Moral Disagreement – A Psychological Account and the Political Implications

Peter Caven

Thesis submitted for the degree of

Doctor of Philosophy

(Philosophy)

March 2015

Department of Philosophy

The University of Sheffield

Abstract

Moral disagreement is not only a philosophically interesting matter in its own right, but is also a highly important social and political issue. Historically, moral disagreement has often led to instability, conflict and persecution. It's easy enough to see how a society of individuals with a consensus on important moral issues can forge a common life together, but it is more difficult to determine how cooperation is to be attained whilst disagreement on key matters persists. If we are to understand how best to face up to moral disagreement, it is vital that we get to grips with what exactly we are dealing with, and what options are feasible.

In this thesis, I articulate a particular conception of what moral disagreement involves, its extent, and what its root psychological causes are. I further demonstrate that this conception has important implications for a contemporary debate within liberal political theory. In so doing, I advocate a descriptive form of value pluralism, whereby individuals are typically committed to a range of distinct values which they implicitly take to have independent normative force. I suggest that individuals both across and within cultural groups weight such values differently, leading to fundamental moral disagreement.

This is explained by my proposed Two-stage Enculturated Affect (TEA) model of moral psychology, whereby values are cultural constructs which gain their perceived normative force through their relation to an agent's affective dispositions. Such affective dispositions are the product of both shared and non-shared environmental influences, as well as our particular genetic endowment. They thus differ between individuals, leading to differences in value-weighting.

My account implies that moral disagreement is both inevitable and often a consequence of affective variation, rather than any defect in either disputant's reasoning. This may present a problem for accounts of liberalism which rely on a consent-based account of political legitimacy. In the final part of my thesis, I show that my TEA model's account of moral disagreement can be drawn upon to help reinforce a modified version of John Rawls's political liberalism.

Acknowledgements

I would like to begin by thanking the Arts and Humanities Research Council for funding my studies. Without the financial support which they offered, this thesis would not have been possible.

Next, I would like to thank my supervisors – Yonatan Shemmer and George Botterill – for their guidance and advice during the course of my PhD. Yonatan has always been extremely supportive and encouraging whilst remaining critical on unclear or erroneous points, which has helped immeasurably in refining my project. George's keen eye for detail and regard for the wider issues at stake has improved the flow and rigour of my thesis no end.

Sheffield's philosophy department has an intellectually stimulating and socially vibrant community which I have been extremely lucky to be a part of, and I'd like to thank all the staff and students for making it such an amazing place to work and study. Special thanks to Charlotte Alderwick, Al Baker, Jessica Begon, Carl Fox, Rich Healey, Stephen Ingram, Katharine Jenkins, Tash McKeever, Jonathan Parry, Angie Pepper, Joe Saunders, Robin Scaife, Jack Wadham, Neil Williams and Stephen Wright.

Parts of this thesis have been presented at various conferences, and the questions and comments which I received in response have helped shape my arguments. I would therefore like to thank the audiences at: MANCEPT Workshop on Moral Conflict and Dirty Hands (University of Manchester); The White Rose Philosophy Postgraduate Forum (Universities of Hull, Leeds, Sheffield and York); Pluralism and Conflict: Distributive Justice Beyond Rawls and Consensus (Fatih University, Istanbul); Forum Scientiarum 'The Evolution of Morality' (Tuebingen University) and Understanding Value III (University of Sheffield). I'd especially like to thank the audiences at the Sheffield Philosophy Graduate Seminar, who have been subjected to most of this thesis over the course of my studies and have always offered interesting and useful contributions.

Finally, I would like to thank my friends and family for supporting me over the last four years. Special mention to Charlotte, Jess, Tash and Tom K, as well as Mum, Dad and Matt. You're the best.

Contents

Introduction

Some believe that abortion is a form of murder whilst others hold that it is no worse than removing a tumour. Some regard capital punishment as barbaric and unjustifiable whereas some take it to be the only just means to punish certain offenders. Some denounce a range of sexual practices as unnatural and perverted, while some claim that nothing that goes on in the bedroom between freely consenting adults can possibly be wrong. Moral disagreement permeates our own society, and is even more extensive when we factor in disagreements between those inhabiting different societies.

We take our moral principles seriously, and although we may reflect on whether or not they apply in particular situations, we rarely doubt that they are fundamentally correct. Thus, when we encounter those who profess different moral views which are incompatible with our own, we find it a troubling experience. We may try to reason with our disputants to bring them around to our way of thinking, whilst they do the same with us. Considerations which the other has overlooked may be brought to bear, and in some cases this will prove effective in changing their mind. However all too often it will lead to a deadlock, typically leaving both parties frustrated yet stalwart in their conviction that their opponent has made a mistake of some kind or other. What are we to make of these situations? How do we explain them from both a conceptual and psychological point of view, and how might we live together in light of them? Answering these questions is the business of this thesis.

In the first part of the thesis, I begin by discussing the phenomenology of moral conflict, and how this underpins an interpretation of moral disagreement as arising from differences in how individuals weight distinct values. Next, I take an empirically informed approach in order to advance our understanding of the scope, extent and psychological causes of such moral disagreement. In the last part, I demonstrate how my conclusions can inform a contemporary debate within political philosophy. My thesis is divided into six chapters, each of which I will put into context and briefly outline below.

Chapter 1 - Moral Conflict and Descriptive Value Pluralism

In Chapter 1, I discuss the phenomenology of moral conflict and argue that it implies a form of descriptive value pluralism. Value pluralism is usually adopted as a metaethical or normative thesis concerning the ontological multiplicity of moral values. I argue for the different but related claim that, purely as a matter of descriptive psychology, people are in fact committed to a range of values, each bearing distinct normative force.

Moral conflict is an intrapersonal phenomenon, which occurs when agents find themselves in a situation where they feel that there are good moral reasons for acting in two (or more) incompatible ways. This is usually understood in terms of individuals feeling there to be conflicting moral duties, obligations or values at stake. A moral dilemma, on the other hand, is an instance of moral conflict where one takes there to be no candidate for a right action; all potential actions are taken to be morally impermissible. Many cases of moral conflict will not constitute full blown moral dilemmas. In most instances, the agent might feel conflicted but can nonetheless judge one course of action to be best, all things considered.

Recognition that situations can exist in which an agent faces conflicting moral considerations predates recorded ethical philosophy itself, as evidenced from examples in Homeric legends. Such examples of situations of moral conflict also feature in the works of the major Ancient Greek philosophers. In *The Republic*, Plato has Socrates invoke an example wherein one has borrowed a weapon from a friend, who subsequently develops murderous insanity and asks you to return his weapon.[1] This example represents a conflict between promise keeping and preventing murder. Aristotle, whilst discussing voluntariness, discusses the case of a man whose family is in the power of a tyrant and is forced to perform base acts in order to ensure their safety. More recently W.D Ross, Isaiah Berlin and Bernard Williams cite the existence of moral conflicts in their arguments for normative value pluralism.

Chapter 1 will take a slightly different tack to these latter thinkers, and argue in favour of *descriptive* rather than normative value pluralism. I will argue that individuals are indeed prone to experience a disquieting sense of moral loss after having resolved moral conflicts when distinct considerations are at stake. This, I suggest, is most especially so when agents are faced with genuine moral dilemmas. However, even when they are satisfied that there is an all-things-considered morally right resolution to a moral conflict, they still undergo a disquieting phenomenology, which Steven De Wijze, dubs 'tragic remorse'. I argue this indicates that

[1] Plato, *The Republic*, 331d

individuals implicitly take a range of considerations to be of distinct moral value, yet that they also typically manage to commensurate values in a particularist manner to deliver an all-things-considered best judgement. In order to resolve moral conflicts, then, individuals assign distinct considerations different levels of overall normative weight, in a context sensitive manner.

I emphasise that this is merely a psychological and descriptive rather than a normative or metaethical claim, and thus doesn't necessarily prove normative or ontological value pluralism. The main import of this chapter is the model of moral judgement which it entails. For the truth of descriptive value pluralism implies that one way in which individuals may morally disagree is through differing in the amount of normative weight that they assign to distinct values. This would lead individuals to commensurate values differently, and thus come to different resolutions to moral conflicts. Such an interpretation of moral disagreement will inform my following chapters.

Chapter 2 - Intercultural Fundamental Moral Disagreement

Chapter 1 left open the question of whether individuals do in fact weight and commensurate values differently from each other or not, which requires an empirical outlook for its answer. In Chapter 2 I move on to discuss intercultural moral disagreement, and survey evidence that it does indeed result from differences in value weighting.

Most of the discussion surrounding moral disagreement in the philosophical literature has continued to focus on its implications for metaethics. Specifically, a common anti-realist argument is that moral disagreement between individuals and cultures is indicative of there being no objective moral truth. Most famously, J.L. Mackie applied the 'Argument from Relativity' as one of his two main arguments in favour of his moral error theory.[2] Mackie's formulation of the argument from relativity has been extremely influential, with many counter arguments and reformulations based on the original notion having followed.[3]

Chapter 2 avoids discussing the metaethical implications of moral disagreement in favour of concentrating on establishing the depth and breadth of the phenomenon itself. I intend to

[2] Mackie, J. L., *Ethics: Inventing Right and Wrong*, London: Penguin Books (1977)
[3] See, for instance, Tolhurst, W. 'The Argument from Moral Disagreement', *Ethics,* Vol.94. No.3 (1987) pp.610-621 and McGrath, S. "Moral Disagreement and Moral Expertise" in Shafer-Landau, ed. *Oxford Studies in Metaethics* 3: (2008) pp. 87-107

demonstrate that cultural groups typically do recognise a similar range of values, but that moral disagreement sometimes stems from them weighting such values differently. In this sense, moral disagreement between cultural groups is sometimes *fundamental.* Fundamental moral disagreement involves two agents who are situated in the same context, possess the same non-moral understanding and are both free of errors of inferential reasoning, yet come to different judgements concerning what is the morally best course of action, all-things-considered.[4]

I aim to substantiate the existence of fundamental moral disagreement between distinct cultural groups, whilst also maintaining that such disagreement occurs between groups which tend to recognise a broadly similar range of values. First I draw on various anthropological examples to illustrate the kinds of moral disagreement – and moral similarities- that we often discern between cultural groups. I then discuss various 'defusing explanations' of such disagreement, which attribute it to factors such as non-moral disagreement, and thus interpret it as only apparent rather than fundamental. I cite various studies which purport to show the existence of fundamental moral disagreement, such as those by Richard Brant, Richard Nisbett and Kaiping Peng, drawing on a recent paper by Alexandra Plakias and John Doris. I then argue against those such as Michele Moody-Adams that more sophisticated formulations of defusing explanations can account for the moral disagreement that these studies reveal.

Chapter 3 - Intracultural Fundamental Moral Disagreement

To many readers, I expect it will come as no surprise that some moral disagreement between cultural groups can be established as being fundamental. However, in Chapter 3, I argue for a less familiar and more radical claim: that moral disagreement is also sometimes fundamental *within* cultural groups. I thus take the position that fundamental moral disagreement is more pervasive than others have suggested.

Discerning whether such disagreement can also be conceived of as fundamental or not is an area of research which has been relatively overlooked, yet is key to determining the ultimate

[4] Moral disagreement of this sort is not necessarily intractable. Two agents who are initially in fundamental moral disagreement might come to weight moral values differently over the course of their life and eventually come to an agreement. However, the relevant point is that this agreement cannot be reached merely by ensuring that both agents share the same context, non-moral understanding and are free of inferential errors in reasoning. It can only be reached if one or both parties revise their fundamental commitments regarding the relative importance of distinct values.

nature and cause of moral disagreement. Here I acknowledge that fundamental moral disagreement is less extensive within cultural groups, especially those which are narrowly defined. But, there are good reasons to believe that it does emerge within the context of such groups, and that its true extent is masked by other factors.

I will scrutinise the sort of everyday moral disagreement within broadly defined, national cultural groups, such as those disagreements concerning the moral status of abortion and capital punishment. I argue that the previously discussed defusing explanations, such as non-moral disagreement, may apply in some such cases. Nonetheless we have good reason to suspect that they do not in all, and are sometimes best explained in terms of differences in value-weighting. Next, I move onto examine moral disagreement within more narrowly defined cultural groups. I concede that such disagreement is less well documented than within broader groups, and that it is indeed less extensive. Nonetheless, there are other factors which might explain this relative moral homogeneity other than there actually being a moral consensus within such groups. Moreover, there is ample anecdotal experience of disagreement with those who share a similar background to ourselves, and moral disagreement found amongst ethical theorists within philosophy represents a particularly good candidate for being fundamental. Finally, I point to the results of moral judgement surveys to provide further proof of intracultural fundamental moral disagreement.

