difficult to imagine let alone find an example of. However, perhaps something very close to this can be found in some particularly severe cases of clinical depression. In the most extreme cases of depression, it is not uncommon for a patient to experience a very high degree of emotional detachment. Similarly, some patients with mild to moderate levels of depression who have been treated with SSRIs (selective

serotonin reuptake inhibitors) have experienced similar periods of emotional detachment albeit to a lesser extent (in fact this is something I have first-hand experience of). The phenomenon is not particularly well understood, but the symptoms generally appear to consist of the patient experiencing an impaired ability to experience strong emotions both positive and negative.

Jonathan Price, Victoria Cole and Guy M. Goodwin talk about precisely this phenomenon. Their paper details a qualitative study gathering data through individual reviews, a group interview and validation interviews, which concluded that emotional side effects from SSRIs are "a robust phenomenon, prominent in some people's thoughts about their medication, having a demonstrable impact on their functioning and playing a role in their decision-making about antidepressant adherence." (Price et al, 2009, 211) During the interviews, eight key framework themes were identified. The general effects on all emotions, reduction of positive emotions, reduction of negative emotions, emotional detachment, 'just not caring', changes in personality, effects on everyday life (helpful or unhelpful), and the final theme being 'it's because of my pills'.

On the general effects on all emotions, the study had this to say:

> Most participants described a general reduction in the intensity of all the emotions that they experienced, so that all their emotions felt flattened or evened out, and their emotional responses to all events were toned down in some way. Very common descriptions of this phenomenon included feelings of emotions being '*dulled', 'numbed'*, ' *flattened'* or completely '*blocked'*, as well as descriptions of feeling '*blank*' and ' *flat*'. A few participants described a more extreme phenomenon, in which they did not experience any emotions at all. (Price et al, 2009, 213).

Almost all of the patients described a reduction of positive emotions to some extent (i.e. happiness, enjoyment, excitement, passion, affection, enthusiasm etc). "Many participants described reduced enjoyment of, for example, social situations, hobbies or interests, beauty and nature, and music and other emotional media." (Price et al, 2009, 213) All of the patients also described a reduction in the intensity of frequency of negative emotions also, which they attributed to the SSRI antidepressant:

> Although a reduction in these negative emotions was usually at some stage a benefit or relief, for many participants it had become an unwanted side-effect, impairing their quality of life. Participants described the need to be able to feel negative emotions when appropriate, such as grief or concern. Some were unable to respond with negative emotions, such as being unable to cry when this would have been appropriate or respond appropriately to bad news. (Price et al, 2009, 213).

Most patients described emotional detachment manifesting in a general detachment from other people and from social interaction. Specifically, they felt reduced **sympathy and empathy** (Price et al, 2009, 213). This also seems familiar to my own experience of SSRIs, which I suppose affords me the opportunity to ask some relevant questions here. Did I feel like my capacity to make what I would now recognise as moral decisions was affected whilst in this emotionally detached state? More importantly, was it noticeably more difficult to convey ethical ideas through moral language? One would think that these would be simple enough questions to answer, but unfortunately they are not, even to one who has experienced this, but if I were to come up with either a yes or no answer then I would probably say 'yes'. The main difficulty in answering this question lies in the realisation that during these periods of emotional detachment, matters of moral decision making did not seem overly relevant at the time for the simple reason that I was rarely if ever interacting with other people in social activities, and so there was little to no pressure to adhere to any moral principles of any kind. All thoughts seemed to be those that referred to what were the most rational courses of action when generally acting around the house on my own e.g. 'I 'should' pre-heat the oven before putting the bread in'. On the rare occasion that I did interact with other people, I was noticeably more polite than usual. However, this politeness did not come from a desire to be so based on the idea that it was 'the right thing to do', but rather that it was a 'sensible' thing to do in order to make the conversation go more quickly. In fact, things being 'the right thing to do' were not generally a relevant concern at the time, rather what was the 'smart' thing to do. Whilst there was a lot to be said for this rational approach to daily life, I ultimately do not regret choosing to stop taking SSRIs and instead treat my depression using a different kind of antidepressant, an NaSSA drug called Mirtazapine, which focuses primarily on raising serotonin levels via treating my insomnia; I found this far more effective in my particular case, though it is important to remember that the best treatments vary on a case to case basis depending on the individual.

I believe this personal testimony of mine bears some resemblance to the framework themes described in the 2009 study. The fourth framework theme described involves patients describing a state of 'not caring' (not feeling overly strong emotions about any one thing in particular). Interestingly, this is described as having both positive and negative results in hindsight (Price et al, 2009, 214). A positive consequence of this is that it reduced the pressure on some patients to perform activities to the impossibly high standards that they set themselves, thus relieving a great deal of stress that would otherwise be experienced. Many patients described their previous depressive state as being rooted in a kind of unrealistic and unhelpful or even pathological perfectionism, inevitably leading to self-criticism and bitter disappointment upon discovery of one's own natural and inevitable imperfection. However, they also felt less motivated to perform well in tasks that they previously excelled at. They also felt less motivated to look after themselves physically as well as they used to. "Some participants felt that their sensible, safety-conscious, side had diminished and they just did not care as much about the consequences to themselves of their behaviour." (Price et al, 2009, 214). In some extreme cases, this went a stage further and started to affect the patients in dramatically negative ways. "A few participants went further, mentioning thoughts of self-harm or suicide that they related, at least in part, to feelings of emotional detachment and emotional numbness. One participant had started to self-harm in an effort to feel emotion [even if that emotion was negative]." (Price et al, 2009, 214). This last case is particularly interesting, as it implies that there is at least some means of motivation for behaviour other than emotion, given that the motivation is directly targeted at feeling an, or **any**, emotion in the first place (I will return to this idea a little in my next chapter).

Without spending too much time going through all of the remaining framework themes, let me instead sum up by concluding that the 2009 study shows us a fairly close example of what an individual functioning with little to no emotional attachment looks like. Now for the important questions: do these cases also display an inability to function in what we would normally recognise as moral behaviour? Also, do these cases display an inability to utilize and/or recognise moral language? Referring again to my own testimony regarding moral language, I do not recall so much as even making use of it when emotionally detached. This was because my perception of everyday life did not seem to manifest in the context of what was 'right' or 'wrong' but rather in terms of what was 'sensible' or 'logical' and what was clearly 'illogical'. Ironically, this did allow me to function as a moral agent of sorts. It simply did not occur to me to carry out overtly cruel acts, as they were generally irrational and/or

pointless and thus not worth doing. However, if I care to think back in hindsight, I realise that non-reciprocal altruism also eluded me at the time. I was, however, able to recognise the moral language of others as moral language, I simply did not utilise it myself. The best way I can think to describe it is that I had little to no moral sentiments, and so I had no reason to convey emotional concepts through moral language. I imagine the experience would be a lot different to someone who had never known strong emotion unlike me who at least had the memory of strong emotion to conceptualise and recognise the meanings behind emotional terminology. However, just imagining a subject who has never experienced an emotion speaking in terms of what is right and wrong in a way that would have the same meaning to you or me is difficult. Regarding the 2009 study, this subject of the effects of SSRIs on specifically moral language is sadly not really touched upon. However, what is a recurring theme in the study is the idea that these patients had a dramatically reduced sense of **empathy** (Price et al, 2009, 213). That is, a reduced **sense** of the feelings of others.

So how can we attach this connection with emotion into the moral language of the expressive fictionalist for example? First, let us look at what an emotion is. This is, unfortunately, frustratingly difficult. Describing emotions in a non-circular way is difficult in and of itself. For example, describing it as a subjective 'feeling' may be unhelpful here. Instead, let us assume for the sake of argument then that an emotion is best described as a kind of alarm bell or warning signal that there is some issue relevant to the subject's survival and/or wellbeing that they would want to take notice of if such things are important to them, as some in evolutionary theories describe it. For instance, suppose we see emotions as Paul Ekman describes, for example the emotion **fear** as related to the subject's awareness and interpretation of potential dangers in their environment. Similarly, we see **disgust** as related to the subject's self defence against contagious diseases, **anger** as related to the subject's motivations to solve problems and/or remove threats quickly and decisively, **sadness** as related to preventing the subject from continuing behaviour that is ultimately unhelpful and/or a hindrance to more productive goals, and finally, **joy** as related to reinforcing behaviour that is helpful and/or beneficial to the subject's health and/or wellbeing (Ekman, 1992, 169)[36]. Most basic emotions seem to stem from these archetypes, for example terror a more extreme form of fear, amusement a more specific form of joy, embarrassment a very

[36] If this sounds familiar, that could be because Ekman was a leading psychological consultant on the 2015 Disney Pixar film *Inside Out*; there are also other 'primary' emotions Ekman describes that do not feature in the film, such as **surprise**.

specific form of sadness (possibly with connotations of surprise as well) etc. Let us also assume for the sake of argument that not only do we have an immediate and very direct sense of our own emotions, but we also have an observable sense of other people's emotions as well: empathy. This is equally useful, because our own individual survival and/or wellbeing **when coexisting in a social group** (e.g. a tribe, a nation or even a world), is intrinsically tied to and in no small way dependent on the survival and/or wellbeing of others. If we combine this with the idea that empathy is somehow connected to what we would recognise as moral concepts (e.g. altruism, i.e. that which relates to our treatment of others), then we can begin to see how a fictionalist may assimilate some of this into their own language (this in particular will help us out a great deal in the later stages of chapter five). First, let us remind ourselves of how a fictionalist describes moral concepts in terms of language. Assuming we are going with Joyce's pretence assertion theory over a tacit operator (Joyce, 2011, 291), then the sentence "kicking cats is wrong", is a pretence assertion, semantically meaning what it says, but in fictional context. Therefore, it not really asserting anything whatsoever, and as stated previously, this raises the question of why the fictionalist would bother to do this. To answer that question let us look at what might be going through the fictionalist's mind when he/she makes such an utterance, and proceed to build our theory from there.

Chapter 4: From Science to Language

My arguments in the preceding chapter(s) should, I hope, suffice to establish that there is a link between emotions and moral attitudes, for instance those of approval and disapproval. If we accept that that is the case, the question that now arises is how this is communicated through moral **discourse**. One possible way, which I have explored extensively in my previous two sections, is to express attitudes of approval and disapproval semantically. However, as we have seen, this explanation is vulnerable to the Frege-Geach problem where we lack a convincing explanation of how moral language can serve an expressive function in unasserted contexts, and it takes very elaborate and ambitious projects, such as quasi-realism, to even try to circumvent that difficulty. However, I argue that the Frege-Geach problem is only a difficulty for expressivism so long as we interpret semantic **content** as expressive rather than assertoric. In other words, we can still be expressivists in terms of our interpretation of the communicative act of moral discourse whilst using alternative approaches to fill in the content semantically i.e. moral fictionalism. However, if we are to marry the two concepts together, then we surely need some means of answering how and why non-realists may use fictionalised moral terminology to assert, or pretend to assert, fictionalised moral phenomena to express emotions and attitudes, such as approval and disapproval. In the following chapters, I will explore how various emotions function in relation to moral cognition, and then moral language. I will do this by introducing two separate models in which emotions may be translated into moral language. First, I will discuss what I call 'the pride, shame, and responsibility model', and later on 'the empathy model'. These two models may even stand as alternatives to each other, yet I would argue that more often than not the two are very closely related, for they are neither mutually exclusive nor even averse to working in parallel with each other, and more often than not they do just that.

First let us discuss the pride, shame and responsibility model. One possible inspiration we may take is from Jesse J. Prinz's 2007 work *The Emotional Construction of Morals*. Prinz, for example, also argues, as Gilbert Harman does and as I do, that morality originates not from the world, but from the psychological, non-moral capacities of living beings:

> We don't have a clothing module; we just get cold and are clever enough to solve that problem by making things to cover our bodies. We don't have a religion module; we just have tendencies to attribute mentality to non-living things and a penchant for stories that violate expectations (Boyer and Ramble, 2001). My conjecture is that morality falls into this category. It is a byproduct of capacities that are not themselves evolved for the acquisition of moral rules. (Prinz, 2007, 270)

This is, ironically enough, not unlike Blackburn and the quasi-realist discourse on projectivsm, whereby non-moral cognitive processes are projected onto the world not the other way around, and where morality is treated as a **construct** rather than a property. Prinz argues that emotions help to form the cognitive processes necessary for moral attitudes to arise. He discusses his theories on how these cognitive processes manifest and the variety of complex systems this involves:

> First and foremost, we need to have certain emotions: other-directed emotions such as anger, contempt, and disgust, as well as self-directed emotions such as shame and guilt, which, I argued earlier, may be related to embarrassment and sadness. Second, one needs the ability to formulate rules. During moral development, we must *transfer* the negative emotions that we are conditioned to experience when we misbehave *to* the misbehaviour. In so doing, we generate a mental representation that disposes us to have negative feelings about a type of behaviour regardless of who is performing it. (Prinz, 2007, 270)

So here Prinz is dividing emotions into at least two categories, those directed at others and those directed at oneself. He also raises a very important point about rule formation. Systems of morality often involve far more than mere approval or disapproval; moral **rules** are an integral component of how morality functions in everyday life. The concept of 'justice', for example, is embedded with layers of additional abstract concepts, such as discipline, punishment, accountability and responsibility. More importantly (for us), these layers of conceptualisation are, more often than not, infused with more emotions like pride, shame and guilt.

For the sake of argument, let us define pride as a feeling of positive self-esteem, either in the form of a general admiration for the self or of a particular set of actions, attitudes or

characteristics perpetrated by oneself. This is not to be confused with another, very similar phenomenon i.e. arrogance, which we may describe as concerning the perceived image of the self **in relation to others** (in this case considering oneself as somehow 'better' than others). To clarify, where we talk about pride (in the context of the pride, shame and responsibility model), we are referring to the emotion/feeling of positive self-esteem, rather than how it relates to how others are perceived in comparison. Now let us interpret shame to be a feeling of negative self criticism, and so in many ways we may see it as an opposite to pride. Much like pride, shame may be targeted at any number of things, from physical appearance to particular actions, attitudes or other feelings one may have in relation to something.

Earlier on Prinz also describes something called 'meta-emotions', this will be something which will come to be rather important later on. According to Prinz, "we don't just value actions; we also value the valuing of those actions" (Prinz, 2007, 120). In other words, like Blackburn, Prinz portrays emotions and attitudes as multi layered, whereby emotions and attitudes may be directed at other emotions and attitudes. According to Prinz, "If your values disagree with mine, I will have a negative response to both your behaviour and your values. The former may lead me to try to alter your behaviour directly, but my distaste for your values will require that I try to alter how you feel" (Prinz, 2007, 120). So in essence a meta-emotion is simply an emotional reaction to an emotional reaction. How this plays out, however, is very interesting, because what appears to happen is that emotional reactions take place not only in response to certain behaviours, but in response to others' emotions (or lack thereof) in response to those behaviours. Later on, he describes something else, this being meta-punishment (Prinz, 2007, 271). This involves not only punishing those who flout moral rules, but also those who allow this to happen. Thus, meta-emotions like disdain or outrage are targeted not at the original immoral act itself (exclusively), but at wilful apathy in response to that act also (Prinz, 2007, 271). These two concepts will help us out a great deal as we begin to unravel how this leads into the pride, shame and responsibility model.

As mentioned earlier, pride and shame are, as Prinz would describe them, self-directed emotions. Now let us look at how these kinds of sentiments manifest in other-directed emotions like **anger** or **disgust**. Prinz, for example, proposes a form of social punishment system which he refers to as 'love withdrawal':

> If you want to train people to be good at withdrawing love, you *could* train them just to behave as if they were withdrawing love, *or* you could train them actually to withdraw love. In short, you could train people to be disappointed (or angry or disgust etc.) when other people misbehave. The latter strategy seems more direct, and hence more efficacious. If I am right, then the best way to implement meta-punishment is to show disappointment (or anger or disgust etc.) at those who do not react to misdeeds with negative emotions. When we do that, we effectively condition people to feel badly about not feeling badly about bad behaviour. Thus, meta-emotion is a form of meta-punishment because it assigns an emotional cost to anyone who fails to have punitive emotions toward those who transgress. (Prinz, 2007, 271)

So already, we can see how one possible model for translating emotions into moral language arises, one based on the social positioning of one in relation to a perceived transgression, and how one's reaction is perceived by others in a social environment. Therefore, the emotional need for the individual to exist harmoniously within one's social environment carries with it certain behavioural imperatives; these include active efforts to avoid violation of behaviours others deem less than acceptable. Failure to conform to these collective standards carries with it other emotions, such as the fear of ridicule, abandonment and/or punishment. In essence, the emotional, social, and perhaps even the physical health of the individual depends significantly on their ability to manage through a complex web of emotional states working simultaneously in conjunction with each other, ranging from pride of one's ability to conform to the social requirements, reinforcing 'good' behaviour, to shame or guilt at one's failure to do so, to then applying the expectation of such emotions in others, then applying anger or disgust towards any apathy displayed by others towards their own need to conform, we may then refer to this as their **responsibility** to do so.