Chapter 4 - The TEA Model of Moral Psychology

The first three chapters aim to motivate descriptive value pluralism and the claim that individuals both between and, to a lesser extent, within cultural groups differ in value weighting. However, this does not explain why fundamental moral disagreement occurs. Without an understanding of the psychological mechanisms underlying moral disagreement, we are still at a loss as to the full extent of differences in value weighting, how it comes to be and whether it can ever be eliminated. The answers to these questions are not only of interest in their own right, but will also determine the precise influence that moral disagreement should have on our normative theory and practice. Chapter 4 focuses, then, on developing my own empirically grounded account of moral psychology - the Two-stage Enculturated Affect (TEA) model.

For much of the history of 20th century, empirically informed psychology has had little influence on moral philosophy. This state of affairs has recently changed, and now both

philosophers and psychologists are engaging in interdisciplinary experimental research which aims to shed light on many aspects of moral judgement. The earliest prominent example of experimental psychology applied to moral judgement can be found in the works of Lawrence Kohlberg. His work emphasised the importance of the development of explicit, conscious reasoning capacities in shaping the moral judgement of individuals. More recently, advancements in moral psychology have highlighted the extent to which much of our moral judgement is a product of automatic, affective processes as opposed to deliberative reason.

Chapter 4 advances a psychological account of the bases of moral judgement which is in line with this contemporary focus on affective primacy in moral judgement. I draw from both Shaun Nichols' and Jonathan Haidt's accounts of moral psychology, but ultimately conclude that each is incomplete when considered in isolation. Thus, insights from both must be combined into a single model. I thereby propose my own TEA model of moral psychology, which combines the fundamental insights from each account and offers a more complete picture of moral psychology. This holds that our affective dispositions are to some extent innate and the product of evolutionary process, and these help shape our moral values, but that these dispositions themselves are somewhat malleable in the face of cultural influences. Moreover, the extent to which a particular moral value is cherished within a particular cultural group helps determine the extent to which a particular norm will originate, survive the process of cultural evolution, and further be treated as morally salient by those within the cultural group.

Chapter 5 – Explaining Moral Conflict and Disagreement

Chapter 5 argues that my proposed TEA model of moral psychology best explains the conclusions of chapters 1, 2 and 3, when compared with competing accounts. Whilst chapter 4 makes the case for the TEA model being well supported by empirical evidence, it does not consider alternative accounts of moral psychology. The argument in its favour is not conclusive until such competitors are considered. Moreover, chapter 4 has little to say concerning moral disagreement, per se, and more is needed in order to demonstrate the relevance of the TEA model for the purpose of my thesis. Chapter 5 aims to demonstrate the plausibility of the TEA when compared to competing accounts by demonstrating that it is not only better supported by the available empirical evidence, but has more explanatory resources at hand to account for the phenomenology of moral conflict and moral disagreement. This will not only buttress the case for my proposed TEA model of moral

psychology, but provides an explanation of the root psychological cause of moral disagreement.

I begin by articulating the two competing accounts of moral psychology which I will consider: Jesse Prinz's Emotional Constructivism and John Mikhail's Universal Moral Grammar. The former agrees with the TEA model that our emotional dispositions are responsible for moral judgement, but holds that they are almost infinitely malleable in the face of cultural influences. The latter, meanwhile, maintains that moral judgement is a product of an innate, dedicated faculty of moral reasoning.

I further point out various reasons for preferring my TEA model to either model, independent of their explanations of moral disagreement. Next, I demonstrate that the TEA model can better explain the phenomenology of moral conflict and tragic remorse than Universal Moral Grammar. I show that the TEA model can also neatly explain the pattern of intercultural moral disagreement, whereas Universal Moral Grammar struggles to explain variation in moral values and Emotional Constructivism finds it more difficult to accommodate intercultural moral *similarities*. Finally, I illustrate how the TEA model explains intracultural moral disagreement to a more satisfactory degree than either Universal Moral Grammar or Emotional Constructivism.

Chapter 6 – Implications for Political Liberalism

Having developed a descriptive account of moral deliberation, moral disagreement and the psychological mechanisms which underpin them, we are now equipped to address the question of what implications this has for normative practice. The answer to this depends upon the extent to which one takes descriptive moral psychology to properly bear upon matters of 'ought' – some might argue that the very fact that we can explain and understand the source of fundamental moral disagreement is normatively irrelevant. Nonetheless, when we are considering how we should live together whilst respecting the diverse judgements of individuals, then our particular conceptualisation of moral disagreement becomes salient. In chapter 6 I will thereby demonstrate how the TEA model of moral psychology and its accompanying account of moral disagreement can inform political philosophy. I will concentrate on the implications my account has for liberalism, being both the dominant

strand of contemporary Western political theory and one for which moral disagreement is particularly salient.[5] I briefly clarify why this is the case below.

What unites the various articulations of liberalism is a fundamental concern for protecting the individual freedom of citizens, combined with the recognition of the necessity of endowing a central political authority with coercive power. In doing so, all forms of liberalism must offer a means by which individual freedom can remain unthreatened by the prospect of state coercion. The vast majority of contemporary liberal theorists do so by drawing on the consent-based notion of legitimacy from the social contract tradition.[6] The basic idea is that laws are legitimate, binding and do not violate freedom just so long as all those subject to them can be taken to have consented to them in some form. Liberal theorists typically recognise that attaining the actual consent of all citizens for any law is impossible, and so too stringent a requirement for legitimacy. Instead they suggest that it is enough that citizens can be said to *hypothetically* consent to the laws which govern their society.[7] That is, coercive power is legitimate if agents would consent to the political authority under certain idealising conditions, such as being fully informed, consistent and instrumentally rational.

Insofar as most contemporary accounts of liberalism are grounded by this kind of account of legitimacy, the problem posed by the prospect of inevitable fundamental moral disagreement is clear. Not only does it imply that no political authority will ever become an object of actual consensus amongst citizens, but it also suggests that none could even hypothetically enjoy universal endorsement. For the principles which justify any political system are ultimately based on a conception of the relative weight of values. Thus, fundamental disagreement concerning the proper weightings of value implies that for any and each proposed political system, some individuals will reject it as incongruent with their value system. Since this disagreement is fundamental rather than merely apparent, liberal theorists cannot sidestep the issue by insisting that all would ultimately consent to a particular authority under certain hypothetical conditions. For even if we make idealising assumptions which abstract away citizen's constraints on knowledge, rationality and consistency, we can

[5] Of course, the psychological claims I have made thus far also have implications for other forms of political theory, but cashing these out would be beyond the scope of this thesis.

[6] This is not to say that all forms of liberalism share these moral foundations. For instance, on one plausible reading of Mill's liberalism, the ultimate justification for privileging liberty is in its instrumental role in terms of maximising utility. More recently, Joseph Raz notably avoids grounding legitimacy upon consent. (Raz, J. *The Morality of Freedom,* Oxford: Oxford University Press (1986)). I will not discuss the implications of the TEA model of moral disagreement for these accounts of liberalism.

[7] Another way to go is through following the Lockean tradition of grounding legitimacy upon tacit consent which, again, I will not discuss here. See Locke, J. *Second Treatise of Government,* Chapter VIII

still expect there to be disagreement about which principles should regulate how society is organised.

In chapter 6 I will assess whether my contention that fundamental moral disagreement is an inevitable consequence of our moral psychology undermines liberalism. Ultimately, I argue that despite the apparent threat that the TEA model's account of moral disagreement poses for political liberalism, it can actually buttress one important variant. I suggest that the TEA model can substantiate John Rawls' claim that individuals are subject to the 'burdens of judgement', aspects of human reasoning which make fundamental moral disagreement between reasonable people inevitable. Rawls rightly claims that widespread recognition of the burdens of judgement amongst reasonable people is crucial if political liberalism is to be practically realisable. However, his presentation of them is, I argue, inadequate. Appealing to the TEA model thereby helps to ameliorate a severe deficiency in Rawls's political liberalism. I further argue that accepting the implications of the TEA model might also require modifying certain other aspects of Rawls's theory in line with political realism, a form of political theory which has long accepted the inevitability of moral disagreement.

Chapter 1: Moral Conflict and Descriptive Value Pluralism

This chapter focuses on the phenomenology of moral conflict and the implications that this has for descriptive accounts of moral thought. I argue that in certain situations individuals tend to experience an internal struggle between incompatible reason-giving factors. I suggest along with Bernard Williams that acting in such situations of moral conflict will typically leave agents with a sense of moral loss. This is indicated by the feeling of what Stephen De Wijze calls 'tragic-remorse'. Utilitarianism and Kantian deontology account for this experience by ultimately attributing it to error and confusion, insofar as they are committed to a monist account of value. Value pluralism, meanwhile, provides an account of moral judgement and the underlying structure of value that informs it which is more consistent with our experience of moral conflict. It therefore serves as a good basis upon which to model moral disagreement between individuals in the following chapters of my thesis.

It is important to reiterate from the outset that I do not aim to advance a direct philosophical argument against the normative legitimacy of monist theories of value, such as utilitarianism and Kantian deontology, in favour of normative value pluralism. In contrast, I mean only to establish value pluralism as the theory of value which provides the best account of our underlying moral thought. Whilst I will frame the issue in terms of the challenge that moral conflict presents for utilitarianism and Kantian deontology, it is only the descriptive component of these theories which I take direct issue with. Such theories typically hold that deep down, we take only one consideration, such as happiness, to be genuinely reason-giving. I mean to deny this descriptive claim, and vindicate the pluralist view, whereby we are commitment to a range of independently reason-giving values which sometimes prove incompatible with one another.

Value pluralism thereby accounts for our disposition to experience multiple, distinct moral motivations which may come into conflict, without attributing this experience to error or confusion on the part of the agent regarding what they implicitly take to be reason-giving. This, I hold, makes it preferable from a psychological descriptive perspective, although it is still open for the advocate of a monist theory of value to argue that we should nonetheless

only take a single value to be genuinely reason-giving on the normative level. As an aside, I do think that such a strategy is problematic, and that my argument could be developed to directly threaten normative value monism, but this is not the goal of this chapter. I only mean to argue that *descriptive* value pluralism offers the most accurate depiction of our underlying moral thought. This justifies my adoption of the conceptualisation of moral disagreement which it implies in later chapters.

In section 1, I discuss the value monist moral theories of utilitarianism and Kantian deontology and why they dismiss the possibility of genuine moral conflict. Section 2 articulates Williams's account of the experience of moral remainders in cases of moral conflict, and how this poses a threat to such monistic theories of value, at least from a descriptive perspective. In section 3 I further elaborate on this phenomenology of moral conflict via De Wijze's concept of tragic remorse. I argue that the particular character of this phenomenology suggests that individuals do not take actions in situations of moral conflict to necessarily be wrong, but only involving of distinct moral loss. Section 4 will buttress this claim, by appealing to the intuitive distinction between what counts as an appropriate decision making procedure for conflicts between instantiations of the same moral value and for conflicts between distinct values. In section 5 I will turn to value pluralism, as articulated by Isaiah Berlin, as an alternative to monist theories of value and argue that adopting a descriptive account of pluralism offers us a better means to explain the phenomenology of tragic remorse. Section 6 will then discuss one aspect of value pluralism which isn't coherent with our moral phenomenology – that of strong value incommensurability. Finally, section 7 will suggest a form of pluralism whereby people do commensurate values in a particularist manner, and spell out how this paves the way for fundamental moral disagreement.

1. Moral Conflict and Value Monism

As mentioned in the introduction, there has being recognition of apparent moral dilemmas and the internal conflict which they provoke in both the ancient and modern world. However, the influential moral theories of utilitarianism and Kantian deontology do not take these phenomena to be particularly revealing of the nature of morality. Proponents of these theories hold that ethical conduct consists in adherence to a single principle. In the case of act utilitarianism, the ultimate moral principle is that one must act in order to maximise happiness. Meanwhile, Kantian deontology holds that we must always act according to the

categorical imperative; we may only permissibly act "in accordance with that maxim through which you can at the same time will that it become a universal law."[8]

The implication of these systematic, principle-based approaches to moral theory is that moral conflicts (and especially moral dilemmas) cannot really exist. For one thing, on the practical level, one of the major aims of these moral theories is to offer determinate action-guiding solutions in situations where the right course is seemingly unclear. Thus, if it were to turn out that there were situations where their principles offered no guidance then they would fail to serve their purpose. However this is supposedly incidental: proponents of these theories do not admit to denying moral conflict out of mere practical convenience. Rather, the fundamental reason that Kantian deontology and utilitarianism do not accommodate the existence of genuine moral conflict lies in the fact that these theories are underpinned by a *monist* conception of moral value. That is, that they consider all moral rightness to be ultimately reducible to a single consideration; for the utilitarian, happiness, for the Kantian, the requirement of rationality, as expressed by the categorical imperative.[9]

For the sake of practicality, advocates of these theories might well suggest keeping distinct moral considerations aside from the ultimate principle in mind. For instance, Kant suggested that multiple imperfect as well as perfect duties could be derived from the categorical imperative; the latter being duties which we are always morally required to fulfil, the former duties which merit praise if fulfilled, but do not merit blame if neglected, at least on some occasions. Meanwhile, rule utilitarians hold that maximising happiness in the long term is best assured by adhering to a code of conduct including rules such as prohibitions upon lying, theft and violence, even in instances where these kinds of action might promote happiness in the short term. There are certainly situations in which such derivative rules could conflict with one another, such as it being necessary to lie or steal to prevent violence, and thus adherents of a monist theory of value might indeed face apparent dilemmas from time to time. However, these conflicts between secondary rules can and should in theory be resolved through proper application of the ultimate principle. As the prominent utilitarian theorist

[8] Kant, Immanuel. trans. Gregor, M *Groundwork for the Metaphysics of Morals* Cambridge: Cambridge University Press (1998) p. 31

[9] Kant offers several formulations of the categorical imperative, but holds that these are merely different expressions of a unitary principle. Some, such as Thomas Hill, have argued that we should nonetheless interpret Kant as an ethical pluralist who acknowledges a variety of moral concerns. (Hill, T.E, 'Kantian Pluralism' *Ethics,* (1992) pp.743-762) On that matter, it can be claimed that utilitarianism can too be pluralistic, insofar as it holds there to be various distinct and incommensurable forms in which happiness takes. I will not pursue the possibility of Kantian or utilitarian pluralism, except to note that these variations of the theories are not my target in this chapter.

J.J.C Smart argues, such considerations represent mere 'rules of thumb' which one should be willing to discount in favour of realising the primary value in situations of conflict. To adhere to a rule even when doing so would clearly fail to maximise utility would to be guilty of "surreptitious rule worship".[10]

Thus, for the value monist any moral conflict we encounter (and especially moral dilemmas) can only be attributed to error on our part; a failure to properly understand what is ultimately required of us. For if there is a single source of moral value from which all other considerations derive their moral significance, then in any particular situation there will be one course of action which will more fully realise this value than any other. For instance, the utilitarian holds that in any particular choice between preventing violence and lying or cheating, there is one option which will maximise happiness and this constitutes the right action. In many such cases it may be extremely difficult to discern which action this is, given the huge range of ways an action might potentially increase or reduce happiness in the long, short and medium term. Nonetheless, the utilitarian conceives this as an epistemic problem rather than a consequence of a genuine conflict taking place.[11] Similarly, Kantianism, with its focus on rationality as the ultimate arbiter of moral rightness, suggests that in any apparent conflict there can only ever be a single action that is prescribed by the categorical imperative as our moral duty. In Kant's words, "A conflict of duties and obligations is inconceivable [for] the concepts of duty and obligation as such express the objective practical necessity of certain actions, and two conflicting rules cannot both be necessary at the same time: if it is our duty to act according to one of these rules, then to act according to the opposite one is not our duty and even contrary to duty."[12]

---

[10] Smart, J.C.C 'Extreme and Restricted Utilitarianism'*, The Philosophical Quarterly*, (1956) pp. 344-354

[11] There could conceivably be a situation in which two different actions would both produce exactly the same amount of happiness. Yet such a situation would presumably not constitute a dilemma as such, as both options would be permissible rather than impermissible. Moreover, these types of situation would be extremely rare and would only constitute a tiny proportion of those which people commonly interpret as morally problematic, such as the examples mentioned above.

[12] Kant, I. *The Metaphysical Elements of Justice:* Part 1 of *The Metaphysics of Morals*, trans. John Ladd Indianapolis: Bobbs-Merill, (1965) pp.24-25. It might yet seem that acting solely in accordance with the categorical imperative could still result in situations of moral dilemma. For instance, we might find ourselves in a situation where we have no choice but to break one of two promises we have made through no fault of our own. Nonetheless, it would seem from this quote that Kant discounts this possibility; presumably, in any potential situation, one action would not actually constitute a violation of the categorical imperative, despite the fact that it seemingly would do so. Regardless, even if a revised Kantianism could accommodate these sorts of dilemmas, this does not undermine my argument; the salient point is that it does not recognise situations which most people experience as morally problematic as involving genuine moral conflict.

Both these moral theories are notorious for their tendency to advise courses of action in certain situations which seem counterintuitive and even morally repugnant. Utilitarianism is often criticised on the grounds that it implies that one should always be willing to sacrifice the well-being of a single individual if it will bring about a larger total sum of happiness for the many. Thus, it would seemingly suggest that we are morally required to torture an orphan child if it would satisfy the desires of many sadists, assuming that it were possible to assure that this act would not become public knowledge or diminish utility in the long term in some other manner. In contrast, Kantianism advises agents to refrain from acts which violate the categorical imperative, such as lying or theft, even if such acts would bring about vastly preferable consequences. The classic example which Kant himself discusses involves an axe-wielding murderer arriving at your door and asking where your friend is, who is currently hiding inside your house. Kant explicitly prohibits lying to the murderer to prevent your friend being killed because this would inhibit their capacity for rational autonomy, regardless of how clear it is that they would (under Kant's understanding of autonomy) fail to exercise this capacity by going on to commit murder.[13]

Many moral philosophers have rejected both utilitarianism and Kantianism as inconsistent with our moral intuitions on the basis of such examples. Advocates of these theories, on the other hand, go to great efforts to either prove that they do not in fact deliver such counterintuitive moral requirements, or else reject the reliability of those intuitions which prove inconsistent with their preferred theory. Nonetheless there is a related and yet more fundamental complaint about value monist act-evaluations than their failure to deliver intuitively satisfactory guidance in particular circumstances. This is that all such theories of value do not do justice to our underlying moral thought under conditions of moral conflict; even in situations when they offer what we might consider to be the right advice, they fail to account for the sense of distinct *moral loss* which we experience.

In the next section, I will go on to discuss this objection. But first it is incumbent upon me to explain the distinction between two different yet related levels of moral thought which I appeal to throughout this chapter. The first level, which I shall refer to as the *explicit level,* is constituted by our most basic, pre-theoretical moral phenomenology. This encompasses our intuitions concerning the moral appropriateness of both our actions and feelings. For instance, the instinctive judgement that killing an innocent child for no good reason is wrong,

[13] Kant, I. 'On a supposed right to lie from philanthropy' in Gregor, M. (ed. and trans.) *The Cambridge Edition of the Works of Immanuel Kant,* trans. Cambridge: Cambridge University Press (1996) pp.605-615

or that to take pleasure in killing a child, even if it were somehow deemed morally justified, is abhorrent, are widely shared aspects of our explicit level of moral thought. This contrasts with what I shall refer to as the *underlying level*. We do not have such immediate conscious access to the underlying level of our moral thought, but we can infer its content from our explicit moral phenomenology. For instance, since people generally judge all situations wherein an individual harms another for no good reason as morally wrong on an explicit level, we can infer that one widely shared component of our moral thought is that, all else being equal, harming others is wrong. We need not be consciously aware that we are implicitly committed to this principle on an underlying level for it to be an important aspect of our moral thought which shapes our ethical intuitions on the explicit level.

2. Moral Remainders and the Challenge to Value Monism

The notion that situations of moral conflict as such threaten moral theory was first explicitly acknowledged by Sir David Ross in *The Right and The Good,* published in 1930.[14] He argued that certain situations present us with two or more conflicting moral duties which, if encountered in isolation, we would be required to act upon. However, he also suggested that in each particular situation there is a solitary duty which overrides those which it conflicts with and which we are morally obliged to fulfil. In explaining this, he introduced the notion of *prima facie* duties. A *prima facie* duty has moral force, and may appear to the agent to be obligatory to satisfy. Yet when two or more *prima facie* duties come into conflict with each other, denying us the possibility of satisfying them all, only one truly constitutes an *actual* moral duty. He accepted that there was no determinate manner of ranking these duties which would apply to every situation, and that no duty was ultimate, entailing that no formal rules of deliberation over which duty took precedence could be produced. Nonetheless, he suggested that reflection could enable individuals to better determine which duties were actual and which merely *prima facie*. Whilst this may seem to offer scant systematic, practical advice, he stated that "Loyalty to the facts is worth more than a symmetrical architectonic or a hastily reached simplicity."[15] For Ross, the experience of moral conflicts is enough evidence of their reality, and they should not be dismissed out of convenience.

[14]Ross, W.D. *The Right and the Good* Oxford: Clarendon Press (1946) pp.16-42
[15]Ross, W.D. *The Right and the Good* p.23

However it was not until the latter half of the twentieth century that the full force of this style of criticism from moral conflict to moral theory was reaffirmed and engaged with. Bernard Williams broached a similar argument over the course of several articles and chapters in his books, beginning with his highly influential piece 'Ethical Consistency'.[16] Here, Williams suggests that the specific character of our experience of moral conflict in some situations indicates that there can be genuine moral dilemmas which do not admit of solution. He draws an analogy between conflicts of inconsistent moral obligations and conflicts of inconsistent desires. Whereas conflicts between inconsistent beliefs can be satisfactorily resolved by appealing to that belief which is most likely to be true, conflicts between desires and obligations cannot be resolved without leaving what Williams refers to as a 'remainder'. Choosing to satisfy one desire or obligation rather than the other will not leave an agent entirely satisfied. Rather, they are left with a sense of something similar to regret or remorse for that which they neglected; ideally they would have wished to have satisfied both desires, or acted on both moral 'oughts', if circumstances had allowed it. This remains the case even if an agent is convinced that they chose the morally best action which they could have taken. The regret is not a consequence of their wish that they had acted differently; they might be adamant that given the same situation they would act in an identical manner, and it would be experienced to a greater or lesser degree whatever the agent had done. In his words, "These states of mind do not depend, it seems to me, on whether I am convinced that in the choice I made I acted for the best; I can be convinced of this, yet have these regrets, ineffectual or possibly effective, for what I did not do."[17] They are the result of having failed to satisfy one of the moral oughts in conflict, which maintains its moral force even after one has decided that it represents a less pressing claim than the ought which one did satisfy.

Ross does make a similar claim in his earlier work, suggesting that the moral force of *prima facie* duties remained even after they did not constitute an actual duty in a particular situation. As he suggests, "When we think ourselves justified in breaking, and indeed morally obliged to break, a promise in order to relieve someone's distress, we do not for a moment cease to recognise a *prima facie* duty to keep our promise, and this leads us to feel, not indeed shame or repentance, but certainly compunction, for behaving as we do; we recognise, further, that is our duty to make up somehow to the promise for the breaking of the promise."[18]

[16] Williams, B. 'Ethical Consistency' in his *Problems of the Self: Philosophical Papers 1956-1972* London: Cambridge University Press (1973)
[17] Ibid, p.122
[18] Ross, W.D. *The Right and the Good,* p.28

Nonetheless, he does not place the same emphasis on this feeling of regret as Williams, who takes it to be highly significant in verifying the significance of moral conflict. It is this focus on our experience of moral conflict and regret which I will concentrate on as a criticism of monist moral theory.

Williams further maintains that experiencing a type of regret after having acted in situations of moral conflict is in fact appropriate and, at the least, not irrational, even from the perspective of an ideal moral agent. Certainly, he argues, those who experience such regret are no less morally admirable for it. In fact he holds that "The notion of an admirable moral agent cannot be all that remote from that of a decent human being, and decent human beings are disposed in some situations of conflict to have the sort of reactions I am talking about."[19] He considers the suggestion that these regrets are a consequence of 'natural' as opposed to moral motivations, and thus imply nothing about the structure of moral conflict. For instance, one might claim that the personal distress Agamemnon felt at sacrificing his daughter to save his army was of course entirely natural, but that it has no bearing on whether the situation qualifies as one of moral conflict or not. However, he rejects this notion, arguing that such distress cannot be so clearly distinguished from moral concern. The source of our aversion to such acts as allowing the death of one's child may be natural in origin, but it is also intimately tied up with our conception of them as wrong in some manner.