In my next chapter, I will examine this model in greater detail and then proceed further by discussing how these emotions may be communicated through fictional utterances, not only expressing pride and shame, but also through other phenomena such as empathy and compassion. With this aim in mind, I would do well to set a few 'sub aims' and key points here:

1) The ultimate aim is to depict a non-realist account of moral language---i.e. one which does **not** rely on any realist metaphysics or epistemology but rather is an account

stemming purely from the philosophy of moral language (specifically within the realms of semantics and pragmatics).

2) The account depicts a fictionalist account of moral semantics---i.e. one which depicts a self-aware fictionalist and is a revisionary account rather than a hermeneutic account (in other words we are describing the language of speakers who have distinctly non-realist/fictionalist philosophical perspectives on moral language, rather than describing the everyday moral language of everyday persons, whom we will presume are either moral realists or are morally agnostic on the issue).
3) Like expressivism and/or quasi-realism, the account depicts a means and/or system of communication whereby information about the speaker's emotion(s) is conveyed through moral discourse. However, unlike regular expressivism, this communication of information is conveyed pragmatically rather than being conveyed within the semantics of the sentence(s) uttered.

To sum up, the aim is relatively simple: to depict a brand of moral fictionalism that also has a story to tell, as expressivism or quasi-realism does, about how information about speakers emotions are conveyed through moral discourse and which is loyally non-realist.

To recap, let us first go over what we have achieved thus far. The first goal was to establish at the very least a plausible case for a non-realist account in the first place. To that end I discussed moral realism (as well as something as dauntingly vast as this can be within a single chapter) as well as some of the most notable counter arguments to it, specifically those of Gilbert Harman and J.L. Mackie. The key point here was and is not to declare in any way that moral realism is false (justifying such a declaration convincingly would no doubt require an entire thesis in and of itself) but rather to simply point out that there is **a case to be made** for a convincing non-realist account. The next stage was to examine what the non-realist is literally saying (and meaning) when utilizing moral language, in other words what is going on semantically? This led me initially to expressivism, at which point we were confronted with the Frege-Geach problem (how can meaning be derived from expressivist utterances in unasserted context(s)?). In seeking a solution to this problem, quasi-realism appeared to be an attractive option promising to deliver everything that a moral realist could wish for whilst still retaining a non-realist status, without committing to any metaphysical arguments, and thus being significantly more economical in its approach. However, whilst this metaphysical economy was indeed attractive, it soon became apparent that quasi-realism was more than

expensive enough in other regards. It also became apparent that there seemed to be a lack of any **clear** distinction between what was being said by Simon Blackburn and what was being said by moral fictionalists, which led us to the debate between Blackburn and David Lewis, who cited this issue. Whilst Blackburn appeared to successfully sever all ties to fictionalism, he appeared to do so at the expense of the crucial, non-realist aspect of quasi-realism.

Moral fictionalism thus became the new prime suspect for a convincing non-realist semantic account of moral language. Fictionalism had both a relatively straight forward story to tell, and one which seemed immune to the ravages of the Frege-Geach problem. In comparison, quasi-realism only had this by no small degree of controversy. Fictionalism, however, lacked the same promises made by quasi-realism, which was a story about the emotional attitudes behind moral language and the communication of those emotions. Nevertheless, it seemed that I had a convincing candidate for my first two goals, and thus the hunt was now on to fulfil the third and final goal by looking into the pragmatics of moral discourse. To do this it was important that I establish a clear and evident connection between emotions and moral language, and so I decided to take a scientific approach in doing so, concluding that whilst I am not convinced of a 'necessary' i.e. logically deductible connection, there seemed enough observable evidence to make a convincing empirical case in this regard. So, finally, this brings us to the last piece of the puzzle, which is to marry this connection to the semantic fictionalist account in a way that makes some philosophical sense.

Chapter 5: Translating Emotions into Moral Language

In chapter four I discussed the self-directed emotions (as Prinz would call them) of pride and shame, as well as some other-directed emotions like anger or disgust. However, I would like to briefly refine Prinz's usage of these emotions before we proceed. This is because I would argue that anger and disgust are not **exclusively** other-directed nor are pride and shame exclusively self-directed. Embarrassment for example may be seen as a highly nuanced and entirely self-directed combination of a whole host of negative feelings, including disgust or anger, with connotations of sadness or even fear as well. In fact, embarrassment is, I would argue, one of the most difficult emotions to actually describe. It is also, I would argue, an essential ingredient for shame, shame being a very specific form of embarrassment manifesting in a realisation of one's failure to live up to a particular standard, directly impacting self-esteem and potentially social status. It is for these reasons that I have referred to this model as one of pride and shame primarily, these being more complex emotions made up of other more basic emotions like anger or disgust or joy, which ultimately fall into and help to constitute one or the other. That said, however, it is still useful to discuss other emotions and their role in doing so. Like Prinz I would also separate these emotions into self-directed and other-directed, but in a slightly different way. I would argue that both pride and shame can just as easily be other-directed as self-directed, we can be proud of someone we have taught, and we can be ashamed of someone who has let us down.

Now let us see how these two mental states (of pride and shame) may work in conjunction with a constructed morality system. As mentioned in previous chapters, it is very important, even for the non-realist, to actively participate in the morality **narrative**. This may be characterised by a collective understanding of actively pursuing things which make us feel proud and avoiding things which make us feel ashamed. Unfortunately, I would argue that when it comes to morality, anxiety about other people's opinions have a lot to do with this, thus there is certainly a level of insecurity attributed to it. This is not always the case, and I would certainly argue that it is entirely possible to derive moral pride from an action without being overly concerned about how that action may be perceived by others. There may even be occasions where we may perform an act of kindness or generosity only to cover it up, because we do not want to be **seen** to be helping that person for whatever external reason. In such cases, the pride we feel is derived from us and us alone, in the sense that we realize that we have done an admirable thing, but do not feel the need to showcase this to others; it is

enough for us to recognise that we are behaving like the people we want to be, and are, therefore, participating in the morality fiction. However, I would, sadly, argue that this is not the majority of cases, even among realists, and that a considerable amount of our acts of kindness and generosity derives largely from fear and insecurity about how we appear to other people. In other words, it is not enough to actively participate in the morality narrative; we must be **seen** to be participating. I would also argue that this has a great deal to do with anxiety about social status, not to mention the consequences of breaking certain laws.

Prinz argues that whilst meta-emotions may not be necessary for such moral senses to exist, and an example he gives is how individuals with autism still moralize despite empathy impairment(s), meta-emotions nonetheless "...play a fundamental role in the emergence of morality through cultural evolution" (Prinz, 2007, 270).

> Using game-theoretic models, Henrich and Boyd (2001) have argued that widespread cooperative behaviour would not be sustained and transmitted over generations if it were not for punishment; a society of co-operators would be over taken by defectors. The same point can be articulated in terms of moral norms; a society would not continue to conform to moral norms generation after generation if wrongdoers were not punished (Stripada and Stich, 2006). (Prinz, 2007, 270)

This is useful, because it reminds us that merely speaking out against an act is one action towards preventing its continuation; another is creating a genuine emotional **cost** to the perpetrator. Now let us picture this in a fictionalist context. The morality narrative, as described by the individual's culture and/or upbringing, outlines certain actions, attitudes and characteristics, some of which the individual is actively encouraged to pursue, whilst others they are actively discouraged from or even actively prevented from doing. Over time, the individual associates so called 'correct' actions, attitudes and characteristics with positive feeling whilst evaluating oneself, manifesting in positive self-esteem upon the realization that their self image matches what they have been brought up to perceive as 'good' (as described by the narrative). We may call this state 'Pride'. This positive association then encourages the individual to reinforce this by continuing to behave in that way. As a result, pride becomes both a self-directed and other-directed emotion driving both a self determination to uphold a certain standard of behaviour as well as to encourage that standard in others. At the same time, the individual also associates the 'incorrect' actions, attitudes and characteristics (as

described by the narrative), with experiences of negative feeling, and in the form of self criticism upon the realization that at least part of their self image matches those, which then discourages them from continuing to behave in that way. Again, this goes both inwards and outwards, driving the individual away from actions they know will induce a sense of shame, and actively discouraging others from those actions as well. Once either pride or shame has been established, this will then inevitably affect how the individual presents the action, attitude or characteristics to others. For example, if someone is proud of a particular action, they may enthusiastically tell others all about it, knowing that this is likely to inspire positive feedback from others i.e. **praise**. If someone is ashamed of a particular attitude, they may actively try to cover it up from others, knowing that if others were to find out about it, then this would incur negative criticism i.e. **blame**.

Now let us look at some potential examples of speech acts utilising these kinds of cognitive processes.

- "I couldn't have lived with myself if I had let that happen" (Self-directed pride reflecting positively on one's awareness of what the moral narrative was, and what was necessary to conform to it).
- "I know you'll do the right thing" (Other-directed pride reflecting an individual appealing to what they see as others' ability to recognise the moral narrative and what is necessary to conform to it).
- "How could you?" (Other-directed shame reflecting disbelief and bitter disappointment in another's failure to conform to the moral narrative, resulting in a kind of love withdrawal (as Prinz would describe it)).
- "I'm so sorry; please forgive me" (Self-directed shame reflecting one's own failure to conform to the morality narrative, and an appeal for a chance to repair the damage to their social status by redeeming themselves and earning the other's trust/love once more).

It is important to remember here that as mechanical and potentially even manipulative as this appears, the majority of the reasoning would appear to take the form of unconscious inferences (something we will return to later on). From the speaker's point of view, the reasoning at work behind the emotions driving these kinds of speech acts may not be

immediately apparent only the fact that they **feel** proud or ashamed. It is also important not to underestimate the efficacy of the fear of punishment, particularly when it manifests in social abandonment. The need to 'fit in' in one's social environment is not a trivial matter; to the contrary it has been essential for our very survival for millions of years. To be seen to conform and contribute (on a moral level) is, therefore, absolutely essential for the non-realist just as it is for the realist.

Pride and shame clearly play a fundamental role in the formation of moral rules also, acting as emotional motivators compelling either the individual, or a collective of individuals, to construct axioms and maxims to adhere to based on what leads to pride and what leads to shame, and encourage others to do the same. At this point, we can apply some good old fashioned projectivism, as this can then be expanded on by **projecting** these same axioms ('Do not steal', 'Be polite', 'Care about others' feelings' etc), onto the world and everyone in it. Referring back to Joyce and the pragmatic motivations behind fictionalism, we may recall how convincing oneself that something bad **will** happen if we don't do x amount of sit-ups a day, or if we smoke too many cigarettes, is more efficacious than more realistically realising that it merely **might** (Joyce, 2011, 303). In a similar vein, the expressive fictionalist can do something similar by projecting a kind of cosmic imperative to conform to the morality narrative and encourage others to do so, acting and speaking 'as if' the universe will find some means of punishing them should they fail to do this. This may be exaggerated for motivational purposes, or it may be based in a genuine concern regarding certain actions and the consequences of acting them out. These consequences may take the form of legal punishment, social abandonment or both. This even helps us to account for moral **disagreements**, as different individuals and different cultures arrive at different axioms and maxims via different circumstances, different attitudes and different unconscious inferences. As Prinz and Nichols put it:

> When emotions are conditioned in the context of behaviour, sentiments are formed and affect-backed rules result (compare Nichols, 2004a). The business about punishment and meta-punishment shows that these relatively simple resources are quite powerful. In principle, any form of behaviour could be subjected to a form of emotional conditioning that would result in the formation of a moral rule (Sripada and Stich, 2006). (Prinz, 2007, 271)

We may call this overarching drive to uphold a moral maxim, be it individual based or collectively based: 'moral responsibility'. We may call punishment meant to either deter or correct transgressions 'justice'. This is not to be confused with 'legal' justice, however. It is important here that we treat morality and law as two different entities to avoid unnecessary confusion. One (morality) we are treating in the context of a constructed, psychological narrative, for the purposes of expressive communication. The other (law) is constructed by political and managerial institutions for the purpose of **enforcing** standards of behaviour necessary for society to function and ensure the safety and welling of its citizens. There is considerable overlap of course. In most cultures, both current and historical, serious offenses such as, theft, rape and unsanctioned violence have been treated as both immoral **and** illegal, and there are some things which nearly everyone agrees are both socially unacceptable **and** actively destructive to a society. However, that which is illegal is not necessarily seen as immoral, and that which is seen as immoral is not necessarily illegal. For instance, it is not illegal to be lazy or impolite, but neither is it seen as 'good'.

With this in mind, let us now look at some more potential utterances made by the expressive fictionalist:

- "Something ought to be done about this!" (Other-directed shame (i.e. disappointment in the establishment) appealing to the social collective to apply some measure against others who have failed to fulfil their responsibilities).
- "I think that I've earned a rest." (Self-directed pride reflecting an appreciation of one's ability to fulfil one's responsibilities).
- "I should've tried harder to stop it!" (Self-directed shame reflecting disappointment in one's ability to fulfil one's responsibilities).
- "Congratulations, you deserve this." (Other-directed pride reflecting admiration for another's ability to fulfil one's responsibilities.)

As we can see here, the expressive fictionalist, by participating in the morality narrative, once again uses speech acts to respond to circumstances in either a preventative or reinforcing manner, just as we will see in the next chapter in the empathy model. In this case, the expressive fictionalist utilises moral responsibility as a motivator, providing emotional benefits for some behaviours and costs to others.

So let us recap on what we have here. First, we have pride and shame, both self-directed and other-directed; this provides our base layer. Next, we have these emotions applied through social conditioning. This may take on many forms, for example, through positive feedback for behaviours conforming to the morality narrative, or criticism for behaviours which abandon it. Next, we (the expressive fictionalist) reinforce this conditioning through systems of reward and punishment, applying genuine physical or emotional costs and benefits to our actions. Next, we combine these elements together, adding them into our own morality narrative. Finally, we communicate this narrative, within the fiction itself, thereby participating in it. In so doing, we communicate far more than mere approval and disapproval, but a whole range of emotional information. In other words, we convey a great deal more about who we are, and how we feel about certain kinds of behaviour, than it says anything objective about the behaviour itself. One may shout, and scream and tell another that they have a responsibility to those around them, or express their pride in someone for doing the right thing. This does not **really** tell us much about the ultimate nature of responsibility or the nature of real moral properties metaphysically, but it does tell us a great deal about that person (who is speaking) and the kind of person that they are. If A tells B that B has a responsibility to treat others with respect, then this tells us a lot about A, for example that they value politeness and social grace, and that they disapprove of people who hurt others' feelings or slander others' reputations. If A expresses pride in B for owning up to stealing a £10 note, this tells us A values honesty and integrity, disapproves of theft and respects those courageous enough to own up to their own transgressions and attempt to make amends. One simple moral utterance, whilst conveying little about the ultimate nature of morality, conveys volumes by way of emotional communication, **regardless** of what the ultimate nature of morality is or is not.

This is, of course, all rather dependent on whether or not A is sincere in what he/she is saying, because the real question is why or even how a non-realist experiences genuine pride or shame about something that they do not **actively** believe in i.e. morality. In other words, given that we are ultimately treating morality as fictional rather than real (ontologically), then where does this captivating drive to conform to this morality narrative come from? Is all this just an elaborate form of psychological manipulation, for the expressive fictionalist to get what they want out of other people? This does not seem at all a desirable conclusion to me. Fortunately, fictionalists like Joyce would seem to think not, and I am inclined to agree.

Critical context plays a vital role in why we both think this; this will be the subject of great discussion in chapter six.

Before this, however, there is a great deal more to talk about in regards to translating emotions into moral language. So far I have outlined one model for doing so based on pride, shame and responsibility. However, this is by no means the end of our story, because the pride, shame and responsibility model may also lead into other models as well, in which emotions like anger or disgust, pride or shame are expressed by way of other psychological phenomena. In the following two sub chapters, I will outline a detailed process by which this may occur. I have called this the empathy model.