In the aforementioned article, Williams tentatively suggests that the vindication of moral conflicts as genuine which he offers poses a challenge to moral realism, as opposed to value monist moral theory. As a consequence, much of the literature on moral conflict has concentrated on countering or supporting this claim.[20] However, in a later piece he uses the experience of moral conflict and the remainder that it leaves as a specific argument against utilitarianism, which applies generally against all value monist theories. In *Utilitarianism: For and Against* he cites some moral dilemmas to help articulate it, including the now famous 'Jim and the Indians' case.[21] Here, he asks us to imagine that a man named Jim, whilst travelling through a South American country, encounters a group of twenty native peoples who are being restrained by a ruthless army commander and his men. The army commander is about to kill these innocents as part of a terror policy to suppress popular protest against the

[19] Williams, B. 'Ethical Consistency' P.123
[20] For instance, see Foot, P. *Moral Dilemmas and Other Topics in Moral Theory*, Oxford: Clarendon Press (2002)
[21] Smart, J.C.C, Williams, B. *Utilitarianism: For and Against* Cambridge: Cambridge University Press (1973) pp.98-99

government. However, he offers Jim the opportunity to save nineteen of the natives if he will concede to kill a single one himself. If Jim refuses this offer then the commander will certainly kill all twenty of them, who are all pleading with Jim to accept it.

Williams holds that in this particular case it would be morally preferable for Jim to go ahead and kill one of the natives, as the utilitarian would presumably advise him to. Nonetheless, his point is that it is not immediately *obvious* that this is the case, and that to take the decision lightly on the basis that killing one rather than letting twenty die would clearly minimise the loss of utility seems inappropriate. To do so would involve neglecting the distinct moral consideration of personal integrity, which one sacrifices when one directly causes a morally wrong action such as murder rather than merely allowing it to happen. He rejects the notion that this amounts to nothing more than self-indulgent moral squeamishness on similar grounds upon which he criticised the distinction between natural and moral aversion in his earlier paper; such so called squeamishness is at least partly derived from the sense of it being a wrong action. Thus, Williams concludes, utilitarianism fails to account for all morally relevant considerations. It implies that it would be inappropriate and irrational for Jim to hesitate and feel bad about killing the one to save the many, which is inconsistent with our explicit moral phenomenology.

Although Williams did not specifically state as much, one can deduce that this argument equally poses a challenge for Kantian deontology. The Kantian solution to the dilemma would be for Jim to refuse to murder the individual in order to save the entire group; the categorical imperative rules out such killing regardless of the severity of the consequences. Nonetheless, even if one disagreed with Williams and insisted that this struck them as the morally preferable course of action, it would again seem inappropriate for the individual to experience no sense of something like remorse or regret for having failed to save the lives of nineteen innocents. Just as utilitarianism fails to do justice to the morally significant value which people invest in their own personal integrity, Kantianism fails to properly account for the moral importance that people place on preventing death and suffering. In sum, as a consequence of their value monism, both theories cannot accommodate the notion of moral remainders and their associated negative emotions in any way other than attributing them to error on the part of the agent.

3. Moral Wrongness and Tragic Remorse

The issue of moral remainders and their significance for moral theory has inspired much discussion in the extensive literature devoted to moral conflict and moral dilemmas since the 1970s. For instance, Ruth Barcan Marcus claims that some situations of moral conflict not only inspire a feeling of remorse or regret for rational reasons, as Williams contends.[22] Rather, the fact that this remorse is regarded as justified indicates that such situations actually force wrongdoing on the part of the agent; there is nothing which the individual could do that would be morally right. She maintains that this is not a consequence of inconsistency of ethical theory; on her interpretation, principles need only be mutually obeyable in a theoretically possible world to count as consistent, and thus the occurrence of dilemmas under our non-ideal conditions does not indicate ethical inconsistency. For Marcus, the fact that moral conflict occurs in the real world, and thus leads to situations whereby whatever we do involves acting wrongly, does not imply theoretical inconsistency. All that we can do to avoid wrongdoing is attempt to structure our lives and the world in which we live so as to minimise the situations in which we may encounter moral conflict.[23]

Phillipa Foot, meanwhile, criticises this suggestion that moral conflicts necessarily involve wrongdoing on the part of those who act in them. She concedes that situations involving moral conflict exist and that resolving them might appropriately inspire a negative emotion in an agent, but denies that this entails that they have done wrong. She points out that sometimes one can feel 'creditable regret' without it implying wrongdoing, invoking the example of distributing the possessions of a dead relative. We may indeed feel regret for giving away objects that we feel connects us to someone that we loved, and it would be appropriate to do so. Nonetheless, this does not imply that such actions are wrong; in fact most would deem it morally admirable and not in the least disrespectful to the dead, especially if one gave them to those in need.[24] Earl Conee makes a similar point in suggesting that although one may experience subjectively justified guilt in acting in a situation of moral

[22] Marcus, R.B. 'Moral Dilemmas and Consistency' in *The Journal of Philosophy*, Vol. 77, No. 3. (1980), pp. 121-136.

[23] This proposal has itself being criticised by Conee, on the grounds that attempting to shy away from certain forms of moral dilemmas might actually be morally worse than accepting the moral wrongness that they supposedly necessitate. Conee, E. 'Against Moral Dilemmas', Philosophical Review 91 (1982) pp.87-97

[24] Foot, P. *Moral Dilemmas and Other Topics in Moral Theory,* Oxford: Clarendon Press (2002)

conflict, and that this guilt is appropriate, it does not imply that objectively justified regret is appropriate and thus that they have committed wrongdoing.[25]

I will not argue whether or not action in situations of moral conflict necessarily involves wrongdoing per se; taking such a position goes beyond the psychologically descriptive ambitions of this chapter. Nonetheless, I do contend that the negative emotion associated with moral remainders does not necessarily indicate that an agent must commit an action which they themselves take to be wrong in situations of moral conflict. In some situations, an agent will feel an emotion similar to regret or remorse and judge it appropriate that they experience such an emotion despite being willing to admit that they were not actually guilty of any wrongdoing. For instance, to take a case very similar to one which Williams' offers in his discussion of moral luck[26], suppose that an individual was driving down a quiet road very carefully, well within a speed limit and in a well maintained car when a child suddenly rushed across the road in pursuit of a stray football. The driver immediately brakes, but it is impossible to prevent their vehicle hitting and fatally injuring the young child. In this situation, the driver would presumably experience something akin to intense guilt and regret for having killed the child. Nonetheless, despite being causally responsible for their death, it would be an extremely harsh moralist who would accuse the driver of 'wrongdoing' in this particular case. Yet whilst most observers might attempt to comfort and reassure them that they have done nothing wrong, we would still judge it appropriate for them to feel bad about their role in the incident, and not merely the consequences that followed from them. If they did not, observers would typically regard them as morally insensitive; most take it to be incumbent on them to experience negative affect with regards to their own involvement in the bad state of affairs, rather than merely the bad state of affairs alone.

What this example shows is that, as a matter of descriptive psychology at least, perceived wrongdoing is not a necessary condition for experiencing some emotion similar to remorse and regret which is regarded as appropriate by both the subject of experience and third party observers. Something like these negative emotions can arise purely from being causally connected to a regrettable state of affairs; they needn't require that the individual conceive themselves as personally blameworthy. However, it is still open for some to object that remorse and regret proper are necessarily associated with overall wrongdoing in our standard understanding of the concept. As a matter of definition, in order for us to experience remorse,

[25] Conee, E. 'Against Moral Dilemmas', pp.87-97
[26] Williams, B., *Moral Luck*. Cambridge: Cambridge University Press, (1981) pp.27-28

we must on some level believe that we have done wrong. In that case, what the driver experiences does not count as remorse. This does seem to be a plausible suggestion, and in that case one can claim that what individuals who act in instances of moral conflict experience, whilst a negative emotion, does not count as true remorse (or regret) either.

More needs to be said about the distinct nature of the negative emotion which we experience when we act in cases of moral conflict. For it does seem right that we can only properly say that someone experiences guilt or regret concerning their actions when they wish that they had done otherwise. Yet people can still feel bad about their actions in situations of moral conflict whilst remaining confident that they would not change their decision if presented with an identical conflict in the future. Williams himself has pointed out that the standard terminology used to describe various negative emotions tied to our sense of responsibility for wrongness is lacking, and doesn't adequately demarcate the full range of different feelings that we are prone to experience. He raises the example of *regret*, which he deems too broad a term to refer to the particular feeling that we are typically disposed to experience when we are non-intentionally yet causally linked to a moral wrong. For we can intelligibly feel some form of regret about incidents that we have no causal relation to whatsoever; I can deeply 'regret' the Chernobyl disaster and all the death and destruction this incident caused without feeling complicit in any way. Yet this sort of regret is distinct from the feeling we would associate with the example above, where even though the driver acted blamelessly, they were nonetheless causally involved with the regrettable consequences. Williams names this distinct moral emotion 'Agent-regret', which is both elicited under different conditions and phenomenologically distinct from the type of regret that we feel concerning states of affairs unrelated to our own actions.[27]

This method of introducing new concepts for specific negative moral emotions associated with our responsibility for wrongness has been utilised by Stephen De Wijze, who coins the term 'tragic-remorse'.[28] Tragic-remorse is an emotion which typically accompanies acts which involve violating a moral principle in the pursuit of the overall right action. De Wijze holds that neither the concept of remorse nor agent regret adequately captures the standard phenomenology of acting in such a manner. Agent regret and remorse involve such feelings as a personal link with a violation of a moral principle or duty, guilt, shame and the necessity of making amends, which are all shared with the concept of tragic-remorse. Yet the former

[27] Williams, B., *Moral Luck* pp.27-28

[28] De Wijze, S., 'Tragic-Remorse – The Anguish of Dirty Hands' *Ethical Theory and Moral Practice* (2004) pp.453–471

emotions also involve the feeling of wishing we had done otherwise, having failed from a moral point of view and the need to reform our character to avoid making such mistakes in the future. These features are not characteristic of tragic-remorse, and although it involves feelings of anguish, shame and guilt, also includes a feeling of pride or relief at having done the overall morally right thing for the right reasons and a commitment to continue to act as such in the future. De Wijze suggests that tragic-remorse is most strongly related to 'dirty hands' situations. These are a subcategory of moral conflicts which come about when we are forced to render ourselves complicit in the evil projects of others, typically associated with political acts considered necessary which nonetheless violate moral principles.[29] Nonetheless, it can be generalised as capturing the emotion which we typically experience in acting in any situation of moral conflict; there is nothing in De Wijze's account to suggest tragic-remorse is exclusively a reaction to dirty hands situations.

I maintain that the fact that people tend to experience tragic-remorse in situations of moral conflict, and that they regard it as an appropriate response, entails that descriptive value monism does not square with our typical moral phenomenology. Whilst an agent may act in a manner that they deem to be overall right in a moral conflict, their negative experience suggests that in acting all things considered rightly they have nevertheless neglected something which they take to be distinctly reason-giving. This does not imply that they regard themselves as committing wrongdoing in terms of an overall evaluation of an act; even in particularly difficult cases, one can feasibly regard either option to be morally permissible. However, an action can involve regrettable neglect of a value taken to have independent normative force without the act being taken to be wrong in itself. Rather than involving perceived overall wrongness per se, the action instead involves *moral loss*; the sense that one has sacrificed one value in realising another, and that what is lost in this sacrifice is not entirely redeemed by what is gained. This represents an underlying explanation as to why some situations prompt us to experience tragic remorse.

4. Coin Flipping and Implications for Value Monism

I will next discuss the difference in attitude that we have between what counts as acceptable decision making procedures in different types of moral conflict. This provides further

[29] What precisely distinguishes dirty hands situations from standard moral conflicts is contentious, with different theorists offering varying interpretations. I will not enter into a discussion on this tangential issue.

evidence that people experience distinct moral loss as opposed to the feeling that they have acted wrongly when acting in situations of moral conflict.

In 'The Diversity of Moral Dilemma', Peter Railton argues that the appropriate decision making procedure in situations of apparent moral dilemma differs depending upon what type of values are in conflict.[30] Some apparent dilemmas involve conflict between different instantiations of the same value. For instance, we may find ourselves in a position whereby two complete strangers whose lives we know nothing about are both in mortal danger, and we can only possibly save one. In the same vein, we may be forced into a position where we have no choice but to break one of two promises of equal stringency which we have made to two equally good friends. In these kinds of situation, Railton suggests we would be justified in adopting a random decision making procedure such as flipping a coin to decide what to do; there is simply no non-arbitrary means of deciding which life we should save, or which promise we should break. Moreover, although there is left what one might deem a moral remainder after we have acted in these cases, our sense of tragic-remorse is likely to be lessened by the thought that there were no grounds for our deciding to do otherwise.

However most moral conflicts do not fit this symmetrical pattern. Rather, they involve a conflict between seemingly distinct moral considerations. For instance, the Jim and the Indians case necessitates a choice between minimising death and preserving one's personal integrity. As Railton argues, it seems inappropriate to utilise a random decision making procedure in these types of dilemma. Whether we choose to prioritise the lives of the natives or our own personal integrity is not experienced as an arbitrary choice, even if it may be one where moral loss is accompanied by either choice. In choosing our actions, we are expressing a commitment to a position on which consideration is of greater ethical significance in the context of the conflict. To flip a coin to make this decision for us would be to avoid making such a commitment, and effectively try to pass off our responsibility for our actions by putting it down to random chance. A moral remainder, and the associated experience of tragic-remorse is inevitable when making a decision in these asymmetrical conflicts. Nonetheless, despite this, it would seem like absolving oneself of a choice which is of genuine import if we were to put it down to chance. And even when we are satisfied that we have attended to the most important moral consideration in our actions, this does not make up for the distinct consideration which we necessarily neglect.

[30] Railton, P. 'The Diversity of Moral Dilemma' in Mason, H.E (ed.) *Moral Dilemmas and Moral Theory,* Oxford: Oxford University Press (1996) pp.140-166

I hold that this account accurately reflects how most individuals implicitly conceive of appropriate decision making in situations of moral conflict. When what is at stake cannot be easily compared, we judge it important to make a decision between which consideration should be prioritised. This sense of certain moral considerations as dissimilar and yet requiring determinate, non-arbitrary choices between them, along with the tragic-remorse associated with moral remainders, indicates that we identify and are morally motivated by a range of distinct values as opposed to a single consideration which all others are ultimately reducible to.

Thus, from our explicit moral phenomenology as encountered when we have to deal with moral conflict, we can infer what underlying structure is shaping how humans experience value. We may not be consciously aware of this underlying level of moral experience; indeed, if we were, then it would not be necessary to make such inferences about it from our more basic reactions to moral quandaries. Nonetheless, the implication of moral remainders and our intuitive sense of how to appropriately react to moral conflict is that we tend to regard a multiplicity of distinct moral values as reason-giving, rather than a single one. This is, I maintain, the most significant inference concerning our underlying moral thought that we can make from our shared experiences associated with moral conflict.

Of course this does not entirely undermine monist theories of value. It is still open for the utilitarian or the Kantian to claim that individuals' multifaceted moral motivations are an inaccurate reflection of ethical reality, regardless of whether they are generally regarded as appropriate or not. People may simply be mistaken to judge it morally appropriate to experience tragic-remorse when acting in situations of apparent moral dilemma.

To elaborate; a utilitarian or Kantian might be willing to accept that agents might experience tragic-remorse after having acted in a situation of apparent moral conflict. They might even offer reasons why it might be appropriate for an imperfect moral agent to experience such a negative emotion. For instance, the utilitarian moral theorist R.M Hare suggests that, given our cognitive limitations, it is preferable for individuals to develop sentimental attachments to 'moral' considerations which only indirectly contribute to utility, and thus experience negative emotions when they are sacrificed in the name of utility per se. He thus holds that "Our common intuitions are sound ones just because they yield acceptable precepts in common cases. For this reason, it is highly desirable that we should all have these intuitions and that our consciences should give us a bad time when we go against them."[31] Nonetheless,

[31] Hare, R.M. 'Moral Conflicts', *The Tanner Lectures on Human Value* (1978) p.186

this is merely justifying tragic-remorse out of the practical consideration that humans are flawed beings. If an idealised agent with perfect instrumental rationality, a faultless ability to calculate predicted consequences and the right moral motivation was to emerge, such sentimental attachments to secondary considerations would be redundant; it is necessarily a consequence of any monist theory that tragic-remorse is ultimately irrational and inappropriate for such an ideal moral agent. This is because, as long as one has acted so as to best realise the single moral value conceptualised by these theories, there is absolutely no reason to feel morally bad for what one has done.

Yet even if such an agent were to exist, I contend that most individuals would regard such an agent as inhuman, and lack something important in our common humanity, to the extent that they did not experience tragic-remorse. Our intuitions regarding the appropriateness of tragic remorse do not, I suggest, stem from the implicit understanding that such emotions ultimately serve to better enable us to realise a single moral value. Rather, they reflect the more simple truth that individuals feel that it is important to recognise multiple, distinct considerations as reason-giving, and that when one of these considerations are put to one side, it is to be mourned. This makes no sense from a value monist standpoint; a consideration is only reason-giving insofar as it realises the single value, which any given option can do to a greater or lesser extent. A value monist could intelligibly feel bad that the single value they posit couldn't be realised to a greater extent, given the situation that they are presented with. But if one picks the option which does so to the greatest extent possible in the situation, there are no grounds for feeling bad concerning the choice that one made. One has acted in the straightforwardly morally best way that one possibly could, given the context, and there is no distinct moral loss involved. As Michael Stocker puts it, the value monist must claim that "there is no ground for rational conflict because the better option lacks nothing that would be made good by the lesser."[32]

Although it is open for the value monist to deny the relevance of our intuitive responses regarding moral conflict, or, like Hare, to invoke an ad-hoc explanation to accommodate them, if we are looking for a psychological descriptive account of value more consistent with our typical moral phenomenology then we should turn to an alternative. The opposing theory to value monism is value pluralism, and the most influential form of it was advocated by the

[32] Stocker, M. *Plural and Conflicting Values*, Oxford: Clarendon University Press (1990) p.272

political philosopher Isaiah Berlin.[33] Although, like value monism, it was initially conceived of as a normative/metaethical thesis, I will be explicating it in order to analyse it in terms of its adequacy as a *descriptive* account of value.

5. Berlin's Value Pluralism

Berlin never attempted to explicitly formulate his understanding of value pluralism, yet one can infer his position concerning the nature of value from his various writings on the history of philosophy. To articulate the position, Berlin focused on bringing to light the underlying role that it plays in the works of particular historical thinkers such as Machiavelli, Vico, Herder and Montesquieu.[34] In his exegesis of these philosophers, he inferred that their conception of value differed significantly from that of the standard monistic enlightenment view. Whilst the latter is characterised by a belief that all human ends are in principle mutually consistent and attainable, since they are ultimately reducible to a single value, Berlin highlights the former's commitment to an understanding of goods as plural, incommensurable and often incompatible. For example, on Berlin's interpretation, Machiavelli held that Christian values such as humility and charity were incompatible with the pagan virtues necessary to maintain a strong and prosperous state,[35] whilst Herder argued that each culture possesses its own unique set of values which cannot all be attained in a single form of life.[36]

Value pluralists such as Berlin suggest that there is a range of distinct sources of moral goodness, each worthy of pursuit in its own right. Though there is a limit to the extent of these values, few pluralists attempt to provide an exhaustive list. As Berlin himself states, "The number of human values, of values which I can pursue while maintaining my human semblance, my human character, is finite – let us say 74, or perhaps 122, or 26, but finite, whatever it may be."[37] Nonetheless, the main significance is not in the details of the particular values which constitute the range, but in the theory of plurality itself. For claiming that there may be more than one ultimate human end admits the possibility that our pursuit of one

[33] Ross' ethical theory was also pluralistic, and came before Berlin's. Nonetheless, it was not as radical as Berlin's exposition of the idea in that it denied the necessity of moral loss in situations of conflict, and did not have the same intellectual impact.

[34] For a collection of Berlin's relevant work on these thinkers and more, see Berlin, I. *Against the Current: Essays in the History of Ideas,* London: Hogarth Press (1979)

[35] Berlin, I. 'The Originality of Machiavelli', in his *The Proper Study of Mankind: An Anthology of Essays*, London: Pimlico Press (1998) pp.269-325

[36] Berlin, I. 'Herder and the Enlightenment', in Ibid, pp.359-435

[37] Berlin, I. 'My Intellectual Path' in *The Power of Ideas,* Princeton: Princeton University Press (2000) p.12

might conflict with that of another, implying that our political policies and individual actions cannot realise all human goods simultaneously. In his words, "To admit that the fulfilment of some of our ideals may in principle make the fulfilment of others impossible is to say that the notion of total human fulfilment is a formal contradiction, a metaphysical chimera"[38]

This notion of value plurality and conflict, although not the central theme of the paper, is also clearly expressed in Berlin's highly influential 'Two Concepts of Liberty'[39]. Here, Berlin suggests that two distinct values come under the common name 'liberty'. He explains that whilst 'negative' liberty is realised when we are unconstrained by external, manmade barriers to action, 'positive' liberty involves self-mastery over one's own internal constraints. Although related, these two forms of liberty are not interchangeable and cannot be mutually maximised within a single society; the full realisation of positive liberty necessarily involves a restriction of negative liberty, whilst maximising negative liberty unavoidably neglects positive liberty. This is because it is sometimes necessary to externally constrain individuals in order to help them overcome internal restrictions; to use the phrase from Rousseau, to 'force them to be free'.[40] For instance, we might forcibly restrict a drug addict from ingesting their chosen substance, thus impairing their negative liberty, on the grounds that helping them overcome their addiction will improve their positive liberty to live a life in line with their second order desires. Although Berlin finds reason to place a general preference on negative over positive liberty, he admits that both are genuine goods in themselves and that the sacrifice of one in pursuit of the other inevitably involves distinct moral loss.

Although Berlin concentrated on the political implications of this thesis, he also suggests that such conflict between incompatible values is a common feature of our ethical experience, and applies on an individual as well as a social level. Just as social values such as positive and negative liberty, equality and security regularly come into conflict, as do personal virtues such as honesty, compassion and loyalty. Thus, his account of value pluralism readily explains the sense of there being a moral remainder as a consequence of our actions in situations of moral conflict, especially when the considerations at stake are asymmetrical. When we act in cases of moral conflict where two distinct values are at stake, we neglect goods which cannot be made up for by the preservation of those which we attend to. In Berlin's terminology, in such situations we make a 'tragic choice'; our act is regrettable and inspires negative emotion as it necessarily involves moral loss. Indeed, he predicts such tragic choices and their associated

[38] Berlin, I. *Liberty*, Oxford: Oxford University Press (2002) p.213
[39] Ibid, pp.166-217
[40]

remainders to be a defining feature of our everyday life and not merely restricted to hard cases of apparent moral dilemma, given the multiplicity and inherently incompatible nature of values.

Value pluralism is thus more in line with the phenomenology of moral conflict than monistic accounts of value. The notion that humans treat a range of distinct considerations which cannot be reduced to a single concern as reason-giving accounts for the experience of moral conflict without resorting to attributing it to irrationality. Given that we tend to regard our feeling of tragic-remorse in response to moral remainders as entirely appropriate and justified, this indicates that value pluralism accords with our intuitive understanding of moral conflict. However, this is not to claim that some form of value pluralism represents the ethical theory which better reflects an objective moral reality in some deeper sense. As previously mentioned, it is still open for value monists to deny that coherence with our moral phenomenology represents the most salient standard by which to judge an ethical theory.

Berlin himself proposed value pluralism as an ontological, metaethical thesis; he claims that "the multiple values are objective, part of the essence of humanity rather than arbitrary creations of men's subjective fancies."[41] Nonetheless, one does not need to subscribe to this conception of distinct values as existing in an objective sense in order to subscribe to the view that it is descriptively true that humans universally tend to recognise a wider range of ends to be morally salient than those posited by monist theories of value. Those who attribute a high degree of salience to our ethical experience in informing moral theory might consider this to indicate that value pluralism better represents the objective moral order. Again, I need not take a strong position on this issue for my purposes. I merely aim to justify the descriptive understanding of what is going on when we encounter moral disagreement that value pluralism implies when I later go on to discuss the issue, on the basis that it provides the most accurate account of our typical moral phenomenology in situations of conflict, for better or worse.