Chapter 5.1 Empathy & Compassion

Over the course of the last two chapters, I have explained one possible means by which emotions may be translated into moral language. In the pride, shame, and responsibility model, emotions like pride and shame play a social conditioning role in the individual to avoid certain behaviours whilst reinforcing others by applying positive feedback to those who conform to the morality narrative, and applying negative criticism to those who do not. In many cases, this may be taken much further by applying emotional costs to those who severely flout the rules of the morality narrative, including social rejection and love withdrawal. The real question for us, however, is how this actually manifests in moral language. Earlier we looked at particular sentences, such as 'I know you will do the right thing' or 'How could you do this to us?'. These **express** emotion states like pride or shame, anger or disgust by projecting those emotions through moral utterance. I would argue, however, that we can take this a stage further, and that we can even outline a **process** by which these emotions are communicated. One means by which we can do this is by exploring psychological phenomena such as empathy or compassion. Empathy, for example, can be treated as a means to having appropriate moral emotions: one may empathise with the victim of perceived suffering, wincing in response. As a result, one may be compelled to feel pity or compassion for them, anger towards the assailant, indignation at their breaching the morality narrative. As such, the mere witnessing of an act perceived as admirable or obscene, as conditioned via the pride, shame and responsibility system, evokes powerful emotions which are then expressed through moral utterance. Over the course of this chapter, I discuss empathy and compassion, and their role in moral discourse. I will then outline, in detail, a ten stage inferential process, beginning with observation, and ending in preventative or reinforcing speech acts.

First, let us examine the study of empathy historically, thereby (hopefully) shedding light on what it is that we are talking about here. The history of the study of empathy is not particularly easy to track, as words with similar meanings have been often hinted at and touched upon ever since the days of Socrates, Plato and Aristotle[37]. It is far more recently, however, that dedicated philosophical research into the specific word and the specific concept

[37] The original source of the modern word 'empathy' (prior to its etymology from the German word 'Einfühlung') is the ancient Greek word εμπάθεια (*empatheia*, meaning "physical affection or passion"). The German word we see later was adapted from this (Gallese, 2003, 171).

has taken place. The word ‘empathy’ most recently originates from a translation of the German word “Einfühlung” meaning (‘feeling into’), as translated by psychologist Edward Titchener in 1909 (as seen in Stueber, 2013). This was following a specific concept brought to light by Theodore Lipps around the same period in his works on the philosophy of aesthetics in which Lipps frequently talked about our ability to read into and feel into various forms of artistic expression (Curtis & Elliot, 2014, 353). Lipps, however, took the concept much further by pitching empathy as an epistemic phenomenon (Curtis & Elliot, 2014, 376). This in turn led to no small measure of debate throughout the early 20th century, largely from Lipps himself; about how exactly our perceptual analysis of other minds takes place, as well as reconciling troubling questions such as Mill’s inference from analogy and the problem of other minds (Stueber, 2013), the former of which is particularly important. The inference from analogy describes the steps taken that enable us to attribute mental states to others by observation and from direct experience of their behaviour (Stueber, 2013). This generally presupposes a Cartesian perspective on the mind, in which we have direct access to our consciousness and which is infallible, and in which knowledge of other minds is indirect and fallible (Stueber, 2013).

The latter of the two examples, the problem of other minds, as Wittgenstein pointed out later on (Wittgenstein 1953, 350), also depends heavily on a Cartesian account of the mind (i.e. knowledge of one’s own mind is self-evident, but knowledge of other minds is suspect). The problem of other minds asks us how we can have knowledge that other people have minds. Empathy understood as the primary epistemic means for receiving information about other people’s minds was revived in the 1980s as well by simulation theorists like Davis and Stone (Davis, 1997, 144-168). Neuroscientists have also held discussions on subjects like mirror neurons, providing empirical evidence for Lipps’ theories on what he calls ‘inner imitation’ (i.e. the mind’s ability to imitate or ‘mirror’ others’ emotions) (e.g. Gallese and Rizzolatti, 2004, 396-403).

In terms of the study of empathy in specific relation to moral philosophy and moral psychology, the works of psychologists Batson and Hoffman are of particular fame. Specifically, Batson devised a series of experiments to test what he calls the empathy/sympathy thesis (Batson, 1995, 1042–1055). This involved trying to determine if empathy/sympathy was altruistically motivating rather than motivating for egoistic reasons. According to the egoistic interpretation, empathy is associated with negative feelings of guilt,

shame or fear of social sanction, something we discussed in the pride, shame and responsibility model. It can also be associated with recognition of the positive consequences of helping behaviour such as social reward and/or feelings of pride. This interpretation is also characterised as "the social exchange theory" (Batson, 1995, 1042–1055). Batson specifically argues against this interpretation in favour of a more genuinely altruistic one and that empathy forms the basis of this. In the experiments, students were asked to listen to tapes from a radio program featuring a woman who has broken both legs in a car accident talking about how she was beginning to struggle in her classes at university. The students were given letters asking them to compare lecture notes and meet with her. In one group, the experimenters attempted to heighten the level of empathy by specifically discussing how the woman was feeling; whilst in the other they specifically portrayed this as irrelevant. The experimenters also adjusted the cost of helping depending on the group in question. The high cost group were also told that she would be in their class after school, and the low cost group were told she would finish at home. The experiment showed that the high cost group were just as inclined to help as the low cost group (Batson, 1995, 1042–1054).

The general consensus within psychology is that Batson's research, whilst impressive, is not necessarily persuasive, however, as the egoistic interpretation that he gives is extremely limiting and his conclusions do not necessarily confront, much less disprove alternative egoistic interpretations, nor does his research alone validate the thesis that empathy is a convincing **basis** for morality (Stueber, 2013). Hoffman on the other hand takes a far more biological perspective on empathy as a physical predisposition towards altruistic behaviour. His conception of empathy is based on various 'modes of arousal', which allow individuals to respond empathically to distress cues from other people. Hoffman also talks about mimicry as a form of social conditioning (Hoffman, 1981, 137).

If there is a connection between empathy and the expression of moral **utterances**, however, that is utterances aimed at communicating attitudes, then it stands to reason that in making such an utterance the speaker communicates, deliberately or not, that they empathise with the object of the sentence. Note here that I am taking the 'object' to mean that which is the object of a sentence, as opposed to the subject of a sentence, grammatically speaking. This is because what is under discussion is a moral sentence which **appears descriptive**, and so the target of the **projected** moral quality is the thing within the sentence that appears as having something done to it. I suspect that in some instances a speaker may also target a subject of a

sentence as well. For example, "...**you** *should not kick the cat*..." here 'you' is the subject, and the **cause for** disapproval, as the speaker knows what kicking involves. Again, this is information that is not specifically stated in the semantics of the sentence. It may, however, be nonetheless communicated if the speaker is projecting the quality of wrongness and if this concept is intrinsically tied to the application of a speaker's empathy. For example, if we take the following **process**:

1) An expressive fictionalist speaker observes a cat being kicked.
2) The speaker is also aware that being kicked constitutes a negative emotional experience.
3) The speaker is therefore aware that the cat must currently be enduring this negative experience by virtue of being kicked. (From 1 and 2)
4) *Observing negative experiences create(s) negative emotions in the observer*. (Assumption).
5) In observing the act, the speaker experiences a negative emotion (e.g. pity, indignation, outrage etc, as seen in the pride, shame & responsibility model). (From 3 & 4)
6) *Emotions motivate either preventative or reinforcing action (other things being equal).* (Assumption)
7) Speaking out in a condemnatory fashion is a preventative action; this may manifest in expressing anger/disgust towards the assailant, and/or expressing intent towards social rejection and/or love withdrawal.
8) The speaker speaks out in a condemnatory fashion. (From 5, 6 & 7)
9) The speaker speaks using moral terminology. (From 8)
10) In using moral terminology, the speaker therefore communicates that they empathise with the subject or object of the sentence, & expresses other directed shame towards the assailant, leading ultimately to the emotional **costs** seen above. (From 8 & 9)

I have deliberately italicised 4 and 6, because of this process these are the two steps in most need of further explanation. The main difficulty with (4) of course is that it is conditional on it being a case of someone who has the empathic trait of experiencing negative emotions when observing the negative experiences of others. However, where exactly does this 'empathy' come from? We may theorize that this empathy is a survival mechanism determined to ensure the survival and/or wellbeing of those around us thus aiding our own by

promoting the health and wellbeing of the **social environment** which we ourselves depend on. However, there may be some individuals who simply lack this trait altogether, for example individuals who are sociopathic. Another angle might be to suggest that if a speaker who lacks any empathy with the subject or object of their sentence speaks about them in a moral way, then there may be grounds to suspect that they are by definition being insincere. Moreover, we should remember that we are attempting to explain how moral language incorporates emotions, and as such we should focus on individuals who feel the relevant emotions we are discussing, and not the people who do not. This is an idea I will look at in a later chapter. (6), however, is the biggest assumption here. Whilst we have explored a great deal about how feelings are psychologically connected to preventative and/or reinforcing action associated with moral thought, we have yet to truly press home, **how** or in **what way** they are linked, and in what way, if any, they are causal. All we know, thus far, is that impaired functioning of one seems to result in impaired functioning of the other. If I were to offer any explanation at all, it would be that moral attitudes are a way of conceptualising how we 'feel' about the well-being of others, and are therefore an important means of communicating empathy.

Unfortunately, I suspect that these are not things to which we can have one hundred percent certain answers to. However, I think that there is enough observable evidence to say with a reasonable degree of conviction that there must be a connection of some kind between emotions and moral decision making of certain kinds, (namely those involving the way we relate to and behave around others). From observable changes in the decisions made by those with damage to their ventromedial prefrontal cortex, to apparent correlations between distress and seemingly altruistic behaviour in multiple animal species, to a lack of such behaviour from patients with severe depression and/or in certain cases of SSRI treatment, there is a great deal of evidence that suggests that what affects a subject's emotions also affects their ability to make moral decisions. How this links in with moral language is I would say contextual after a fashion in that we refer to these concepts of empathy as 'moral' ones (i.e. relating to how we feel about others) in conversation.

This is not to say that there is not a wealth of recent literature on the connection between empathy and morality; there certainly is. There are some contributions that are also sceptical about the importance of empathy. For example, Paul Bloom argues that empathy is too 'narrow in its focus, rendering it innumerate and subject to bias' (Bloom, 2017, 24).

According to Bloom, empathy 'can motivate cruelty and aggression and lead to burnout and exhaustion' (Bloom, 2017, 24). Bloom challenges empathy on the grounds that it is insufficient to produce results when compared to other characteristics such as compassion (Bloom, 2017, 24). Indeed, there may be some weight to the argument. Just because we may empathise with somebody or something does not necessarily mean that how we react will be in everyone's best interest. For example, I may empathise with someone who is suffering and be motivated to speak out against it, and yet in doing so perform more harm in the process if it is something that cannot be prevented, or is necessary for some subtle reason. By contrast, we need not necessarily understand or empathise with someone to react with compassion. For example, one can be polite and courteous to others, even towards people we regard as detestable. However, I think that we should avoid getting carried away here. What I argue for is how a non-realist speaker may use empathy to communicate emotions. What I do not argue is that this is always a good idea, or that it is always beneficial. I argue that one of the defining features of moral language is that it is **imperfect** and often speaking out or in some other way taking a stance on some 'moral' issue can often cause significant problems for the person speaking. So whilst examples like Bloom's go a long way to outlining the limitations of empathy, they do not invalidate it as a tool of recognising situations which motivate preventative or reinforcing speech acts.

Before we proceed with the next step, let us briefly discuss one more important issue and that is the difference between empathy and compassion. Unlike empathy, compassion does not necessarily require an understanding of what the one recognised as suffering is experiencing per se, only an understanding that that being **is suffering**. For example, there are many experiences endured by people which we have never experienced and in all probability will never experience. There are also some experiences which are only experienced by a particular group of people be that determined by sex, race, gender, body type and so on, and in many of these cases, complete and total empathy by those who have never had, and will never have these experiences is, to a certain extent, impossible. This does not mean, however, that we are totally incapable of responding to the needs of those whose experiences are completely alien to our own, only that recognising them can be more difficult. Empathy is such as a useful phenomenon in this regard simply because it allows us to identify with others on a deeply personal level. It is much more difficult to understand the suffering of someone (or something) that is experiencing something you have little to no concept of.

However, just like with empathy, compassion works on the principle of understanding that suffering is something that we ought to prevent, and so whilst compassion alone may be less efficient than empathy in that it is more difficult to **identify** what is and what is not 'suffering' in a person or being that we cannot so easily identify with, if we can still **recognise** it as suffering, then the result is much the same as the empathy case, in that it triggers emotions in us which motivate preventative or reinforcing action. For example, from a cisgendered male perspective, I think that attempting to completely empathise with a woman experiencing child birth would be futile at best (and probably a little insulting), because by definition it is not something I will ever go through nor am I physically equipped to even conceptualise what the experience is like, nor would I care to[38]. However, just because this is the case does not mean that I would be totally divorced from the situation were my wife or girlfriend to go into labour, nor would I fail to be overcome with a wide variety of very strong emotions were this to happen. My point is, complete understanding of what the other is experiencing is not strictly necessary, but rather recognition of what the other needs from you in that moment.

How then does this affect the empathy model? Essentially it does not, except to slightly modify (4) from an assumption about the speaker mirroring, to a certain extent, what the cat must be going through, to simply being concerned that that is happening, and wanting it to stop.

This may seem like a more genuinely altruistic account, but it should be noted that it has two distinct disadvantages compared to the original empathy model:

1. The assumption made by (4) is far narrower, as it requires a person to experience negative emotions **despite** not identifying in any way with the other (in the example case 'cats').

2. Because it is much harder to recognise suffering in an animal that we do not identify with like we do a person, it is less likely this will actually occur.

In other words, empathy is more reliable than simple compassion, because empathy **shows** us, albeit in a limited way, what the other is experiencing rather than merely alerting us to the

[38] The equal and opposite could also be said of biologically male specific phenomena such as prostate or testicular cancer.

information that they are suffering. It therefore **compels** us, in a much more potent way, towards that preventative or reinforcing action in response to it. In very extreme cases, and one in which there is already a very high degree of emotional investment, such as the wife or girlfriend going into labour, it is easy to see how a powerful emotional reaction galvanising one into action would be almost automatic. However, with the cat case it is less clear (I think) that this would actually be the case given the lack of prior emotional investment. By contrast, a non-realist who makes an active attempt to empathise with cats that are being kicked would, I argue, be far more compelled to speak against it.

Lastly, it should be noted that this ten stage process can, and often does proceed in both directions. In other words, it can lead to praise just as easily as it can lead to blame. For instance, an expressive fictionalist may observe another rescuing a cat from being kicked, follow the process and communicate (to the rescuer) empathy with the cat, admiration for the rescuer's compassion, their bravery, respect for their moral integrity and so on. They may simultaneously express other directed shame towards the attackers as seen in the original example. All this and more may be communicated by a single sentence: "It's horrifying that they would do such a thing; thank goodness you got here when you did!"

As we can see, there are thus multiple ways by which we can translate emotions into moral language, ways which are not mutually exclusive. The general idea is, however, that emotions triggered by observation compel speech acts which then express (by projecting onto the world) moral attitudes. In my next chapter, I will examine this empathy model on a much deeper level, and attempt to piece together a story about how this process ultimately comes together.

Chapter 5.2 Combining the Models

Now that we have seen both models in action, we should see how the two may work in conjunction with each other. To do this, I will break down the empathy model and examine each stage of the ten step process in detail. In doing so, I will (hopefully) be able to piece together a story about emotions and their role in moral thinking and moral discourse.

First, let us look at the empathy model once again. In my last chapter I put forward the following ten stage process as an example of how marrying an emotional perspective to the semantic fictionalist account might be achieved:

1) An expressive fictionalist speaker observes a cat being kicked.
2) The speaker is also aware that being kicked constitutes a negative emotional experience.
3) The speaker is therefore aware that the cat must currently be enduring this negative experience by virtue of being kicked. (From 1 and 2)
4) *Observing negative experiences create(s) negative emotions in the observer*. (Assumption).
5) In observing the act, the speaker experiences a negative emotion (e.g. pity, indignation, outrage etc, as seen in the pride, shame & responsibility model). (From 3 & 4)
6) *Emotions motivate either preventative or reinforcing action (other things being equal).* (Assumption)
7) Speaking out in a condemnatory fashion is a preventative action; this may manifest in expressing anger/disgust towards the assailant, and/or expressing intent towards social rejection and/or love withdrawal.
8) The speaker speaks out in a condemnatory fashion. (From 5, 6 & 7)
9) The speaker speaks using moral terminology. (From 8)
10) In using moral terminology, the speaker therefore communicates that they empathise with the subject or object of the sentence & expresses other directed shame towards the assailant, leading ultimately to the emotional **costs** seen above. (From 8 & 9)

Prinz mentions empathy in *The Emotional Construction of Morals*, referring to it as ‘mind reading’; “the ability to attribute mental states to others” (Prinz, 2007, 270). This allows us,

he argues, to acquire information about social norms, as well as promoting pro-social behaviour. It also allows us to acquire "norms about norms", turning observers into educators who "...can recognise our emotions and urge us to feel sorry for our bad behaviour..." (Prinz, 2007, 270) As we will discover later on in this chapter, this helps to **connect** the empathy model to the pride, shame and responsibility model. Social and psychological conditioning increases the likelihood that one will act in accordance with these norms, because, as we will see later, a failure to do so results in negative experiences such as emotions like guilt, or the disapproval of others manifesting in anger or disgust, leading to other consequences such as social abandonment or physical punishment. According to Prinz, the disposition to experience guilt provides us with stability, "...because we would feel guilty if we didn't feel guilty" (Prinz, 2007, 270). Therefore, the establishment of rules help to establish meta-emotions, which act in a similar way to an insurance policy deterring misbehaviour.

Before we proceed on this, however, let us first investigate our initial cognitive process in great detail. As mentioned before, I have also italicised the two most vulnerable assumptions necessary for my ten stage process to function. Before I address these particular concerns, however, I would do well to elaborate on, to an extent at least, the thought processes behind each component of this process.

The initial step is that an observer, who just so happens to be a fictionalist and knowingly so, observes, for example, a cat (or any other sentient organism) being kicked (or otherwise receiving violent or offensive treatment of some kind). This observation is important, because whilst I am attempting to piece together a non-realist case for moral discourse, I am not for a moment going to suppose that the **subject** matter of this moral discourse is not itself real. My point here is that whilst the language used to describe the moral characteristics of cat kicking voiced by a fictionalist may be non-realist, the fact remains that the observation of the kicking of the cat that inspire(s) such language is a **real** phenomenon, and so what we have is a non-realist language that nevertheless addresses **real** issues.

The second step is marginally more controversial, that the observer recognises that the kind of violent behaviour that they are observing constitutes negative physical and/or psychological experience(s) in whoever is receiving it. The main difficulty in justifying this step is that it must by its very phrasing take a somewhat one dimensional approach to the issue of pain and suffering. For example, can we continue to hold this as a plausible step

given that we also observe that an individual may actually **consent** to receiving pain and/or suffering? One option here might be to adjust the phrasing slightly to target specific instances of non-consensual ill treatment. However, I am not entirely sure this is necessary. Even if a painful experience was desired, it remains a painful experience, and so I believe this step still holds for its intended purpose. The ultimate aim of outlining the process is not necessarily to account for how a fictionalist **should** react when observing a cat being kicked but rather how one simply can do. It is important also to remember that moral reactions to observations are not always perfect or even appropriate to the reality of the situation. Often we will observe what we think to be an immoral act and react to it in a condemnatory fashion without realising the full context of what it is we are observing and thus whether or not our condemnatory reaction is appropriate to the situation or not.

If we accept these two steps, then we should accept that if the observer has the ability to deduce from a general principle that being kicked constitutes a negative experience, and that they are observing a cat being kicked, that the cat must, by this same principle, be having a negative experience. Here we encounter a problem, because here we are portraying the expressive fictionalist as making an inference, and this might run counter to some theories on empathy in which it is portrayed as distinctly non-inferential. This may also be a problem with empathy in the sense that if there is any ambiguity in our working definition, then we end up holding hostages to fortune without even realising it. I argue that it is therefore important to outline what we are essentially taking 'empathy' to mean here. By 'empathy' I mean, as Prinz does, the ability to **detect** what other people are feeling, and in some way experience and/or mirror those same emotional states in one's own mind as a result. There is nothing in this definition that necessarily prohibits an inferential means of moving from a simple observation to an emotional reaction to that observation; the observer may not be consciously aware of this inference, but that does not mean that it does not take place. It is also important to outline what we mean by 'inference' also. By 'inference' I mean the ability to reach a conclusion on the basis of evidence and reasoning. There is nothing in this definition that demands that such inferences are always made consciously. In other words, I argue that even if empathy does not always **appear** inferential, as on a conscious level emotional reactions just seem to occur in response to certain observations, on a subconscious perceptual level, it stands to reason that the human brain engages in some kind of inference when processing sensory information in order that additional information may be accessed. This kind of **unconscious inference** is also not without significant evidence in perceptual

psychology[39]. Optical illusions are one example of unconscious inferences in which observational 'premises' lead to conclusions which ultimately prove to be false. In a similar vein to fictionalism, Helmholtz also refers to theatrical performance as an example of unconscious inferences leading to emotional reactions to situations that are not really taking place.

> An actor who cleverly portrays an old man is for us an old man there on the stage, so long as we let the immediate impression sway us, and do not forcibly recall that the programme states that the person moving about there is the young actor with whom we are acquainted. We consider him as being angry or in pain according as he shows us one or the other mode of countenance and demeanour. He arouses fright or sympathy in us [...]; and the deep-seated conviction that all this is only show and play does not hinder our emotions at all, provided the actor does not cease to play his part. On the contrary, a fictitious tale of this sort, which we seem to enter into ourselves, grips and tortures us more than a similar true story would do when we read it in a dry documentary report. (Helmholtz, 1925, 28)

So according to Helmholtz, even material of a fictitious nature may incur an unconscious inference which leads to emotional reactions even more profound than genuine events; the difference is created by our ability to **experience** the event visually (or other perceptual means, e.g. a radio play) (Helmholtz, 1925, 28). This is also very similar to the issues I raised in section two in chapter six about the phenomenon of being more emotionally invested in a fictitious television program than in a news report. One may detach oneself from a dry news report, and convince oneself that the events being described do not apply to them, and yet convince oneself that they have a great deal of stake in events not really taking place. Again, these kinds of reactions are consistent with how we are currently portraying empathy as a means of detecting and mirroring other people's emotions. There is nothing in this definition that demands that these people or these events be real, or that on an unconscious level the process of moving from observation to reaction is devoid of unconscious inference.

[39] The term 'unconscious conclusion' was first coined by German physicist and polymath Hermann von Helmholtz in 1867 in *Handbuch der physiologischen Optik* (*Treatise on Physiological Optics*) as translated by James P.C. Southall (Helmholtz, 1925, 28). It has since been referred to as 'unconscious inference' and used by Edwin. G. Boring (as seen in Boring, 1950) and Daniel T. Gilbert (as seen in Gilbert, 1989).

This brings us to one of our more vulnerable assumptions: that observing negative experiences in other beings produces negative emotional experiences in the observer. The reason I have marked this as a vulnerable assumption is that this is, again, an assumption about the meaning of empathy and what empathy generally entails. It also should be painfully obvious to any reader that this is anything but a universal. Not only is it clearly possible to observe pain and suffering without any sympathy or concern whatsoever, but there even appear to be individuals who seem to lack such reactions altogether (e.g. such as is the case with sociopaths). Therefore, I would argue that this assumption be seen as something resembling the antecedent of a conditional. **If** the observer can identify with the subject of their observation in such a way as to experience negative emotion upon observing it, **then** the rest of the account should make sense, and the next conclusion, that the observer experiences a negative emotion, can then be reached. If we accept this assumption, however, then we should be able to easily accept the conclusion that in observing the act, the speaker experiences a negative emotion, such as pity, outrage, anger or disgust. This conclusion should also be aided in no small way by referring to our discussions on social conditioning in the pride, shame and responsibility model. As seen in that model, individuals are raised over a long period of time to experience pride for behaviours which conform to the morality narrative, and shame for those which abandon it. This will greatly influence which emotions are likely to be experienced in response to observing how others maintain that narrative.

The next assumption may also be somewhat controversial, however, that emotions motivate either preventative or reinforcing action, and by this I specifically refer to successful motivation where action occurs, assuming there are no countervailing motives sufficient to successfully counteract it. The problem is, of course, that it is not set in stone that emotions are indeed inherently motivating, despite what we discussed in regards to social conditioning in the pride, shame and responsibility model. There is also the issue of conflicting emotions; even if we assume that the negative experience of observing an abused cat motivates preventative action, there may be conflicting intentions which motivate the observer away from this. It is simply not clear that observing a terrible thing means that the observer will **necessarily** do anything about it. Therefore, once again, it is perhaps best if this assumption be seen as like unto the antecedent of a conditional claim. **If** the observer is motivated by their emotion(s) to act, **then** the rest of the account should make sense.

I trust that the step where speaking out in a condemnatory fashion about a thing counts as preventative action of some kind is suitably non-controversial. If so, we should be able to reach the point where if all of the above counts as true, then the observer will speak out in a condemnatory fashion, assuming that they do not instead intervene in an alternative way. It should be noted that here is where we really start to see the pride, shame and responsibility model's role in the empathy model. As we have seen, one very effective means of affecting the situation being described is to convey a sense of real emotional **cost** to the listener. This may be expressed by other emotions like anger or disgust, components of other-directed shame. This implies a degree of social rejection on behalf of the speaker in regards to the listener. For instance, remember one of our example sentences from the beginning of chapter five "How could you?" This can be a very potent sentence despite not really conveying much semantically. Semantically the sentence appears to be a request for information about how the action was achieved, which in the case of cat kicking ought to be self-evident (the kicking of cats is not physically difficult to achieve). That is not the implication; however, rather it is that psychologically bringing oneself to do such a thing 'ought' to be extremely difficult given that, according to the morality narrative, the assailant 'ought' to feel a great sense of shame, and in addition holds a responsibility to refrain and prevent such things from happening. So despite not asserting anything semantically, the sentence itself lying within a context of pretence surrounding the morality narrative, it nevertheless communicates a great deal of information about the speaker and their feelings towards what they directly observe.

It also stands to reason if we posit that an expressive fictionalist is operating within a context where a vast majority of the people they are addressing are either moral realists or are morally agnostic on the issue. Certainly if I were addressing someone I knew to be an unwavering realist, then it makes sense to use a language that I know they will understand, and be motivated by, if my purpose is to either reinforce or prevent whatever it is they are doing by creating a sense of real emotional cost. In other words, if I say that kicking the cat is wrong, then I stand a much greater chance of the abuser listening to me and ceasing the abuse than if I merely say that I do not like the fact that the cat is being kicked, especially if I know that the abuser happens to be a moral realist. This is simply because the abuser may or may not see my own negative experience as a result of my empathy as remotely relevant to their actions. However, they may view what I am saying as relevant if they believe that it contains some emotional **cost**. To borrow another idea from the original fictionalist account, I describe what the fictionalist does here as the usage of a particular degree of 'critical context', as

Joyce would put it (Joyce, 2011, 290), and I will discuss this a lot more in my next chapter. If we accept this, then we can easily accept that the speaker will choose to speak out in a condemnatory fashion.

The ninth step of the process comes from a conclusion from equivalence following from only one other sentence. My assertion here is simply to say that to speak out in a condemnatory fashion is directly equivalent to (or equal to) using **moral** terminology of some kind. This is assuming we are talking about a kind of condemnation that is regarded as worthy of special attention or concern regarding the genuine safety or wellbeing of oneself or others as opposed to the kind of condemnation that occurs when we express aesthetic disapproval or disapproval of poor etiquette. As stated above, this terminology also seems to be linked to preventative or reinforcing action and that it makes sense to use words like 'bad' or 'evil' in a world of realists who believe that these words have literal or referential meaning if the purpose is to convince them to stop whatever it is they are doing. This should also work in the opposite direction as well, as words like 'good' or 'kind', 'brave' or 'benevolent' naturally seem to carry more significance to any individual raised in an environment where possessing these 'virtues' or 'values' make one similarly 'valued', whether they happen to be realists or not. However, this brings us to a very important question about the expressive fictionalist, which also brings us to our final conclusion.

My final conclusion (concerning the combined model) is that if we accept all of the above, then in using this kind of moral terminology, the speaker communicates information about their emotions and thus their ability to empathise with the subject in question, as well as expressing other information such as other-directed shame, as seen in the pride, shame and responsibility model. As previously mentioned also, this may be done in the opposite direction as well; a speaker may follow a similar process to express pride in another's ability to conform to the morality narrative, and/or express empathy with their struggles in doing so. This conveys not only emotional information but also reinforces a sense of solidarity or collectiveness as well.

The question is, however, how deliberate is this? No doubt this process can be, and no doubt often is, deliberately carried out by non-realist speakers. However, I would argue that more often than not it is not. If we suppose for a moment that a speaker is doing this deliberately, and is deliberately **implying** that they have empathy through moral terminology, then that

raises important questions about the motivations for speaking about morality in a realist sounding way. It may even lead to rather pessimistic conclusions. For example, we could say that the fictionalist only speaks the way they do due to selfish motivations e.g. in order to manipulate others (the vast majority of whom are realists) into doing what they want (in fact if we were to look at this again from a psychological perspective, this idea is not without some fairly compelling evidence). However, I argue that this is **usually** not the case and that what is actually going on is far more complicated, and this involves a lifetime of conditioning to certain 'critical contexts', something which, once again, shall be discussed in my next chapter. In doing so, I hope to shed some light on the question of how these emotions manifest in varying levels of situation and examination, ranging from emotionally detached conversations about morality and the nature of it, in contexts like a philosophy seminar, to highly emotionally charged ones in the context of extreme situations as they are occurring and as they are being experienced in that particular moment.

Chapter 6: Critical Context(s) and the Problem of Deception

As I began to discuss in my last chapter, I feel it is important to return to the subject of critical contexts, because what we ultimately want is a thorough account which includes how an expressive fictionalist may speak about ethical values both in standard conversation as well as in highly critical, academic conversations about meta-ethics. Therefore, I would like to see how the Joycean account of critical contexts or as Joyce calls them 'critical perspectives' (Joyce, 2011, 289), and which I discussed in the second section of my thesis, applies also to my account of expressive fictionalism. Ultimately, I would like to make the case that the standard use of expressive fictionalist language more often than not is not deliberate but is rather a result of speaking in less critical contexts than when we talk about those same values in academic discourse. I also argue that these less critical contexts display some dependency on situational emotion states. The two main concerns in this chapter are 1) why the fictionalist would speak using moral terminology when they do not believe that they name anything real and 2) whether the fictionalist is being deceitful or manipulative when they do this. These are both important questions, because ultimately they might, if left unaddressed, cause serious problems for expressive fictionalism. It might seem plausible to suppose that someone using terminology which does not reflect their beliefs about the nature of ethics may do so specifically to mislead others. However, I argue that it would be problematic to apply this interpretation to all instances of expressive fictionalist discourse. What I do not wish to do is to portray expressive fictionalism as a form of elaborate lie spinning, but rather as something more akin to storytelling. The purpose behind telling others about a narrative that is not (in the strictest sense) true is rarely to mislead, but rather to communicate things above and beyond semantic content; one extensive example of this is literary fiction. In other words, in order to avoid the problem of deception, it is important not to look at expressive fictionalism at face value but rather to look beyond what is said literally and at how and in what **context** it is said.

In section two chapters six and seven, I mention Joyce's account of critical contexts in which a speaker may speak of things they do not believe exist. To recap, I will briefly go over Joyce's thought experiment once again in regards to colour fictionalism. Supposing an individual has a philosophical opinion or belief about the metaphysics surrounding the ontology of colour and has come to the conclusion that colour does not really exist. In other words, when it comes to the ontology of colour they are akin to an error theorist, but they

nevertheless continue to use colour terminology in everyday conversation. Note that I make no commitments towards any claims about the actual ontological nature of colour, rather what is at issue for me is the fact that this individual believes, philosophically, in an error theoretic account of colour ontology and yet continues to speak about colour in a way that does not resemble this account on a regular basis. It is thus an issue specifically related to the language surrounding colour and light. The individual continues to talk about colour as if it is a real thing even though they do not believe that this is scientifically or philosophically accurate. This, as I have said multiple times before, is an issue at the very heart of most objections to fictionalism. How does it make sense to talk about things realistically when one categorically believes that these things are not real? In response to this mystery, Joyce establishes a concept of varying degrees of critical context:

> ...it would be too bizarre to hold that an individual, who has never given the issue any careful thought whatsoever, but thinks and acts in accordance with theory *T*, does not really believe *T* simply because if he *were* to think carefully about it, he would deny it. But if we add that at some point he has adopted a critical perspective and therein sincerely denied *T*, and remains disposed to deny *T*, were he again to adopt that perspective, then he disbelieves *T*, regardless of how he may think, act, and speak in less critical perspectives. (Joyce, 2011, 290)

This, unfortunately, is the majority of what Joyce has to say on the matter in 2011. However, I believe there is much more to be said about critical contexts than this, as I believe this is by no means the only conceivable example of it and that our everyday language is, in fact, saturated with them.