At this point, it might be useful to re-emphasise the various ways in which value pluralism and monism could apply and say more about how they relate, in order to clarify the nature of my argument. Firstly, value pluralism or monism could hold on a purely descriptive level. According to descriptive value monism, humans ultimately treat a single value as legitimately

[41] Berlin, I. 'My Intellectual Path', p.12

reason-giving.[42] All of our apparently distinct moral concerns can be reduced to care for this one value, even if we are not always conscious of this link. An example of a descriptive value monist position is psychological hedonism, most famously espoused by the utilitarian philosopher Jeremy Bentham.[43] Psychological hedonism holds that humans ultimately strive for pleasure and the avoidance of pain on an underlying level; all motivation is guided by this concern, even apparently altruistic actions reflect an implicit understanding that such behaviour will maximise our pleasure. Descriptive value pluralism, in contrast, holds that those ends which we treat as reason-giving are irreducibly distinct. For instance we might treat both loyalty to one's family and impartial justice as reason-giving, and these considerations are not reducible to a third factor such as their contribution to our happiness.

However, these positions are distinct from normative value pluralism and monism. The theories of utilitarianism and Kantianism which I have articulated are monist in this sense. They do not necessarily deny that individuals, as a matter of psychological fact, experience multiple distinct ends as reason-giving. Rather, they hold that we should consider only a single value as legitimately reason-giving, regardless of whether this is a natural tendency of human psychology or not. Normative value pluralism suggests the contrary; in order to act morally, we need to consider a range of values as worthy of consideration and act in accordance with them all, insofar as we are able to do so.

Now, there is logical space for any combination of these two forms of value monism/pluralism; accepting descriptive value pluralism does not commit one to normative value pluralism, nor does advocating descriptive value monism force one to subscribe to normative value monism. As I have just explained, a utilitarian or Kantian can consistently combine descriptive value pluralism with normative value monism. Hypothetically, one could also endorse a position which combined descriptive value monism with normative value pluralism. For example, a theory might hold that humans take only their own pleasure as reason-giving but that in fact they should strive to pursue a multitude of other moral goods, or satisfy a range of distinct principles, other than maximising ones pleasure. Although both the sorts of positions could fall foul of the 'ought implies can' principle, depending on the

[42] I use the term 'reason-giving' throughout this chapter as it relates to those motivations which we endorse on a second order level. Thus, it rules out motivations such as a drug addict's striving for heroin, or the kind of self-interested desires which we do not approve of and attempt to overcome when they conflict with considerations which we recognise as having moral worth. The descriptive value monist need not deny that we are motivated by many distinct ends; they need only maintain that we only perceive one of these ends as ultimately morally worthwhile, and thus legitimately reason-giving.

[43] See Bentham, J. *Introduction to the Principles of Morals and Legislation* Oxford: Clarendon Press (1907)

level of psychological flexibility towards what we take as reason-giving that we conjecture, they cannot be precluded purely on the basis that psychological facts should dictate our prescriptive standards. Such preclusion would run up against Hume's law – the famous dictate against deriving an 'ought' from an 'is' without any independent reason for doing so.[44]

Having said that, there is also a sense in which the matter of whether either descriptive monism or pluralism holds true does have a bearing upon the status of normative monism or pluralism. For most ethical theorists implicitly assume that there is at least some indirect relation between our descriptive moral thought and moral reality. Ethical philosophy does not generally attempt to merely codify common sense morality; there is usually a degree to which it attempts to refine and correct what we believe to be the right grounds for action and feeling. Nonetheless, it would be an odd ethical theory which held that our pre-theoretical understanding of what is of moral worth and therefore legitimately reason-giving is completely misguided. If our ethical intuitions are universally mistaken, it is not clear from where we are supposed to derive our understanding of moral truth, and most moral philosophers, explicitly or implicitly, take moral intuitions as an important source of evidence when evaluating ethical theories. In Ross' words, "the moral convictions of thoughtful and well-educated people are the data of ethics just as sense perceptions are the data of a natural science"[45] Moreover, as John Rawls frames it in his discussion of reflective equilibrium, any ethical theory can only be inconsistent with so much of our moral experience until it becomes implausible and in need of revision.[46]

Thus, whilst I intend only to argue for a form of descriptive value pluralism, I believe as an aside that this does in fact count against normative value monist theories, such as utilitarianism and Kantianism as they are traditionally interpreted. As I make clear, this does not constitute a knock down refutation of these theories, but it does go some way towards undermining them.

[44] See Hume, D. *A Treatise of Human Nature: A Critical Edition* David Fate Norton and Mary J. Norton (eds.), Oxford, Clarendon Press (2007) 3.1.1.27
[45] Ross, W.D. *The Right and the Good,* p.41
[46] Rawls, J. *A Theory of Justice: Revised Edition,* Oxford: Oxford University Press (1999) pp.18-19

6. The Problem of Value Incommensurability

There is a further aspect of Berlin's value pluralism which seems to be less consistent with our moral phenomenology in situations of value conflict; that of the supposed *incommensurability* of values. Value incommensurability is a difficult concept to make sense of and much work has gone into articulating it.[47] Here I will attempt to offer a brief account of it in order to explain why it does not cohere with our experience of deliberation in situations of moral conflict.

The claim that values are not only plural and conflicting but incommensurable represents the most radical and contentious aspect of Berlin's value pluralism. Value incommensurability entails that there is no means of rationally comparing different values and that they cannot be weighed against each other by means of some common currency. As he suggests, "The world that we encounter in ordinary experience is one in which we are faced with choices between *ends equally ultimate*, and *claims equally absolute*, the realisation of some of which must inevitably involve the sacrifice of others."[48] (Emphasis mine) To illustrate, value incommensurability rules out claims such as the following:

Value *x* is worth more than/less than/equally as much as value *y*.

Such comparisons make no sense to the value pluralist who subscribes to an understanding of values as strongly incommensurable. For claiming that one value is greater, lesser or even equal to another implicitly involves appealing to a further, higher order value to ascribe each its relative worth. This fails to appreciate the nature of values as distinct entities, which are not reducible to a common measure of evaluation.

On the strongest understanding of this idea, one cannot claim that a particular choice to ascribe greater relative worth to values such as compassion, honesty or personal integrity is better grounded than any other. Without any ultimate principle to guide us, we cannot make any better or worse choice, all things considered, when incommensurable values clash. This would serve to undermine any hope of making clear and unqualified normative judgement that we regard as right in such instances. For there would be no preferable way of resolving the matter – a particular choice could only be said to be better than others relative to a particular value. For instance, in choosing whether to lie to protect the feelings of another or

[47] See Chang, R. (ed.) *Incommensurability, Incomparability, and Practical Reason*, Cambridge, MA: Harvard University Press (1997) for a collection of articles on incommensurability
[48] Berlin, I. 'Two Concepts of Liberty', p.213

not, we might judge lying to be most compassionate course of action and telling the truth to be most honest, but there would be no way of saying which action is most 'right'. If this was true, there does not seem to be much productive that we can say when it comes to ethical prescription beyond highlighting the exact value trade-offs that we face in decision making. Whether we should proceed to make such trade-offs is a matter we could not tackle in a non-arbitrary manner.

Yet this strong account of value incommensurability is not reflected in how people tend to deliberate in situations of moral conflict. It is of course plausible that perceived values are distinct to such an extent that acting in such a way as to realise one does not make up for the loss of another. As argued earlier, the sense of moral remainder and tragic-remorse associated when attending to one of two or more conflicting considerations indicate that people do not easily weigh competing values by reference to a common factor. Nonetheless, people do not regard their judgements in most situations of moral conflict to be arbitrary. In fact many instances of moral conflict admit of a solution which seems obviously right to most people, despite the distinct nature of the incompatible considerations. For instance, few would hold that the choice between withholding and returning borrowed weapons to a friend who has subsequently gone mad in Plato's famous case is a morally arbitrary one. We might well admit that there is distinct moral loss involved in refusing to return the weapon to the crazed friend, insofar as the agent involved breaks a promise to return them. Nevertheless, it would strike most as ludicrous to suggest that the values involved are incommensurable to the extent that this option can be evaluated as no better nor worse than returning it; it seems entirely clear that withholding the weapon is morally preferable. As E.J. Lemmon puts it in discussing a similar case, "Someone who thinks that it would really be better to return the gun must either hold the importance of a man's giving his word to be fantastically high or else hold human life to be extremely cheap, and I regard both these attitudes as morally primitive."[49]

Moreover, as discussed earlier, even in harder cases where people are less certain of the best course of action, they are generally able to make a decision in a manner which they regard as non-arbitrary. Indeed, whilst after acting they may experience the tragic-remorse associated with moral remainders and regard this experience as appropriate, they do not cease to believe that they have acted in the best manner possible. In contrast, as Railton suggests, the situations in which the choice is regarded as entirely arbitrary are those in which the conflicting considerations at stake are identical in kind and of equal weight. These are the

[49] Lemmon, E.J. 'Moral Dilemmas' in *The Philosophical Review* 70 (1962) p.147

only moral conflicts in which it seems appropriate to adopt an arbitrary decision making process such as flipping a coin. Yet if values were truly treated as incommensurable then this should hold true for all moral conflicts. Clearly, people manage to implicitly commensurate values in some manner or other.

In fact, Berlin himself does not seem to adopt such a strong conception of value incommensurability in his understanding of value pluralism as that sketched above. An indication of this is his explicit support of liberalism and of the value of negative over positive liberty in particular, despite his admission that both represent distinct values of genuine worth. In fact, he seems to believe that value pluralism itself supports liberalism, suggesting that "if value pluralism is a valid view… then toleration and liberal consequences follow."[50] If values were genuinely incommensurable to the extent outlined above then this would be an incoherent viewpoint. For whatever potential advantages Berlin might cite in favour of negative liberty, a strong interpretation of incommensurability renders the prospect that one value or set of values might be decisively preferable over another inconceivable. Indeed, George Crowder, Gerald Gaus and John Gray have reconstructed multiple possible arguments which Berlin might be attempting to deploy in favouring liberal over non-liberal values, and unanimously conclude that if one takes value incommensurability seriously then they all must ultimately fail.[51] Far from lending support to non-interference, Crowder even suggests that pluralism might actually undermine the case for negative liberty. For as he claims, if we accept a strong account of the incommensurability of values, "it is always open to the pluralist to ask, why not the illiberal option?"[52]

7. Commensurating the Incommensurable

On the basis of philosophical charity, then, we shouldn't attribute radical incommensurability to Berlin's value pluralism – to do so would be to commit him to grave inconsistency. In fact, in response to Crowder's claim that "choices among incommensurable values 'are "underdetermined by reason""[53], a short article co-authored by Berlin and Williams explains that they do in fact believe there to be a means by which seemingly incommensurable values

[50] Berlin, I. 'My Intellectual Path' p.13
[51] Crowder, G. 'Pluralism and Liberalism', *Political Studies*, Vol. 42, No. 2, (1994) pp.293-305, Gaus, G. F. *Contemporary Theories of Liberalism*, London: SAGE (2003) pp.42-50, Gray, J. *Isaiah Berlin*, Princeton: Princeton University Press (1997) pp.150-156
[52] Crowder, G. 'Pluralism and Liberalism', p.304
[53] Berlin, I., Williams, B. 'Pluralism and Liberalism – A reply.' *Political Studies* (1994) p.306

can be weighed against each other in a reasoned manner. They hold that Crowder has confused the notion that no single value, such as justice, has rational priority over all others in every possible instance, with the very different claim that when two values clash in a particular instance that reason has nothing to say about which takes priority. They maintain that whilst the first claim is of course a consequence of value pluralism's ruling out of a 'priority rule', the second is not implied by their doctrine. Thus, although it remains the case that "practical decision could not in principle be made completely algorithmic"[54], in contrast with monistic ethical theories, pluralists are not condemned to the implausible position that any choice between incommensurable values is entirely arbitrary.