First of all let us attempt, as best as we can, to recap what a critical context is in the first place. Based on what Joyce has mentioned, I would best describe a critical context as its name implies, that is as a conversational context in which the subject matter is under particularly significant critical and/or academic scrutiny. For example, suppose that in one conversation I describe distinctly anti-realist ontological beliefs about something as concrete as a physical object (such as this laptop). For example, I may take some ontologically sceptical approach or even a completely nihilistic one, and yet in an entirely different conversation, which does not happen to be about philosophy but rather about the fact that I use said laptop to answer emails & write thesis chapters, I will describe the laptop in a way

that directly contradicts the beliefs I conveyed in that earlier, more philosophical conversation. My point is that it seems clear from this observation that the level of realism we attribute to the subject of a given sentence seems to depend very heavily on the context in which we are referring to it and the critical nature of the conversation in which we utter it.

Therefore, in addition to what Joyce has to say on this subject, I would like to continue with some previously mentioned analogies and thoughts as well as adding some new ones. Colour fictionalism is a useful springboard into a discussion of critical contexts in everyday language, but what truly interests us for the purpose of this thesis is where we speak in these contexts specifically during moral discourse. Suppose that we choose to believe in the ten stage process of expressive fictionalist discourse I outlined in my previous two chapters. Given the frequency with which this process may occur in which a fictionalist speaks in a way resembling a moral realist despite having specific ontological views that contradict moral realism in a critical context, for the purposes of maintaining some semblance of meaningful dialogue with moral realists, to what extent and with what frequency should we expect this process to be consciously deliberate?

For example, suppose I am speaking in a philosophy conference on sports ethics and I am discussing match fixing or doping or some other related controversial subject. To raise a meta-ethical discussion at an applied ethics conference would no doubt be distracting at best and actively counterproductive at worst. I may make a conscious decision to speak using **realist** sounding moral terminology. However, I think such an assumption would be misleading. The aim is surely not merely to entertain or humour the other speakers and listeners but rather to also engage with them, for instance to persuade and/or debate them. The fact that I have a meta-ethical view surrounding the ontology of moral language does not alter the fact that I have strong emotional feelings surrounding what it is I am talking about. Imagine then that I feel very strongly about a particular philosophical point to do with doping. I may speak both passionately and with a degree of apparent moral earnestness on the subject. Am I being insincere? I would not say so. I would argue that I am merely attempting to persuade those listening to adopt my point of view, and that the non-realist nature of my more critical ontological view on the morality of doping is not, at that point at least, at the forefront of my mind in the context of that level of conversation. In other words, there are at least two layers of reasoning behind the specifics of why I use realist sounding moral terminology in this kind of setting. First, moral terminology seems to be the most effective

terminology to use when motivating a majority of others to one's own point of view (this is not to say that it will **always** be effective e.g. sociopaths simply may not care no matter what I say or how passionate I sound). Second, repeated use of moral terminology is intrinsically habit forming. In other words, I use this terminology so often that I do not have to think overly hard about the fact that I am using it. In other words, I am speaking in a **less** critical context than I would be if I were talking about meta-ethics rather than applied ethics. By 'less' I mean specifically, in this case, less critical of moral **language**. I may, however, be extremely critical of certain applied ethical views during a conversation on applied ethics. At best, I would argue that I am being **inadvertently** insincere. Assuming this is the case, does it make sense to suggest that it is necessary to make a conscious choice to speak resembling a realist, given the nature of this kind of language? I argue that it is not.

If the language I use were to accurately resemble my ontological meta-ethical views on whether morals are real or not, then I suspect my language would very quickly become either convoluted or at the very least less persuasive to the points that I am trying to raise. We could potentially counter this by supposing that were a particularly strict non-realist to speak in such a setting, one with strong emotional feelings towards the subject matter in question if not any belief in 'real' moral truth conditions, then they would endeavour to speak as objectively as possible without utilizing any moral terminology whatsoever. In other words, they would merely quote the facts without expressing any opinion on them, but would also be selective in the facts that they disclose. However, I would argue that whilst this may be a perfectly common argumentative strategy, not only is it by no means limited to non-realists, but that even the most devout of anti-realists are very much in the habit of both having and conveying moral opinions towards the facts that they disclose. One implication of what Joyce says is a phenomenon in which one's ontological beliefs appear to shift depending on the context in which one is speaking. For example, if we temporarily return to our colour fictionalism analogy, when pressed on what he **actually** thinks about the nature of colour, David replies that it is not a **real** thing. However, when engaging in casual conversation, David answers the enquiry about the colour of his mother's eyes replying that they are green. When pressed again in response to this, on whether or not he thinks his mother's eyes are **actually** green given his earlier confession that he does not believe green is a real thing, he begrudgingly replies that he does not, directly contradicting an earlier statement made in a less critical context. My question then is this: when David was asked about his mother's eyes

in the less critical context, did he believe the answer that he gave? I argue that he did, and I suspect that Joyce would also argue this as well.

To help explain why I think this, I would like to discuss one final concept. Suppose that I came home one day from a philosophy seminar on meta-ethics in which I discuss at length my non-realist views on the ontology of moral values only to discover that my house had been burgled & important family heirlooms or mementos of lost loved ones stolen while I was away. In other words, I am confronted with something which would have very **severe** emotional implications for me. Under such extreme circumstances, I suspect that the moral outrage that such a thing could even happen that I would inevitably express would certainly **feel** very real to me at that given moment despite what I said about moral ontology earlier or what I might say about it at a later date when I eventually calm down from the distress and anger that I would temporarily be experiencing. More importantly, in such an extreme situation, I suspect that it would be difficult for anyone to come to the conclusion that the realist **sounding** language being tearfully uttered is insincere. Indeed, I suspect that there are instances in every non-realist's life where a particularly extreme or traumatic experience occurs, they express a passionate, realist sounding moral sentiment and in doing so **believe** everything that they say at the time without any hint of insincerity.

So what is going on? If we look at this from a very strict Joycean analysis, then clearly traumatic or otherwise highly emotionally charged conversations seem particularly devoid of critical context, because other, more important aspects of those conversations **temporarily** take priority. In other words, if I am trying as hard as I can to convey the severity of my emotional situation to others in an attempt to motivate them to help me reclaim the important lost items, then the ontological nature of the moral values I am referring to are doubtless not particularly important to the pragmatic **objectives** of that conversation. Similarly, if we assume from my analysis that I am motivated to speak at all by strong emotions myself, then they will also be at the forefront of my conscious experience of that conversation. In other words, I describe what happens as a kind of '**emotion override**', in which any critical perspective I may otherwise take on my own moral language is temporarily overridden by the emotional experience and rendering me, in effect, a fully functioning **temporary** moral realist for the duration of that conversation. As to why this override occurs, I suspect that that is relatively straight forward. Not only are most of us raised in an environment surrounded by realist sounding moral language, but from both media portrayal and from genuine and

constant interpersonal experience, as discussed in previous chapters, rarely are these moral values expressed **dispassionately**. In other words, we have been conditioned, through experience and observation, to use moral language to persuade, inspire, comfort and confront by talking about it in both an emotional and a realist sounding manner, which means that when we are faced with a particularly emotion engaging conversation, such as one following the discovery of lost mementos of deceased loved ones, then the objective of the moral language of that conversation to motivate some kind of real life response demands a degree of empathetic focus which should override any critical perspective we may otherwise take on that language in a different context. I say 'should', because I argue this is what **usually** happens, but we should also not fail to forget that there are some individuals who are unusual outliers, and who are without the ability to behave so empathically when observing another's distress.

Chapter 7: Moral Progress

If we accept that expressive fictionalism, when analysed within a high degree of critical context, amounts to concluding that the moral language is used with pretence assertions (for the purposes, for instance, of emotive communication and social cohesion), then this may have some wider implications, specifically regarding whether or not there is a case to be made for the presence or detection of moral progress or moral improvement. To put it simply, how would moral progress be possible if moral language does not refer to or assert anything ontologically real? In the following two chapters I shall explain my reasons for believing that moral progress is possible within the narrative of an expressive fictionalist. Rather than trying to explain elaborate reasons why moral progress exists or is **true** in a quasi-logical sense, such as is the case with quasi-realism for example, I argue simply that there is no reason to suppose anything of that kind. However, that does not mean that we cannot or do not **treat** moral progress as a real thing even if in reality it is nothing more than a particular way of perceiving things, which is what I demonstrate in chapter nine. I will also attempt to argue, in a similar way that Blackburn does, that we can construct a case for an analysis of moral improvement of attitude by looking at the overall consistency of what is expressed, among other things. In particular, the case I will make includes an endorsement of a kind of anti-hypocrisy principle. Again, however, I argue that all of this lies within the context of the fiction rather than existing independently.

In my second chapter concerning moral progress, I will also look closely at the arguments of Catherine Wilson from her text "Moral Progress without Moral Realism". Wilson puts an intriguing epistemological spin on the subject, arguing that there is also a case to be made for moral progress if we compare the retrospective perception of new moral beliefs over old with a similar narrative of scientific understanding of the natural world. According to Wilson, we can track a narrative of epistemological progress fuelling moral attitudes in a way very similar to what we see in the progression of theories in natural sciences. Just as there are pragmatic reasons for elevating the oxygen theory of combustion to the status of truth, replacing the phlogiston theory of combustion, so too are there pragmatic reasons to do the same thing with the theory that slavery is morally indefensible (Wilson, 2010, 106). This is also compatible with my own argument from anti-hypocrisy in the sense that it is beneficial to all agents involved, in most situations, to be as informed as possible to ensure that the moral attitudes expressed are consistent and do not create double standards. There are,

however, also some problematic issues with this approach which I will attempt to deal with. There are also some axiomatic assumptions which I am using, most notably that it seems given by the very nature of the idea of a 'moral' rule that it should reasonably apply to the needs and desires of the subject in question, that it should appear well constructed, consistent and that it should not create unnecessary double standards (this being a similar definition of 'moral' to the one used in Hare's *Freedom and Reason*) (as seen in Hare, 1963).

To begin with, we must ask the simple question of how moral progress may be possible under an expressive fictionalist analysis. My first response to that question is that given the nature of expressive fictionalism not as a necessary error theoretic account, but rather as the account championed by the moral agnostic then one kind of progress is already theoretically possible.[40] What I mean of course is that there is room under this analysis to suppose metaphysically **real** (in one sense or another) moral facts and/or properties, which exist independently from the fictionalist language used, but that the language does not actually **refer to** them, being based semantically on pretence assertion. It is theoretically possible for moral progress to be made if the fictional language in some way mirrors what is metaphysically real. However, such mirroring would be coincidental, as again the language does not refer to the fact or the property, it merely pretends to. As such, I do not find this either sufficient or persuasive, and I certainly question the qualitative value of 'accidental' moral progress as being at all in the same league as moral progress as a result of conscious self reassessment, experience and attention to past events, both personally and academically; this is, I clarify, what we are really looking for when we talk about moral progress. It becomes more difficult, however, to account for this kind of moral progress if we choose not to presume anything whatsoever about moral ontology; the moral agnostic chooses not to rely on the existence of any realist notions of moral truth to help them out here, much less anyone's epistemic capacity for providing any reliable information on them even if they do exist.

If we cannot rely on moral ontology or moral epistemology to directly validate moral truth conditions, then what can we use to assess moral sentences? Can we assess them? One thing

[40] Harkening back to my chapter on Jonas Olson, I describe the moral agnosticism in expressive fictionalism as the commitment (or lack thereof) of one who is neither convinced nor fully dismissive of the proposal that moral properties exist. To the expressive fictionalist, such a proposition is largely irrelevant, because what this kind of **revisionary** fictionalism is describing is not what necessarily is or is not the case metaphysically but rather what is simply treated as such in order to express some kind of attitude towards the relevant action or circumstance.

that we can look at is the consistency of moral sentences (this will mean 'consistency within the pretence' in the case of the fictionalist). For example, if the moral justification I give for not wanting to be kicked is that people ought to generally refrain from kicking other people, because being kicked is painful and unpleasant, then it would be hypocritical to then proceed to kick people for no reason. General moral rules which we (fictionally or not) assign to large groups of people are, therefore, fairly easy to grasp. What becomes more complicated, however, is when we assign very specific rules to specific groups, or individuals or oneself. For example, recall the case I used in section two chapter three where a monk takes a vow of celibacy but does not complain when one of his friends, who has not taken any such vows, gets married and starts a family. This does not seem to be overly hypocritical; the assignment of this celibacy 'rule' is deeply personal and deeply specific to he who takes it. Were the monk himself to get married and have children, however, without relinquishing his vow, then this surely would strike people as hypocritical. It does, however, become more difficult to make reasonable cases of hypocrisy against people when we consider that different people, or for instance different animal species, have different needs and different desires. Whilst I would certainly argue that kicking a cat is at the very least a similar, if not equal offence to that committed by a human kicking another human, it seems intuitively obvious to us, for example, that putting a collar around the neck of and assigning ownership of a cat is not in the same vein as doing this to a human being.[41] Generally, we do not regard the practice of keeping pets as morally problematic, and yet the idea of keeping human pets is regarded as morally abhorrent. This sort of issue is important to understand, because it helps to demonstrate the limitations and subtle nuances of consistency as a tool being applied to the issue of hypocrisy.

So it seems as though a rigid form of consistency **alone** is not enough to make an attractive case for moral progress. If, for example, we say that the way we treat adults should be consistent for children as well, then this would no doubt be an undesirable conclusion. Instead, it would be better if we utilized a more loose form of consistency whereby assigned moral rules are consistent **unless** there is a morally **relevant** difference that requires some kind of exception clause. In other words, moral 'rules' do seem to require some kind of tailoring for whom, specifically, they are intended. What appears to be more questionable,

[41] It should be noted here that assigning ownership of human beings without their consent has existed in human history for a very long time. The difference is we call that slavery, and we do not advocate it in mainstream 21st century moral tradition or in international law.

however, and more open to analysis is the extent to which physical, environmental and/or neurological differences between individuals really are relevant to the moral rules being assigned. What is also open to analysis is the factual accuracy of the alleged differences which supposedly call for this difference in treatment. For example, it is difficult to reasonably argue, within the context of a 21st century scientific understanding of neurobiology, that there is any reasonable evidence for a conclusion that women are any less capable or competent in political or managerial roles than men are. Therefore, the fact that this has been a commonly held belief that has been held for significant periods of time prior to 20th century feminist movements should strike us as questionable at best. The question is, where exactly do we place the blame for this? Do we use this as justification to accuse the ancient world of a kind of 'moral laziness' or 'moral hypocrisy', or do we argue that they had less scientific knowledge about the subject, and so assumptions of this kind where not unreasonable within that historical context? Unfortunately for ancient peoples, I am inclined to lean towards the former conclusion for the simple reason that to make any such assumption in the face of ignorance strikes me as epistemologically irresponsible.[42] In other words, if they had no reasonable evidence, whether by rational deduction or by empirical validation that women were any less capable or competent at managing political institutions than men were, then the default assumption should have been that they were not, and in fact there are several very famous exceptional instances of ancient women in very high political positions which directly provided evidence to the contrary even at the time (Balliet, 1974, 89).

I would, however, clarify that many of what we would regard nowadays as bad conclusions were not always seen as such within their historical contexts. There have been attempts at rational arguments on the subject of women in politics throughout history just as there have been on many matters we now regard as morally problematic. For example, in ancient Rome it was largely believed that slavery was morally acceptable for the simple reason that one allowed oneself to become a slave (Funk, 2011, 15). In other words, if enslavement was unacceptable then death would be preferable. If a slave did not want to be a slave, then they did not have to be, they could simply commit suicide, and this would be seen as a very honourable thing to do, and so they believed the lack of hypocrisy stemmed from the following attitude, "I would rather die than be a slave, so the fact that my slaves clearly don't

[42] It is not a lack of scientific knowledge that is necessarily problematic, but rather the presumption of a conclusion in the absence of evidence based validation or rational inference and a presumption that is primarily made for the pragmatic purposes of maintaining the status quo at the wilful expense of the liberty and potential of a particular group of people.

think this, not having killed themselves, proves that they are different", so just as animals do not express a desire for liberty, those slaves that did not commit suicide were placed within a similar category (Funk, 2011, 23). Whilst this is not a particularly philosophical approach, it is one which is endorsed by ancient philosophers, from Aristotle's theory of 'natural slavery', to Seneca, to the ex slave and stoic philosopher Epictetus in the first century AD, the stoics, for instance, famously advocating benevolent treatment of slaves but never actively opposing the institution of slavery, believing that it did not necessarily impede the stoic ideal of liberty of will (Funk, 2011, 23). My point, of course, is that whilst the common justification of slavery from antiquity does indeed resemble a logical argument, it is not a very well constructed one, as it is clearly something of a red herring.[43]

By now we should be able to piece together some criteria for a qualitative analysis of moral rule assignment. The rule should reasonably apply to the needs and desires of the subject in question, it should appear well constructed, consistent and it should not create unnecessary double standards. Most importantly, it should reflect some factually accurate picture of what the needs and desires of the subject are within the context of the situation that they are in. For example, the needs and desires of a child are quite different from those of an adult, and they are usually relevant to how we treat them in, say, a parental, disciplinary or educational context, whereas the needs and priorities of an adult woman in a political or managerial context are rarely if ever so different from those of most men in the same position that her gender should be regarded as remotely relevant to how she is treated or how seriously she is taken. Returning to the matter at hand, however, does all of this help us address the issue of moral progress from the perspective of an expressive fictionalist? If we take moral progress to be something within the fiction we are projecting, then the answer would seem to be yes, but the question then becomes what exactly it is that we are pretending. I argue that what is being pretended/projected is the use of consistency and anti-hypocrisy being applied to recognised rules and social norms. The idea is that we can see changes in the rules that an individual or a society **accepts** as being progress if the new rules conform better to these kinds of criteria, and that we have a tendency to **call** that moral progress.

[43] Just because something as extreme as death may or may not be preferable to slavery does not make slavery at all desirable or acceptable, provable by the fact that many slaves attempted to escape captivity. By the same logic, I could argue that it is acceptable to commit banditry by threatening members of the public with a firearm, because if they really didn't want to be robbed, then they would fight back and risk me shooting them. This is **clearly** a fallacious argument. You cannot justify an action or situation just by the threat of something worse.

Chapter 8: Wilson on Moral Progress

In the last chapter I discussed how we may assess moral progress from the perspective of an expressive fictionalist. This involves establishing a kind of non-hypocrisy principle by which we can examine the consistency of moral belief states as applied to varying situations and people as compared to how one also applies them to oneself. One of the key features of this is recognising that it is not enough to simply be consistent in self assigned moral 'rules' when applying those rules to other people, because, for example, the rules one adheres to as an adult may not be exactly the same as those one assigns to a child. In other words, the rules that we assign require some tailoring to the needs, vulnerabilities, beliefs and emotions of the individual or group. This then leads us to examine the factual accuracy and/or **relevance** of supposed criteria by which we differentiate the needs and treatment of others. For example, we do not hold children morally responsible for their actions in the same **way** that we do adults, because we realise that they are at an earlier developmental stage psychologically. By contrast, there are no reasonable grounds for affording a similar liberal attitude to harm caused by people of a particular race, gender or religion, as these factors are not relevant to whether or to what extent they are responsible for their own actions like any other adult. In this chapter I will continue this theme of epistemological factors and how they contribute to a means of assessing moral progress from the perspective of an expressivist (i.e. where moral discourse is treated as a communication of attitudes). In particular I will reference some of Catherine Wilson's work on this subject, who I believe gives us a very elegant analogy to how we make progress in our understanding of the natural world. However, I will also discuss why this analogy suffers from a similar problem to that of quasi-realism; it tries to give the moral realist everything they want without committing to any realist metaphysics, but in this case it becomes borderline relativistic and **semantically** ambiguous. I believe we can take the pragmatic element from this, however, but that we still require some semantic explanation, which I will attempt to provide myself from the position of an expressive fictionalist, which I believe fares better in this respect.

As mentioned, a good method we can use to assess moral progress from an expressivist viewpoint is an overall progression of **epistemological awareness** of information relevant to the construction of moral systems aimed at avoiding hypocrisy, inconsistency and double standards. This appears to be a recurring theme in the work of Catherine Wilson in "Moral Progress without Moral Realism", and from which I take some inspiration here. Wilson

argues that moral belief changes share certain key features with theoretical progress in the natural sciences, and that the process of convincing one's peers that one belief is formally and empirically superior to another means that the seemingly better theory will often be promoted to the status of truth, that is **accepted** as true (Wilson, 2010, 106). For example, consider the following two sentences:

1a) M used to subscribe to the phlogiston theory of combustion, but now she has come to favour the oxygen theory.

1b) M used to subscribe to the oxygen theory of combustion, but now she has come to favour the phlogiston account. (Wilson, 2010, 105)

According to Wilson, 1a) is a story about epistemological progress, whilst 1b) is one of regress (Wilson, 2010, 106). Certainly we can understand why 1a) is progressive whilst 1b) is regressive if we assume that beliefs improve by matching reality or corresponding facts. However, surely this is, by definition, not the case for moral claims unless we are arguing for some level of moral realism. If the assumption from which we are working is that there are no moral facts or reality to correspond to or to corroborate moral claims, then what exactly is the difference, in terms of either progress or regress, between a and b? So it seems that this cannot be the direction we must go to understand what Wilson is saying. Rather, the strategy appears to be to reverse this direction by saying that what we take the facts to be, what we assign as 'true' is just whatever beliefs survive some process which is driven by pragmatic concerns. Wilson goes on to give a similar account of moral progress and regress to her account of scientific progress and regress:

2a) M used to hold that slave labour was sometimes necessary for a society, but now she holds that no one should own slaves.

2b) M used to hold that no one should own slaves, but now she holds that slave labour is sometimes necessary for society. (Wilson, 2010, 106)

However, it is not enough to simply say that these two examples are analogous without explaining why it is that Wilson thinks this. Surely there are important differences. In the case of scientific progress, we have **measureable** evidence from experiments, which

demonstrates that the oxygen theory is more factually accurate than the phlogiston theory. For a moral anti-realist and/or a moral agnostic, we do not have a similar apparatus or working knowledge of moral facts to act as any frame of reference to verify or measure the superiority of anti-slavery attitudes to pro-slavery attitudes. So, what methodology does Wilson use to compare the scientific progress of example 1, with the supposed moral progress of example 2 in such a way that they seem analogous? As Wilson puts it, "what process of confirmation established these truths?" (Wilson, 2010, 106)

Wilson's answer to this conundrum involves what she calls 'unidirectional narratives', in which, in **both** the scientific and the moral case, we can describe a narrative of **belief change**, in which both are treated as examples of progressive thought, but in which the understanding of this is **retrospective** (Wilson, 2010, 106). For example, it would be unfathomably arrogant to accuse 18th century scientists of being unintelligent simply because they did not understand the relationship between oxygen and combustion whilst in the 21st century this is common knowledge to most high school students. It is, however, important to remind oneself that the scientific method that we use today is far more sophisticated than it used to be not by accident but by a rigorous process of trial and error stretching back centuries. Put oneself in the 18th century and be given only the data and the methodology that was available at the time and one may very well come up with a similar conclusion. In other words, the phlogiston theory is a perfectly reasonable and logical explanation of why things burn, **given** the data and the methodology that they had. With the addition of developments to the scientific method and acquisition of additional data, we have retrospectively reached the conclusion that the oxygen theory is a better explanation, and so we have elevated it to the status of truth, replacing the phlogiston theory. We would also do well to remember with a degree of humility that should further developments in data or methodology arise, the oxygen theory might itself be replaced by an even more reasonable explanation. This narrative of scientific progress portrays the apparent 'improvement' in scientific belief not as correspondence to ultimate fact or ultimate reality per se, but rather more how that conclusion is reached, and how it is **perceived** retrospectively.[44] In **that** sense it is analogous to the moral case. We can observe a similar unidirectional narrative of progress in applied ethics and moral philosophy on the topic of slavery whereby we judge new attitudes superior to old based on innovations in the kinds of arguments used or the acquisition of new perspectives, and much like the

[44] The scientific community is not in the business of declaring that their conclusions are most definitely true. To the contrary, they will be the first to say that their conclusions are as **close to** the truth as we can **currently** manage given the data and the methodology currently available.

scientific case these new attitudes are elevated to the status of 'truth' and replace previous attitudes.

As Blackburn has a particular way of depicting moral truth, so too does Wilson depict **both** moral and scientific truth as being based on things other than correspondence to reality or nature, most notably retrospective perception. Both are driven by pragmatic rather than semantic concerns. In other words, what is being said is not as important as what is being communicated. From an expressivist point of view this seems to be, in the moral case, to express some approving or disapproving attitude towards the subject of discourse. In the scientific case, it is to convey the best possible explanation for observed natural phenomena. At face value this appears to be a significant difference, but when we start to unravel these apparently distinct cases more similarities begin to emerge according to Wilson (Wilson, 2010, 100), and this is to do with theories and belief states. In order to communicate some belief state about oxygen there must be, as Wilson puts it, a logically ordered nexus of wider beliefs about chemistry and about how the natural world works, or at the very least to be 'following or imitating some causally related source that does hold such a nexus' (Wilson, 2010, 100). According to Wilson, scientific claims about oxygen are interpretable 'only within the wider theory of chemistry' (Wilson, 2010, 100). Similarly, moral claims, whilst independently non-referential (not corresponding directly to a moral fact), serve as a proxy to a very similar nexus of belief states:

> The conjecture that capital punishment is wrong has several rivals to contend with: the conjecture that capital punishment is morally permissible as well as the conjecture that it is obligatory for certain crimes. These theoretical conjectures too stand proxy for entire sets of beliefs about the awfulness of certain crimes, about desert, deterrence, and other matters. My conjecture that torturing cats is wrong is dependent on my beliefs about what actually happens when a cat is being tortured, and it is in competition with rival conjecture that torturing cats is plain good fun because cats are insensate machines, or because the feelings of cats do not matter. It is the isolated moral statement that is considered truth apt, but behind it is a theory about the way the world is and what ought to happen in it. (Wilson, 2010, 100)

So what seems to be happening, according to Wilson, is that both scientific and moral claims act as proxy utterances summing up opinion on a large nexus of interacting belief states about

their retrospective perception of some important matter about how the world is, and what ought to happen in it. However, even this, personally, I find unsatisfying.

I agree that it is very convenient for us in the 21st century to accuse the ancient world of a kind of moral blindness in regards to something like the institution of slavery, and that we would do well to remember that we only **think** this because we live in a time period which fosters particular brands of the cultural values of liberalism, dignity and egalitarianism, and reflects retrospectively on a long and fascinating history of civil rights and abolitionist movements. Put yourself in ancient Greece or Rome, and listen to what Aristotle or the Stoics say on the subject of slavery, and the practice of slavery might not seem so barbaric, but rather as what is completely normal for that time period. I would, however, add to Wilson's argument by clarifying, to myself as well as others, that this need not necessarily be an argument for moral relativism. Just as we can argue that there are good philosophical and pragmatic reasons for judging the oxygen theory superior to the phlogiston theory by appealing to developments in scientific methodology, evidence and research data, so too can I do something similar with the moral case by appealing to developments in moral philosophy and the study of history. However, it is important from the position of both the anti-realist and the moral **agnostic** (where there are no known/knowable moral facts to **verify** our conclusions), for us to find some additional explanation as to how this works exactly and to place such explanations and arguments within their proper **semantic** context. For example, what is the point in justifying the abolition of slavery if, in **reality**, slavery is neither good nor bad; it simply **is**?

Again, this seems to be an important difference between scientific discourse and anti-realist moral discourse that Wilson, much like Blackburn, does not quite confront head on. In both approaches the goal seems to be to be able to give the moral realist all that they want without actually committing to any realist metaphysics. However, just as we discussed in section two, this is a very slippery fish to catch. Even as a pragmatic expressivist, what is actually being **communicated** when we say that slavery is wrong is surely to convey a disapproving attitude towards slavery. In this case, we do not need moral facts to verify this, because our own attitudes should be clearly apparent to we who are experiencing them. However, this is surely **not** analogous to what is being communicated in scientific discourse about oxygen. In scientific discourse, generally speaking, realist facts are being asserted semantically. We can say that the scientific community is not in the business of declaring absolute **certainty**

regarding the truth or the reality of these facts, rather in terms of what is probable given the evidence, but that does not mean that they are not presented or intended to reference what is ultimately real and true. So unless the expressivist has some notion of truth in which moral attitudes are comparable with scientific facts on a semantic level, then this is a big problem. Blackburn attempts to do this of course and runs into all of the difficulties I discussed in sections one and two. We could, alternatively, try to go in the other direction and portray a non-realist picture of scientific discourse in order to make the two more comparable and then work on the analogy within the pragmatic content(s), which seems to be what Wilson attempts to do. However, this too is problematic. Scientific discourse is generally regarded as being distinctly realist in nature, both semantically and pragmatically (the goal being to communicate what is most probably true fact and not to express any approving or disapproving attitude towards it).

What we can do, however, is present a fictionalist narrative of moral progress discourse which is analogous to scientific progress discourse. Just like any kind of fictionalism, however, this does require a degree of bullet biting. Referring to my chapter on critical context(s) (see section three chapter six), if we take a hard line fictionalist approach to moral progress then we are ultimately forced into the conclusion that, in a maximised (and extremely rare) level of critical context, there is no compelling evidence in favour of any moral claim about progress, and there is no reason to think that any moral assertion about it is true, rather it is a fictional concept. This is what it means to take an agnostic stance on moral metaphysics. If we take an anti-realist stance, then we should by rights go even further and say that there is no such thing as moral progress any more than there is morality in general. Despite what Blackburn may claim, this does seem to be an inescapable result of being a non-realist; to be a non-realist is not to believe that something is ultimately **real**, and to be a non-realist about moral language is to believe that moral discourse does not **refer** to anything real. However, and this goes both for non-realism and for moral agnosticism, this does not mean that moral discourse (including discourse about moral progress) is meaningless, insignificant or vague, any more than fictional discourse about Spiderman or any other fictional object is. Pragmatically, we use fiction to communicate above and beyond the semantic content of a sentence even if semantically, we are saying very little. Referring to earlier discussions of expressive fictionalism, I would say that whilst ultimately it may or may not be a true fact that slavery is wrong, I am still inclined to say so, 'as if' it were so, because not only do I disapprove of it, and I wish to communicate this, but I also want action to be taken against it

by others including moral realists, and because when I am not looking at it in a maximised critical context (which is most of the time), I perceive it exactly as I would any **real** fact about my environment.[45]

So in essence what we can have is the same kind of fictionalist narrative of moral progress that we do for moral assertion(s), moral utterances being pretence assertions which look a lot like other, more **fact** oriented utterances (e.g. scientific ones). Similarly, the process of conveying a narrative of moral progress is much like telling a story about something which is very similar to scientific progress. In other words, moral progress can be seen as part of the morality fiction. We can assert (or pretend to assert) that 2a) is moral progress, and 2b) is moral regress; these statements are untrue in a maximised critical context but are nevertheless treated as true in the same way that scientific facts are treated or any other claim about how we perceive the world to work. To be analogous, the two do not need to be given the same ultimate truth status; they need only **behave** in a similar way. For example, if I were to say that the way Spiderman conceals his identity as Peter Parker is analogous to a spy concealing their identity from their family, one is fictional (it has not happened) and the other is not (it does happen; spies in fact do conceal their identities from their families, else they would be poor spies). So we need only **treat** moral progress as a real thing to make this analogy; we need not say that it is ultimately the same in every possible sense. With that in mind, we can then begin to look at all the ways in which they **are** similar.

[45] It may or may not be true that slavery is wrong; as moral agnostics we should not discount the possibility that there is a moral fact that could conceivably verify this one way or the other, but we should also not **rely** on this, as what we are ultimately looking for is a way of accounting for realist sounding language without any **actual** realism.