However Berlin and Williams do not offer any details as to how we can weigh values in such a reasoned manner. They clearly believe that it is important that we take the contextual details of each particular situation of conflict into account when coming to a decision about which value takes priority there, but beyond this they are vague. Crowder interprets their rebuttal of his complaint as an indication that they hold a view similar to that of Aristotle. That is, although the prospect of a systematic, principle based normative theory is impossible, given the plurality of salient values and situational factors at play, one can develop a kind of moral sensitivity and practical wisdom which enables one to make reasoned ethical choices in situations of conflict. The most that we can hope for in terms of ethical principles are only very roughly accurate generalisations, which are always liable to come unstuck and admit of exceptions given the variable context of conflicts. For the most part, we must rely on our own sense of judgement when we face situations when principles conflict and fail to offer determinate advice. Conflicts between values are thus to be resolved in a manner akin to that recommended by moral particularists; through judging each particular tricky case on its own merit rather than appealing to a universally applicable principle.[55]

This sort of account, though very vague, does seem to be how in fact people implicitly make moral judgements in situations of value conflict. As discussed, people do not regard their decisions as entirely arbitrary, and feel it would be inappropriate to simply flip a coin to decide when two distinct values are at stake, even in hard cases. However, neither do they decide what they judge to be the action which is all things considered morally preferable by prioritising a single value at the expense of others across all cases and determining how best to realise it. Rather, individuals typically take a middle course. They tend to take individual

[54] Berlin, I., Williams, B. 'Pluralism and Liberalism – A reply.' p.307

[55] For detailed discussion of Moral Particularism, see Little, M. (ed.), *Moral Particularism*, Oxford: Oxford University Press (2000)

situations and, consciously or not, weigh conflicting considerations up against each other in a context-sensitive manner. Their resolution may leave moral remainders, yet generally allow them to come to judgements which they feel to be all things considered for the best. Although some situations will involve particularly difficult trade-offs between values which people feel to be of almost equal importance in the context, they can usually manage to make judgements which minimise their sense of moral loss involved, if not eliminate it entirely. This implies that despite the fact that people do take many moral considerations to be distinctly reason-giving, they do in fact have some vague means of commensurating values; they implicitly adopt a model of moral deliberation which incorporates the distinctness of value pluralism with the determinacy of judgement that monism enables. It is simply that this form of monism that they tacitly adhere to is not as clear cut as popular monistic theories of values prescribe. Again, this is a descriptive rather than a prescriptive claim; I am not suggesting that this is an objectively preferable means by which people should morally deliberate, merely that this is the method which people tend to implicitly use.

This account of how people typically morally deliberate in situations of moral conflict sheds light upon the nature of moral disagreement. If people reach their moral judgements via a combination of developing a full understanding of the contextual details of a situation and then weighing the importance of the conflicting values in the context, then disagreement could arise at two different stages. Firstly, individuals may not share the same level of understanding of the descriptive details of a given situation, and thus not be aware of what value trade-offs are involved. For instance, to take Plato's case of returning weapons to a friend, a person who was unaware that the friend had gone mad would probably regard it as wrong to refuse to return the weapons. In their mind, there is only one salient value at stake here – that of repaying one's debts. They would thus morally disagree with the more fully informed individual, who would judge it wrong to return the weapons on the understanding that the friend had gone mad and might use the weapons to go on to commit senseless violence. However, this sort of disagreement does not necessarily reflect any deep, intractable difference in their ethical attitudes. It would most likely be a consequence of non-moral misunderstanding; if they shared the same knowledge concerning the contextual details of the situation, then they would probably come to the agreement that it would be wrong to return the weapons to their friend.

Yet the account of moral deliberation sketched also concedes the possibility of a more fundamental disagreement, which does not admit of such easy resolution. For it does not rule out the possibility that two individuals may share identical knowledge of a situation of value

conflict and be equally sensitive to each contextual detail involved and yet make opposing moral judgements. They may simply disagree about the relative importance of each value consideration in that particular situation. For instance, in the case where one has to choose between telling a lie and hurting someone's feelings, two individuals might possess precisely the same understanding of the context of the situation and both be acutely aware of the trade-off involved. Nonetheless, one may feel that in this instance the value of honesty trumps compassion and thus judge it right to tell the truth here, whilst the other might hold the opposing view. The former may simply invest more importance in honesty relative to compassion either *in general*, and thus hold that it trumps the value of compassion in this particular instance and in many other instances where it conflicts with other values, or *in this particular situation*, whilst the latter weights these values differently.

This proposed model of moral deliberation and disagreement may still seem vague, and further articulating the above example might help clarify it somewhat. Take two agents, James and Amanda, who face a moral conflict; each has to choose between lying to save their mutual friend Kevin's feelings, or to tell the truth and thereby cause him distress. Let us imagine that for the sake of this example, the relevant contextual details of the situation each faces are to be taken as completely identical, and that both have the exact level of non-moral knowledge of said details, although this would admittedly be impossible to say for a pair of any real world situations. Now, let us say that Amanda judges that it would be wrong to lie to Kevin and accordingly tells the truth, whilst James judges that it is morally preferable to lie and save Kevin's feelings, which he does. On my account, the best explanation for this discrepancy in their judgements is thus; James assigns a higher weight to the distinct value of compassion relative to honesty in this particular situation than Amanda does. This is not to say that James would always hold that compassion trumps honesty across multiple situations, and that Amanda would in contrast always feel that honesty trumps compassion. There may even be situations where Amanda would judge compassion as more important and act accordingly where James would disagree and prefer honesty: the relative importance of each value in each situation for each party might be highly dependent on particular contextual details, some of which will have greater salience for one party than the other.[56] Nonetheless, let us say that in this situation, Amanda assigns x units of moral weight to honesty, and y units to compassion, whilst James assigns x units to compassion and y to honesty, where $x > y$.

[56]This being said, my TEA model's account of moral disagreement articulated later suggests that individuals will generally place greater importance on the same values across multiple situations.

As previously explained, this model is part value pluralist, part value monist. It is pluralist to the extent that each individual would experience honesty and compassion as distinct reason-giving factors, and thus suffer a sense of distinct moral loss in whichever way they act. Yet it is monist to the extent to which these values can in fact be commensurated on an extremely vague and abstract level. This enables agents to make judgements about the relative moral importance of distinct values in particular situations and further judgements about what would overall be for the best, thus minimising the sense of moral loss which they experience when acting upon these judgements.

Of course, the example conjectured above is not necessarily possible on the account of moral deliberation I suggest; it is merely a theoretical possibility. It could still very well be the case that if everyone held the exact same non-moral knowledge of a situation, then moral disagreement would never occur. Individuals might implicitly weight each value exactly the same given an identical context. Nonetheless, if it were to be shown that some moral disagreement cannot be attributed to differences in non-moral understanding, then we could comfortably attribute it to differences in how individuals weight the importance of different values against each other. We would then require an explanation as to why some individuals were prone to make such varying value judgements despite sharing the same non-moral knowledge. In the chapters which follow, I intend to show that in fact some moral disagreement cannot be reduced to non-moral disagreement, and provide such an empirical explanation as to why people weigh the importance of values differently, before determining the normative consequences from the perspective of liberal political theory.

Conclusion

I have argued that monist theories of value such as utilitarianism and Kantianism cannot explain our experience of moral conflict without attributing it to implicit error regarding our own moral commitments, which runs against our intuitions concerning our appropriate reactions to it. We typically experience tragic-remorse when we act in particularly hard cases of moral conflict, even when we are sure that we have done the right thing, and regard the feeling of this emotion as appropriate, whereas value monists must judge this to be irrational and unjustified. Whilst this does not necessarily tell against normative value monism, it does suggest that value monism is false on a descriptive level.

Value pluralism, meanwhile, provides a descriptive account of the underlying structure of our moral thought which better accommodates the explicit phenomenology of moral conflict. This account suggests that humans are prone to experience tragic remorse when choosing in situations of moral conflict because it invokes the sense of distinct moral loss. Thus, individuals are not somehow misguided to experience such an emotion. It does not stem from a confusion, which they wouldn't experience if they were ideally situated so as to understand that they in fact were committed to a single moral value, rendering such an emotion groundless. Rather, tragic remorse is a natural and appropriate response given the multiplicity of ends which people implicitly take to be reason-giving. However, people do manage to come to judgements about what is right and wrong in situations of moral conflict and do not think that any decision made would be a purely arbitrary choice. Thus it is more plausible to suggest that values, although distinct, might not be experienced as strongly incommensurable as is suggested in the most radical versions of the theory.

Descriptive value pluralism therefore offers a good basis upon which to conceptualise what could potentially be going on when we encounter moral disagreement. When people disagree over moral judgements they are not merely disagreeing about how best to realise a single moral value. Rather, they are implicitly treating the situation as one of value conflict, and could either be disagreeing about non-moral factors concerning the context of the conflict, *or* it could be a consequence of the individuals involved differing over the relative importance of distinct values. I will use this understanding of moral disagreement, derived from my exploration of the phenomenology of internal moral conflict, to inform my discussion of such disagreement between individuals in later chapters.

Chapter 2 – Intercultural Fundamental Moral Disagreement

In the last chapter I argued that agents tend to treat multiple, distinct moral considerations as reason-giving, on the basis that individuals generally regard the experience of tragic remorse as an appropriate reaction after acting in situations of moral conflict. However, individuals manage to resolve the conflicts which often arise as a consequence of this underlying phenomenology of value by implicitly weighting each consideration differently. They then judge acting in accordance with the weightiest moral concern as all things considered right even if this leaves them tinged with the sense of distinct moral loss. Agents thus adopt a form of moral deliberation which is value pluralistic without strong incommensurability.

I further suggested that this hints at a potential explanation of the basis of moral disagreement; when two individuals make diverging moral judgements, it could be the case that each assigns considerations different weights. For instance one agent might place more emphasis on the importance of compassion whilst another might place greater emphasis on honesty, which could lead the former to judge lying to save someone's feelings in a particular situation as morally right, whilst the latter would judge this to be wrong. However, in section 7, I was clear to note that this is not necessarily the implication of the account of moral deliberation I sketched. From what I've argued, it could yet be the case that people place an identical weight on each value. To the extent that they disagree, individuals may only differ over the contextual factors which determine the relative salience which they ascribe to each value in the particular situation.

In this chapter, I argue that intercultural moral disagreement between agents cannot always be reduced to such contextual disagreement or otherwise dismissed as merely apparent, but is often fundamental, indicating a genuine conflict over the relative importance of values. In order to do so, in section 1 I survey a range of anthropological and historical evidence which exemplifies the diversity of moral attitudes between members of distinct cultural groups. Section 2 critically discusses various defusing explanations that such cultural diversity is not necessarily indicative of genuine moral disagreement from theorists such as David Brink.

Next, section 3 reviews Richard Brandt's study on the ethics of the Hopi tribe, which purports to reveal genuine moral disagreement, along with Michele Moody Adams's response. In section 4 I articulate and expand upon John Doris and Alexandra Plakias's argument in favour of fundamental intercultural moral disagreement, which draws on more recent evidence from studies conducted by psychologists/experimental philosophers such as Richard Nisbett and Steven Stich. I will conclude that this evidence provides the basis for a solid empirical case for fundamental moral disagreement between members of distinct cultural groups, before turning to the next chapter's task of vindicating intracultural fundamental moral disagreement.

1. Apparent Moral Disagreement Between Cultural Groups

First we must scrutinise the basis for the claims of those who hold that there is, in fact, moral disagreement between those of different cultural groups. The precise meaning of the term 'cultural group' is a controversial matter, with many different definitions applying depending on the discipline and theorist. In this context, I take it to refer to something like a group of people who have been exposed to similar cultural influences during the course of their life due to a shared locality, history, customs and traditions.[57]

Moral disagreement is clearly most evident between different cultural groups and as a general rule the more historically and geographically distinct one cultural group is from one another, the greater the discrepancy between their moral norms. Practices such as polygamy, ritual sacrifice and cannibalism, almost universally regarded as self-evidently immoral in European cultures since the classical age, have long been known to be endorsed and practised by peoples of other times and places. Moreover, it has become increasingly apparent through increased contact with geographically isolated communities and a greater understanding of history that such differences in moral judgements between members of different cultural groups are not rare aberrations, but the norm. This insight has been particularly buttressed through the findings from fieldwork conducted by cultural anthropologists and ethnographers, especially since around the beginning of the 20th century. These studies involved observing and integrating into the culture of societies of which little was previously

[57] It is important to note that I do not define cultural groups in terms of the shared values of those within the group. If I were to define cultural groups in this way then my later claim that individuals within the same cultural group sometimes fundamentally disagree would be false by definition.

known, and highlighted the extent to which their values and beliefs differ from that of our own.