Chapter 9: Applying Expressive Fictionalism---A Historical Narrative

To recap what we have discussed thus far, if we want expressive fictionalism to be complete, then we want to be able to convey a narrative of moral progress, but in doing so we cannot compromise our own non-realism, otherwise we become, as Lewis describes of Simon Blackburn, queasy-realists (Lewis, 2006, 319). Thus far in this section I have provided a form of fictionalism which also contains additional layers of pragmatic discourse which acts in a way similar to expressivism or quasi-realism; the idea is that we project qualities of wrongness through moral utterance in order to communicate emotions and attitudes, such as approval and disapproval, but unlike quasi-realism, expressive fictionalism implies that this is done through the lens of fiction and does not refer to anything real or quasi-logically true; ultimately, it is all pretence. In the last two chapters I explained that we can also apply this idea to moral improvement and moral progress, and this involves applying (or pretending to apply) rules of consistency and aversion from hypocrisy or double standards. In this chapter, I will temporarily enter the fiction and turn on the expressive fictionalist projector in order to demonstrate how this is done. In this case, I will use as an example one means of applying this principle of comparing moral progress with scientific progress that I discussed in my last chapter on Wilson, and that is is by looking at human history through the lens of this morality fiction.

Whilst I may run the risk of constructing something akin to a pseudo Marxist reading of history, looking with a somewhat euro centric pair of eyes at empires of antiquity followed by medieval feudalism, industrialisation and capitalism followed by early modern socialism, or seeming to throw my lot in with the likes of Thomas Hobbes in his position on civilisation as opposed to that of Rousseau, I would nevertheless argue that there is an apparent progression throughout human history towards greater understanding of nature and of the universe. In a very similar way, there is also greater understanding of socio-economic systems and in turn greater understanding of conflict(s) and how and why they arise. This can be interpreted as a greater understanding of people in the sense of how they work, physically and psychologically, what the similarities and differences are between different groups of people and how relevant they are within various political, social and scientific contexts. Within the morality fiction, assuming what we have claimed to be the case in previous chapters, this creates a narrative of moral progress analogous to that of scientific progress. Whilst this trend does appear to be ultimately progressive in nature, it is not, however, without several

significant and famous backlashes, just as there has been an overall progression with **occasional** regression of technological and scientific progress throughout history. I will discuss this further later on.

There are, however, some additional issues with the analogy that Wilson makes between something like the historical narrative of slavery and the phlogiston theory of combustion that I would like to add before continuing, and I would like to make them now, because they are at this point very relevant to the how we are portraying the morality fiction. As I have mentioned before in my previous chapter, something like the Roman belief that slavery is justifiable because one allows oneself to become a slave by failing to honourably commit suicide is unlike the phlogiston theory in that it is not an argument based on any real lack of evidence, nor is it a credible attempt to explain the nature of things. To the contrary, the fact that slave escapes and even armed slave rebellions occurred should ideally have at least led people to question the argument somewhat, and so the fact that this does not seem to have happened very often seems to imply that, just like many uncomfortable or **inconvenient** moral issues today, such as intensive animal farming, the environment, third world exploitation or foreign civil warfare, there was a degree of wilful apathy amongst Roman citizens in regards to the plight of slaves. In addition, ancient philosophical attitudes towards the institution of slavery, such as those of Aristotle (Wayne, 1987, 390-410), were sometimes and unapologetically born out of pragmatic necessity. It is fairly easy in a 21st century age of machines and technology to condemn the ancient world for the practice of slavery, but it is worth considering that without those institutions, it is unlikely that many of the great civilisations of the ancient world would have ever existed. By contrast, the phlogiston theory of combustion, far from being an argument intended to preserve the status quo put forward from a position of wilful apathy for the purposes of avoiding inconveniencing people's comfortable lives, was never a particularly politicised argument, rather it was merely an attempt by 18th century chemists to explain why things burn.

The reason I bring this up now is because this highlights an important consideration for moral rule assignment as analysed by the expressive fictionalist. If we want to analyse the quality of a moral expression by the factual accuracy of the subject matter that inspires the respective attitude, then aspects like political convenience and political allegiance must surely play an important role in the morality fiction as it is portrayed by individuals. On the other hand, it is simply not enough to say that the fact that everyone accepts something as morally acceptable

proves that this is an attitude of any real quality, because this has absolutely no baring whatsoever on the **fact**, for instance in the case of Roman slavery, that millions of men, women and children suffered and died as a direct result of this collective attitude. By intuition alone, this certainly seems unsatisfactory. When we say that a moral rule 'ought' to be based on some factually accurate picture, current popular opinion may be taken as an indicator, but never as sufficient justification, for the simple reason that individuals that make up large populations are often very divorced from moral issues that do not directly and negatively affect them or anyone close to them, often deliberately, and this must surely influence their arguments somewhat. After all, whilst our modern societies are vastly different from the small hunter gatherer tribes which human beings first evolved into, hundreds of thousands of years ago during the stone age, our brains are virtually identical in evolutionary terms, and we have simply not yet evolved to cope with the awfully inconvenient truth that our actions now have social consequences far beyond those who live in our immediate vicinity.

What does seem apparent from Wilson's argument, however, is that the trend of scientific, moral and technological progress, as we tend to see and portray it, appears to be linked to peoples' **understanding** of the world and those who live in it, and all of these things are related to one another. Indeed, what is science and technology for if not for at least **aiming** at providing people with an easier, more opulent and generally better quality of life, not to mention increasing their capacity for survival? Healthcare is perhaps the most obvious example of this, but also things like transportation, electricity, communication, all of which are at least **intended** to augment our day to day lives and our capacity to be productive (note: I say 'intended' because this is not always the result). Even weapons, whilst harmful by definition, are rarely if ever invented to cause harm for the sake of harm, but rather are intended to protect and defend one group of people from another and/or to forcibly secure resources for that group (albeit at the expense of another). Hunting is also another primary use of a weapon dating back to prehistory. Similarly, morality too appears to hold a very similar function in that it appears to provide better quality of life and better survival prospects by focussing on things like social cohesion and on mutual cooperation and understanding. However, if we are talking about progression in all of these things, then we would do well to consider its opposite; what about regression? This is also important to ask, because unlike Wilson, I argue that scientific and technological progress is not unidirectional but bidirectional; it does go backwards sometimes, and so if we want to claim, by way of the morality fiction, that moral progress behaves in a similar way, and is therefore in some sense

analogous to science and technology, then we should consider this as a potential aspect of it within the fiction.[46]

Before we can answer the question of moral regress, however, we should see if scientific or technological progress indeed appears to hold up in both directions. Can we have scientific and/or technological **regress?** Most certainly if we define this as a loss of either some useful methodological approach or more overtly in the case of a loss of technological capability. Whilst the description of the period between the 5th and the 11th centuries following the decline of the western Roman empire as "the dark ages" is a controversial one amongst modern historians given the continued achievements of the Byzantine Empire, the Sassanids in Persia, the Tang Dynasty or Moorish accomplishments in Spain, it has been generally accepted that in much of north western and central Europe (such as Gaul, Britain, Italy and parts of Southern Germania), there was a severe and dramatic loss of knowledge regarding architecture, engineering and healthcare as well as things like history, art and literature, not to mention previous century's worth of decline in military technology (Postan, 1966, 5).[47]

The question then becomes: can we see similar regressions in **moral** attitudes throughout history? Even within the morality fiction, this is far more difficult to say, especially for a non-realist. Historically, losses in science and technology are often the result either of economic problems which cause less incentive to preserve knowledge and technology, such as was the case in the decline of the Roman infantry soldier by the 4th century AD, or from particularly destructive political conflicts which results in people and places of learning being destroyed, such as may have been the case in the destruction of the library of Alexandria.[48] Similarly, as developments in new technologies replace old ones, knowledge of how those previous

[46] In theory, I could claim that it is not even bidirectional but multidirectional in the sense that there may be a loss in science and technology for a period of time followed by the invention of alternative inventions in a different direction, rather than a reinventing of technology now lost forever. One example of this might be the chemical ingredients for Greek fire, used in the early medieval period in Byzantium and beyond, the recipe of which is a well known historical mystery, and yet there have been several chemical agents which act similar, or even better, as incendiary weapons invented since then. (Pardington, 1960, 42)

[47] Western Roman legions of the 4th century were a far cry from the Marian and Augustian reforms half a millennium before---as the empire expanded, military spending meant that they could not equip each soldier with the classic rounded scutum, pilum, gladius and lorica segmentata of the early imperial era, and eventually legionaries were reduced to far less effective flat shields, a maile shirt and a spear (as seen in Cowen and McBride, 2003).

[48] Unfortunately, when exactly this occurred and who was responsible is also a controversial historical point (Lerner, 2001, 30).

technologies worked often declines.[49] If we want to compare this to apparent regresses in **moral** attitudes, which to recap for our purposes we are taking to mean a decline in adherence to principles of non-hypocrisy and avoidance of double standards, and/or an adherence to moral belief states constituted by beliefs based on factually inaccurate information, then I believe the answer is: yes, this does also happen. For example, we can and do observe widespread backlashes against progression in some particular direction politically and socially which may be to conserve previously held moral attitudes, many of which may be based on either blatant double standards or on scientifically inaccurate information. At the risk of sounding too topical and political, that may even be what happened in the United States in 2016 in what appears to have been a severe backlash against globalisation, multiculturalism, firearm legislation, feminism and lgbt rights following the Obama administration. This is also not a new phenomenon historically.

So in conclusion, I would argue that not only is moral regress possible within the morality fiction, but it has indeed happened on many occasions throughout history. This is usually the result of some radical change, whether it is political, social, economic or migratory and as a result a major breakdown of social and/or political cohesion between multiple groups or factions. Indeed, this is perhaps the story of all conflict(s) in a nutshell, even on a personal level, where some major change in circumstance that cannot be easily prepared for or some major misunderstanding results in some kind of conflict between friendships or relationships. However, moral progress is also possible in the sense that certain events and levels of mutual understanding can similarly foster greater degrees of cohesion and team work. On a broader historical level, I would argue, as Wilson does, that the trend of increased epistemological awareness has not only increased mankind's scientific and technological expertise, but also our moral capacity overall. To avoid becoming too Eurocentric, whilst there is very little that has happened in history that does not ever occur **somewhere** in the 21st century: slavery still occurs (illegally), as does ethnic cleansing, the wilful destruction of settlements, warfare, economic disparity and exploitation not to mention more common social problems like racism, sexism and religious prejudice, and crime too exists in abundance, none of these exist with anything like the same frequency or distribution as we see commonly in previous centuries, and, more importantly, the common **attitudes** towards these things seems to have

[49] It is questionable at best how many in the mainstream public today know how to use many historical household tools and appliances used as recently as the Victorian era given the invention of electricity, warm running water and gas central heating.

gradually changed. These things are no longer natural and necessary aspects of the human condition to be accepted or even glorified; rather they are seen and are portrayed (or projected) as evils to be actively kept to an appropriate minimum if not completely eliminated. The rise of international law has also had a profound impact on the relative **scale** of these issues. This is not to say that there are not entirely new problems that face modern people, such as the wilful apathy towards environmental responsibility in the wake of increased urbanisation, population and rate(s) of climate change, but I suspect that even such things as this are, like slavery, a simple matter of whether and how much people understand the issue, how relevant and significant it is to them, and if they are aware of or willing to cope with alternative ways of doing things as to whether or not they will feel motivated to do anything about it. One of the main limiting factors in any assignment of a moral principle within the morality fiction, such as legal autonomy or environmental responsibility, is in how well it is fully **understood** by those who have the ability to affect it. It is not enough to simply 'know' of a problem, one must fully 'understand' the gravity of the situation and what can be done about it. This is, after all, the purpose of **any** language, moral or not, fictional or not, to communicate and create understanding of information of potentially great importance.

Chapter 10: The Final Result

The aim of this thesis was to piece together an account of the kind moral language used by non-realists, that is, those who take either an anti-realist stance on the metaphysics of moral issues or are agnostic on those issues. To do this, I needed an account which deals with the semantic content of what the words in moral sentences literally mean, and I also needed one which deals with those elements concerning pragmatic discourse and the motivational reasons behind why this kind of language is used, in what context(s), what kind of information is conveyed in the process and in what form. I also wanted an account which deals with the emotional state(s) of those speakers when engaging in moral discourse, owing to an observation I have made over the years when looking at not only **what** people are saying when they engage in moral argument, but **how** they are saying it. I do not believe it is any coincidence that moral arguments are rarely made dispassionately, and I believe that emotion plays a vital role both in a cognitive sense in how moral thoughts are processed psychologically, and also in how these thoughts are communicated through moral language.

What I have attempted to provide, in answer to these observations and enquiries, is 'Expressive Fictionalism'. 'Expressive Fictionalism' is, in essence, a theoretical way of explaining how we may **create a story or narrative** about something we call 'ethics' in order to give meaning, direction, purpose and persuasive power to our emotions and our attitudes, even though, on inspection in the most maximised of critical contexts, it is but a perception of the nature of things we have attitudes towards rather than an observation of any natural properties beyond what we merely invent for our own purposes. This is not to say that the creation of morality is necessarily 'selfish'. The idea is that we use the morality fiction to communicate our feelings, and this may be of use to us, but equally it can also get us into a great deal of trouble[50]. It is, therefore, a marriage between fictionalist semantics, and pragmatic expressivism.

We began way back in section one with non-realism. This involved providing some convincing arguments giving us cause to doubt a realist interpretation of moral language (the

[50] I may bring voice to my objections towards something which everyone else believes is a good idea; I may be fully aware that my voice will not be listened to and that speaking out will only get me in trouble, but still conclude that speaking out is the right thing to do. Whether or not this 'right thing' is indeed a natural property of the thing in question is not the point, the point is that I **disapprove** psychologically, **therefore** it <u>feels</u> bad to stay silent, and good to publically condemn it, and that is the **motive** behind these moral utterances.

idea that moral adjectives refer to natural properties), and to give us some basis for considering a non-realist interpretation. We briefly looked at error theory (the idea that moral adjectives attempt refer to natural properties and that the speaker is somehow mistaken in doing this), and moved away from that theory owing to the belief that such an account, whilst straight forward, does not always help us explain why non-realists continue to use moral language despite their sceptical beliefs about the nature of morality. It also does nothing to explain the passionate nature of moral discourse that I mentioned earlier. What does help in all of these things, however, is **expressivism** (the idea that moral language is a means of communicating attitudes). However, we still needed a theory which accounts for what the content of a moral sentence translates to, and as we later discovered, when we apply expressivism semantically, issues start to arise, most notably the **Frege-Geach** problem.

Enter Simon Blackburn, our champion with a brand of super efficient and super economical semantic expressivism to get round the Frege-Geach problem: **'Quasi-Realism'**. The Quasi-realist project aims to circumvent the Frege-Geach problem by providing a platform for a form of logic of expressive utterances rather than a form of logic based on assertions, thus providing a means by which we can make valid inferences even when using premises that include unasserted contexts. What we also get from this is a fascinating story about the motivations behind non-realist moral discourse in the form of communication of attitudes of approval and disapproval. This is particularly useful, because it allows us to take something from the quasi-realist project even though, as we see later on, it is actually far more expensive than Blackburn and the quasi-realists initially expected. As we see in chapter five of section two in particular, in Blackburn's response to David Lewis's paper on quasi-realism and fictionalism, quasi-realism is not the same creature as fictionalism, but in highlighting the differences, Blackburn reveals a potentially fatal flaw in the quasi-realist project: in order to preserve the 'realism' it sacrifices its own expressivism in the process; the debate also reveals an inability to confidently demonstrate attitude **inconsistency**, and in pushing for an account of **moral truth**, Blackburn seems to abandon simple expressivism in search of something else, what exactly that may be becoming more and more nebulous as he does so. Nevertheless, the essence of the quasi-realist project, to provide a platform for explaining expressivist moral language and circumventing the Frege-Geach problem is extremely useful, because, as we see later on, it very much helped us out when we later combined it in a more pragmatic form with Joycean fictionalism.

Enter Joycean fictionalism then, a form of non-realist semantics in which we see a similar situation of moral utterances made in unasserted contexts, whereby there is the pretence or illusion of assertion. In this situation, there is no real need to demonstrate attitude inconsistency in the same vein as the quasi-realist is required to provide; it is enough to demonstrate theoretical inconsistency **within the context of the fiction**, and not to provide genuine inconsistency of assertions that are **not really being made**. This straightforward yet elegant solution provides some useful evasive manoeuvres away from the Frege-Geach problem. However, as I explain in section two, fictionalism is at the same time a frustratingly one dimensional approach in that all we are getting is some useful semantics and not a complete theory about what is actually being communicated when we speak about moral issues in the form of fictional utterances. It is one thing to say that the literal translations behind the words we use are as they appear, but they do not refer to anything that really exists in the same sense that scientific discourse might. Therefore, in terms of moral discourse in a fictional context, the semantics are largely dealt with, but the greater question of the meaning behind the **communicative act** of speaking in moral terms is what now interests us, and for that I refer back to **expressivism**.

What seems intrinsic to the **idea** of a moral principle by which we guide ourselves to act is to have an **attitude** (or at the very least some strong emotional feeling) towards whatever this principle is being applied to. To provide some evidence of this, rather than to proceed on this kind of conjecture, I spent a great deal of time in section three looking at the scientific relationship between emotions, behaviours and mental states associated with the theme of morality, for example altruism and empathy. I also examined what happens when emotional sensitivity is impeded and the effect that has on those same ideas of doing 'right' by others. Whilst I do not provide a purely necessary or logically deductible connection, the correlations that I investigated are, I claim, striking enough to suggest a causal link between emotional sensitivity and moral thought, and therefore by proxy, moral language. I then used this as a base upon which I began to construct an inferential process linking the sensing of others' emotions with approving and disapproving attitudes, and then preventative or reinforcing action in response to it. In doing so, I attempted to paint a logical picture of the kind of processes involved in a non-realist observing that which they approve or disapprove of, processing that and speaking out, for or against it, using the fiction of morality as a contextual platform. I then developed this idea of translating emotions into moral language via two models, one based on pride, shame and responsibility, the other on empathy and compassion,

the former of which relying on a complex web of emotions created via social and psychological conditioning. The idea is that a non-realist must be able to communicate their emotions in a way that is not only understood, but also listened to; else the mere act of communicating those attitudes becomes trivial at best, or worse completely irrelevant. To do this, the expressive fictionalist conveys not only approving and disapproving attitudes towards actions and emotional responses to them, but also a sense of emotional **cost** resulting from perceived transgressions. This cost manifests in criticism, social rejection, or in some cases real life punishment. Similarly, in an environment where the dominant, mainstream platform of moral discourse appears to adhere to the conventions of moral realism, then whether listeners are realist, non-realist or morally agnostic, that is the language convention that is already used, and the means by which important (by one measurement or another) decisions are made. Therefore, whether a non-realist or an agnostic believes in natural moral objects or properties or not, if they care on any level what happens to individuals, it still makes a great deal of sense to speak in moral terms, even if it is used merely in the context of a **useful fiction**. In other words, what I am saying is that whether or not there is some natural basis for moral thought and language, or whether it is more akin to a construct of our own imagination, giving form to our emotional pleasure or displeasure at certain behaviours or attitudes we observe, what is even more important is the way in which moral language is **used**, and the effect that it has on behaviour and on society more broadly, whether it is actually based on anything real or not.

This can be easily seen when we ask questions like, 'is moral progress a thing?' To speak as a strict non-realist, in a maximised critical context, the simple answer is either 'no', or at the very least, 'there is no evidence to suggest that it really is'. However, this does not mean that even such a non-realist is totally devoid of having an **opinion** on whether or not he/she **approves** of the way their world has been unfolding. Such a tragic lack of feeling or imagination would be unfortunate indeed. It would be akin to asking what a person thinks about a fictitious character's actions in a story, only to be met with, 'I don't care, they don't exist, and my opinion is irrelevant'. Clearly, the person being asked has completely missed the **point** of the question. In fiction, we speak 'as if' to reveal things about ourselves and about each other. If I say that Romeo and Juliet were victims of infatuation rather than love, then that tells you absolutely nothing about any real life couple in particular, but it tells you a great deal about my feelings on the matter of short vs. long term relationships. I argue that moral discourse operates in a very similar way. If I say that kicking that cat was 'wrong', that

does not really tell you anything about the nature of that act on any metaphysical level, other than that it was an event which took place. It does, however, tell you everything you need to know about my reaction to it, the political stance that I am taking, and potentially even the actions I intend to take in response to it, which is, I argue, far more useful and relevant information to communicate to you. Similarly, if I say that mankind has made great progress not only technologically, but also morally since the dawn of human civilisation, that tells you far more about me, who I am and what I believe in, than it does anything about human civilisation itself, which is, in reality, neither progressive nor regressive, neither good nor bad, it simply is. So, in essence, what we have here in this analysis of meta-ethics and moral language is a system of storytelling designed to tell the listener less about the thing in question and more about the **person** telling the story, what they believe in, what they value and what is important **to them**.

In the final analysis then, I would certainly argue that my original objective has been **fulfilled**. I have examined the case of non-realism, and I have investigated the various problems and major theories that exist on the subject and the strengths and weaknesses of those theories. In response, I have created a new theoretical approach designed not only to answer some of the most pressing questions in non-realism, and in a way that I do not believe has been attempted before or to this extent, but also to introduce new questions and a new avenue for future research. 'Expressive Fictionalism' is an account which deals both in the fictionalist semantics of non-realism and in the pragmatics of the communicative act of speaking up for moral attitudes from the position of an expressivist. It is also compatible with a scientific approach towards moral thought and moral language and in doing so combines armchair philosophy with evidence from empirical research. I argue that it succeeds in avoiding the Frege-Geach problem, and even more importantly succeeds in not compromising its own non-realism in the process. It provides us with a means of explaining non-realist moral language in a non-error theoretic way, as well as a platform for a creative narrative about moral improvement and moral progress. Finally, and perhaps most importantly, it is not a position **against** moral realism per se, but rather it is a position that **withholds judgement** on the existence of moral properties. There is nothing in my thesis which directly contradicts theories about the nature of ethics in either metaphysics or epistemology, but rather it stands apart from them, in no way relying on them, as a separate entity, and in doing so, I leave those theories about the metaphysical nature of ethics in the hands of other philosophers.

Bibliography

Almog, Joseph, and Paolo Leonardi. *Having in Mind: the Philosophy of Keith Donnellan*. Oxford University Press, 2012, p.79

Ayer, A. J. *Language, Truth and Logic*. Victor Gollancz, 1938, 103-111

Baldwin, T., 1985. "Ethical Non-naturalism," in Hacking 1985. 23–45.
——, 1990. *G.E. Moore*, Routledge: London and New York.

Balliet, Barbara J. *Women, Culture and Society*. Kendall Hunt, 1998, pp. 89-96

Batson, C.D., R.R. Klein, L. Highberger, and L.L. Shaw, 1995. "Immorality From Empathy-Induced Altruism: When Compassion and Justice Conflict," *Journal of Personality and Social Psychology*, 68: 1042–1054

Beaver, David I. "Presupposition." Stanford Encyclopedia of Philosophy. Stanford University, 30 July 2007. Web. 2 Mar. 2017.

Blackburn, Simon. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford Clarendon, 1984

Blackburn, Simon. 1993 *Essays in Quasi-realism*. New York: Oxford UP, 1993

Blackburn, Simon. *Ruling Passions: a Theory of Practical Reasoning*. Clarendon Press, 2009

Bloom, Paul. "Empathy and Its Discontents." *Trends in Cognitive Sciences*, vol. 21, no. 1, 2017, pp. 24–31., doi:10.1016/j.tics.2016.11.004.

Boole, George. *The Laws of Thought*. Prometheus Books, 2003, pp.3-4

Boring, Edwin G. (1950). *A History of Experimental Psychology*. New York: Appleton-Century-Crofts).

Burgess, John P. And Gideon Rosen, (1997). *A Subject with No Object*. Oxford: Clarendon Press.

Clark K.L. (1978) 'Negation as Failure'. In: Gallaire H., Minker J. (eds) *Logic and Data Bases*. Springer, Boston, MA

Cowan, Ross, and Angus McBride. *Roman Legionary: 58 BC - AD 69*. Osprey Publishing, 2003

Curtis, Robin; Elliott, Richard George (December 2014). "An Introduction to Einfühlung". *Art in Translation*. **6** (4): 353–376

Davies, Ella. "Humpback Whales Intervene in Killer Whale Hunt." *BBC News*, BBC, 9 May 2012, www.bbc.co.uk/nature/17991601.

Davis, M. and L. Kraus, 1997. "Personality and Empathic Accuracy," in *Empathic Accuracy*, W. Ickes (ed.), New York/London: Guilford Press, 144–168.

Dunaway, Billy, and Tristram Mcpherson. "Reference Magnetism as a Solution to the Moral Twin Earth Problem." *Ergo, an Open Access Journal of Philosophy*, vol. 3, no. 20190214, 2016, doi:10.3998/ergo.12405314.0003.025.

Egan, Andy. "Quasi-Realism and Fundamental Moral error1." *Australasian Journal of Philosophy*, vol. 85, no. 2, 2007, pp. 205–219., doi:10.1080/00048400701342988.

Ekman, Paul. "An Argument for Basic Emotions." *Cognition and Emotion*, vol. 6, no. 3-4, 1992, pp. 169–200., doi:10.1080/02699939208411068.

Funk, Angela Marie. "Epictetus, Stoicism, and Slavery." *University of Colorado, Boulder CU Scholar* (2011): 15-23

Gallese, V., C. Keysers, and G. Rizzolatti, 2004. "A Unifying View of the Basis of Social Cognition," *Trends in Cognitive Science*, 8: 396–403.

Gilbert, Daniel (1989). "Thinking lightly about others: Automatic components of the social inference process". In Uleman, J. S.; Bargh, *J.* A. *Unintended thought*. New York: Guilford. pp. 189–211.

Grice, Paul (1975). "Logic and Conversation". In Cole, P.; Morgan, J. *Syntax and semantics. 3*: Speech acts. New York: Academic Press. pp. 41–58

Grice, Herbert P. *Studies in the Way of Words*. Harvard Univ. Press, 1989, 22-40

Hare, R. M. *Freedom and Reason*. Clarendon Press, 1985.

Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford UP, 1977, 3-10

Hoffman, M., 1981. "Is Altruism Part of Human Nature?," *Journal of Personality and Social Psychology* 40: 121–137.

Hume, D. 'The Sceptic'. In David Hume, Selected Essays, edited by S. Copley and A. Edgar. Oxford: Oxford University Press, 1993

Hume, David. *A Treatise of Human Nature*. Createspace, 2016 (Origionally published 1738), pp.167

Jenkins, C. S. "Lewis and Blackburn on Quasi-realism and Fictionalism." *Analysis* 66.4 (2006): 315-38.

Jordon, Jeff. "Pragmatic Arguments and Belief in God." Stanford Encyclopedia of Philosophy. Stanford University, 15 March 2018. Web. 28 Feb. 2019.

Joyce, Richard. "Moral Anti-Realism." Stanford Encyclopedia of Philosophy. Stanford University, 30 July 2007. Web. 6 Feb. 2017.

Joyce, Richard. "Moral Fictionalism." *Philosophy Now* (2011): 287-313.

Joyce, Richard. *The Myth of Morality*. Cambridge University Press, 2007.

Landau, Russ Shafer (2010). *The Fundamentals of ethics*. Oxford University Press. P.292

Lerner, Frederick Andrew, Stuart B. Schimmedl, and Caroline F. Schimmel. *The Story of Libraries: From the Invention of Writing to the Computer Age*. New York: Continuum, 2001

Mackie, J. L. Ethics: Inventing Right and Wrong. Harmondsworth: Penguin, 1977.

Mackie, John Leslie. *Ethics: Inventing Right and Wrong*. Harmondsworth, Middlesex, Eng.: Penguin, 1977.

McPherson, Tristram, and David Plunkett. *The Routledge Handbook of Metaethics*. Routledge, 2018.

Neale, Stephen. "Descriptions." *Routledge Encyclopedia of Philosophy*, doi:10.4324/9780415249126-x011-1.

Ostertag, Gary (ed.). (1998) *Definite Descriptions: A Reader* Bradford, MIT Press. (Includes Donnellan (1966), Chapter 3 of Neale (1990), Russell (1905), and Strawson (1950)), pp.3-4

Partington, James Riddrik. *A History of Greek Fire and Gunpowder*. 1960, pp. 1-90

Pitman, R. L., Deecke, V. B., Gabriele, C. M., Srinivasan, M., Black, N., Denkinger, J., Durban, J. W., Mathews, E. A., Matkin, D. R., Neilson, J. L., Schulman-Janiger, A., Shearwater, D., Stap, P. and Ternullo, R. (2017), Humpback whales interfering when mammal-eating killer whales attack other species: Mobbing behavior and interspecific altruism?. Mar Mam Sci, 33: 7–58. doi:10.1111/mms.12343

Price, Jonathan, et al. "Emotional Side-Effects of Selective Serotonin Reuptake Inhibitors: Qualitative Study." *The British Journal of Psychiatry*, 2009, pp. 211–217.

Prinz, Jesse J. *The Emotional Construction of Morals*. Oxford Univ. Press, 2007, pp. 270-273

Postan, Michael M. *The Agrarian Life of the Middle Ages*. Cambridge University Press, 1966

Rader, Melvin. *A Modern Book of Esthetics: An Anthology; Ed. with Introd. and Notes*. New York: Holt, Rinehart and Winston, 1973.

Roojen, Mark Van. “Expressivism and Irrationality.” *The Philosophical Review*, vol. 105, no. 3, 1996, p. 311-335., doi:10.2307/2185703.

Ross, W. D., 1923. Greek text and commentary on *Aristotle's Metaphysics*, vols 1 and 2. Oxford: Clarendon Press.
——, 1928. Translation of Aristotle's *Metaphysics*. 2nd edition. For example, in Richard McKeon (ed.) *The Basic Works of Aristotle*. New York: Random House, 1941.

Russell, Bertrand. *On Denoting*. Blackwell, 1905, p. 479

Schueler, G. F. (1988) "Modus Ponens and Moral Realism." *Ethics* 98.3: 492-500

Shafer-Landau, Russ. *Oxford Studies in Metaethics*. Oxford: Oxford UP, 2011

Smyth, Nicholas. “Resolute Expressivism.” *Ethical Theory and Moral Practice*, vol. 17, no. 4, 2014, pp. 607–618., doi:10.1007/s10677-014-9495-y.

Stanley, J., 2001, “Hermeneutic Fictionalism”, in French and Wettstein (2001), pp. 36–71

Stanford Encyclopaedia of Philosophy (http://plato.stanford.edu/entries/fictionalism/#Bib) --- Fictionalism 4.2 (http://plato.stanford.edu/entries/nonexistent-objects/) --- Non-Existent Objects 3.2

Stevenson, C. L. (1937). "The Emotive Meaning of Ethical Terms". In Stevenson, C. L. Facts and Values. Yale University Press (published 1963), 83-141

Stueber, Karsten. "Empathy." Stanford Encyclopaedia of Philosophy. Stanford University, 2013. Web. 6 Feb. 2013.

Tacitus, Cornelius, and A. J. Woodman. The Annals. Indianapolis: Hackett Pub., 2004.

Uzgalis, William. "John Locke." Stanford Encyclopaedia of Philosophy. Stanford University, 24 Apr. 2017. Web. 24 May 2017.

Van Der Schaar, Maria, *G. F. Stout and the Psychological Origins of Analytic Philosophy*, Springer, 2013, p. 8

von Helmholtz, Hermann (1867). *Handbuch der physiologischen Optik. **3**.* Leipzig: Voss. Quotations are from the English translation produced by Optical Society of America (1924–25): Treatise on Physiological Optics.

Wayne Ambler, "Aristotle on Nature and Politics: The Case of Slavery," Political Theory 15, no. 3 (Aug. 1987): 390-410

Wilson, Catherine. "Moral Progress Without Moral Realism." *Philosophical Papers* 39.1 (2010): 97-116

Wittgenstein, L., 1953, *Philosophical Investigations*, Oxford: Blackwell, p. 350

Wright, C., 1988a, "Realism, antirealism, irrealism, quasi-realism," *Midwest Studies in Philosophy*, 12: 25–49.

Young, Liane, and Michael Koenigs. "Investigating Emotion in Moral Cognition: a Review of Evidence from Functional Neuroimaging and Neuropsychology." *British Medical Bulletin*, vol. 84, no. 1, 1 Dec. 2007, pp. 69–79.