Nonetheless, clear universal themes with regard to what ethical codes regulate can be discerned, with moral norms concerning such matters as violence, sexual relations, reciprocity, and equitable distribution of resources being particularly prevalent across cultural groups.[58] This is a significant point which is worth highlighting; there is certainly evidence of a strong degree of commonality in terms of the sort of considerations that cultural groups tend to moralise. To put it in Berlin's terminology, there does in fact seem to be a 'human horizon' of widely shared values – the range of moral diversity we encounter between cultural groups is constrained by human nature. When discussing moral disagreement we must keep in mind that it typically occurs within this backdrop of a common range of moral concerns. This point shall become more salient when addressing the question of how and why humans come to adopt the values that they do in my later chapters. Yet for now, the relevant issue is that variations between the norms which address such concerns are often found to be pervasive. In fact some have suggested that for every supposed cross-cultural moral universal one might identify, there exist past or present cultural groups which provide counter examples.[59]

To take an extreme example of a cultural group which has a seemingly alien moral code to that of most contemporary societies, let us look to an infamous portrayal of the Ik people of north eastern Uganda. The norms and practices of this group were examined and publicised by the ethnographer Colin Turnbull in his 1972 work *The Mountain People*.[60] Turnbull, who spent several years conducting first hand fieldwork within the tribe, portrays the Ik community as having a fiercely individualistic, ruthless and uncooperative cultural climate. He claims that individuals were expected to do whatever was necessary in order to survive the inhospitable conditions which they faced; selfishness to the point of theft and blackmail was not regarded as blameworthy, whilst helping others was seen as a sign of weakness rather than praiseworthy. Most noteworthy for Turnbull was his observation that bonds of familial

[58] For a review of evidence for this claim, see Sekhar Sripada, C. "Nativism and Moral Psychology: Three Models of the Innate Structure that Shapes the Contents of Moral Norms" in W.Sinnott-Armstrong (ed.) *Moral Psychology Volume 2 - The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press. (2008) p.322.

[59] See, for instance, Prinz, J. "Is Morality Innate?" in W. Sinnott-Armstrong (ed.) W. Sinnott-Armstrong, (ed.) *Moral Psychology Volume 1: The Evolution of Morality – Adaptations and Innateness* Cambridge, MA: MIT Press (2008) pp. 367 –406.

[60] Turnbull, Colin M. *The Mountain People*. New York: Simon & Schuster, (1972)

kinship were particularly weak. Whilst most cultural groups tend to place a great deal of importance on moral obligations towards one's kin, the Ik would routinely abandon their babies and elderly relatives to face starvation or to be eaten by wild animals. Children as young as three would be permanently expelled from their households, and left to join 'age bands' of their peers who would be forced to learn how to survive without adult guidance. On first blush, then, this might seem to be a good example of a cultural group which is prioritising values in a very different manner to our own.

However, Turnbull emphasised that he conducted his study during a period of mass starvation, the area having suffered from devastating drought for a succession of years, and he suggested that the cruelty and selfishness endorsed by the Ik was a consequence of such extreme hardship. It might be argued that their practices do not necessarily reflect deep moral disagreement over the importance of the values of community, family and cooperation, but a practical, Hobbesian response to desperate circumstances of scarcity. Whether such a conditional waiving of moral norms reflects actual moral disagreement or not shall be considered in greater depth later. In any case the accuracy of the conclusions and methodology of Turnbull's study has been criticised, with some arguing that the supposed selfishness and cruelty of the observed practices was not representative of the community as a whole.[61] Therefore sweeping claims about the endemic nature of moral disagreement between different cultural groups cannot be derived from such exotic and potentially unreliable examples.

Richard Miller makes this point, highlighting the difficulties in making inferences concerning moral diversity across cultures based on "pathetic refugees or demoralised remnants of a defeated society, too desperate to subject themselves to moral constraints or burdened by challenges for which they were not remotely prepared", such as the Ik, denying that they can "present informative contrasts to our standards of behavior and assessment."[62] He instead points to the Tiv people of Nigeria, who operate under far less severe conditions and maintain a fairly stable and successful way of life, yet nonetheless seem to wholly reject the relative weighting typically assigned to moral values within western cultural groups. According to the description of their culture offered by the anthropologists Paul and Laura Bohannan, the Tiv are tribal agriculturalists who live within a loose confederacy of extended family compounds.

[61] Heine, B, 'The Mountain People: Some Notes on the Ik of North-Eastern Uganda' *Africa: Journal of the International African Institute*, Vol. 55, No. 1, (1985), pp. 3-16.
[62] Miller, R.W. *Moral Differences – Truth, Justice and Conscience in a World of Conflict* Princeton: Princeton University Press (1992) p.21

There is no presiding central authority other than local councils of elders, whose role is restricted to mediating disputes in informal tribal courts, or so-called '*jirs'*. Such mediations, however, demonstrate a marked lack of concern for impartiality and what we might consider fairness in favour of promoting other values; resolutions are reached by primarily attending to the concerns of group harmony and family loyalty. For instance, marital disputes are resolved in a manner which is expected to result in the least disruption of the social order, rather than in terms of addressing who was guilty of wrong-doing within the couple or concern for the welfare of any children involved. Meanwhile, in *jirs,* family members of the disputants are expected to only provide testimony which portrays their kin in a good light rather than speaking the truth. If a witness were to willingly reveal facts which damaged the case of their family members, it would be regarded as an abhorrent violation of their moral duties towards their relatives. Similarly, the degree to which an act of violence or theft is regarded as blameworthy by the Tiv depends upon the relation of the accused to the victim; harming or stealing from one's extended kin or friends is deemed to be far more deserving of punishment than doing the same to a stranger.[63]

Of course family loyalty and group harmony are regarded as morally valuable to some extent in most societies. Some proportion of individuals within all cultural groups might even suggest that loyalty to one's family and a concern for stability can, in some instances, override the need for impartiality, and judge that the Tiv's way of resolving justice is in some respects justified. Nonetheless this attitude is publicly endorsed to a far lesser extent, at least in most contemporary western cultural groups – for the most part, people regard impartial justice and fair treatment to be paramount in such matters. We might understand, and even on some level admire, the parent who lies in court to save their child a life sentence in prison, or the judge who maintains the harmony of a community at the expense of delivering verdicts which are unfair to individuals. Yet presumably few would claim that they would be wrong to do otherwise, as would the Tiv. In contrast, there is a general attitude within contemporary western cultural groups that the morally preferable option lies in sacrificing family loyalty and stability when they conflict with the demands of justice. This indicates that although we too recognise the moral force of such concerns, as a cultural group we do not ascribe it the same level of salience as do the Tiv.

However one might object to this inference made here. For the fact that certain moral norms are publicly endorsed within a cultural group does not necessarily imply that the value ranking

[63] Bohannan, P. *Justice and Judgement Amongst the Tiv* Oxford: Oxford University Press (1968)

these norms imply is universally shared by all members of the group. An individual can outwardly claim that they endorse the moral judgements most common within their community whilst, deep down, and free from social pressure, admit that they regard them as wrong, and would prefer a different set of norms to reign. Thus one might argue that there is no conclusive basis for concluding that the Tiv generally rank values in a different order to most western cultural groups purely by looking at what moral norms they collectively follow in resolving disputes. It might be that family loyalty and group stability are valued by individuals within western cultural groups just as much as they are by members of the Tiv, relative to impartiality. It could simply be social dynamics which causes most westerners to endorse moral norms which enshrine the latter, whilst most Tiv endorse norms which promote the former.

This is an interesting suggestion, and one which I will return to later when discussing moral disagreement within cultural groups. Yet for now I take it that the fact that a particular moral norm is endorsed on a public level within a cultural group indicates that there is at least *prima facie* basis for assuming that, in general, individuals within the group morally value that which the norm promotes. Further, the particular set of moral norms which is endorsed in a cultural group can indicate the extent to which individuals within the cultural group prioritise values relative to one another. Say that a set of moral norms governing a social practice, such as conflict resolution, which realises impartiality, is generally endorsed within one cultural group, whilst a set of different moral norms concerning the same social practice, which instead realises family loyalty, is endorsed in another cultural group. In this situation, we have reason to believe that the members of the latter group *generally* value family loyalty more, and impartiality less, than members of the former group. This is entirely consistent with some individuals within each cultural group weighting group loyalty and impartiality in a different manner to their peers, and thus deep down rejecting the moral import of publicly endorsed norms in their group. On this interpretation, then, the example of the Tiv suggests that even in cultural groups where there is a relative lack of scarcity and instability, moral norms are such that they seem to reveal a very different weighting of values to that which we are more familiar with.

Moving on, to provide further evidence for the claim that there is a huge range of moral diversity across cultural groups it might help to examine how differently a single practice is morally judged in different times and places. Let us take the example of incest, a practice which is sometimes assumed to be universally condemned within all human societies and thus a good candidate for something which might be thought not to be subject to cultural

moral disagreement. Sigmund Freud famously argued that we are naturally predisposed towards incestuous desires, and that all societies had constructed artificial prohibitions upon its practice as a means of repressing them.[64] However others claim that humans possess an innate, adaptive psychological faculty which acts to minimise the incidence of incest. This psychological predisposition, it is suggested, cultivates a strong aversive response to the prospect of sexual contact with those whom we have lived in close proximity during the first three years of life, and leads to the development of taboos against it within all cultural groups. This is known as the 'Westermarck effect', after the Finnish anthropologist Edvard Westermarck, who was the first to suggest it as a universal feature of mankind.[65]

It is indeed true that incest among immediate family relations is often moralised against across cultures in some form or another, and, as I will later argue to be the case for other almost universal themes in morality, this is most likely a consequence of there being some innate disposition to develop an aversion towards it. However, the extent to which a relationship is regarded as incestuous and/or considered a moral violation varies considerably from group to group. Contemporary western Judeo-Christian cultural groups tend to take a hard moral line on incestuous relations. Marriage between second cousins is often regarded as morally suspect, and there is a strong taboo against first cousin marriage. Yet other cultural groups of both past and present have held that marriage between cousins of any degree is permissible. For instance, marriage between cousins, such as Isaac and Rebecca, is documented in the Hebrew bible without any signs of moral condemnation of the practice. Moreover, first cousin marriage is strongly encouraged in some parts of India, Pakistan and the Middle East; it has been estimated that more than half of all contemporary Pakistani couplings are composed of first cousins.[66]

Incestuous relations between more immediate family members, such as siblings, are far rarer than those between cousins in almost all cultural groups, and this may be thought to be due to the universality of moral norms condemning it. Yet the historical and anthropological record is not so unanimous; evidence suggests that brother-sister and parent-child marriage was not only the province of the elites in some pre-modern societies but that it was actively

[64] See Freud, Sigmund. *Three Essays on the Theory of Sexuality*, trans. James Strachey. New York: Basic Books (1962)
[65] See Westermarck, E. *A History of Human Marriage*. New York: Macmillan (1891)
[66] Modell, B., and Darr, A. 'Genetic counselling and customary consanguineous marriage.' *Nature Reviews Genetics*, 3, (2002) pp.225-9

encouraged at all levels of society by Ancient Zoroastrian scripture.[67] Even more significant, a comprehensive study of a diverse sample of various cultural groups by Nancy Thornhill found that only 44% had norms against immediate family incest, and those that did enforced these norms to a highly variable degree.[68] It would thus seem that although humans generally tend not to practice immediate family incest, there is yet much disagreement concerning the morality of such an act between different cultural groups. Whilst some groups condemn the practice as immoral, others regard it as morally neutral or even laudable.

So it would appear on first glance that moral disagreement between cultural groups is clear – many groups exhibit behaviour which indicates an entirely different set of moral norms from many others, whether in conditions of scarcity or not, and moral norms often taken to be universal in actuality vary between groups. This can plausibly be interpreted as reflecting the fact that different cultural groups endorse different weightings of distinct values. However it is possible to yet maintain that such diversity in moral codes is not, in fact, a consequence of fundamental moral disagreement, but of something else, which is consistent with a moral consensus on the relative weight of values on a deeper level. I will now sketch and respond to the various arguments which claim that the sort of evidence discussed above does not indicate that there is actual moral disagreement between cultural groups.

2. Defusing Explanations of Moral Disagreement

Much of the apparent moral disagreement between cultural groups such as that discussed could potentially be attributed to differences concerning non-moral knowledge, instrumental reasoning or other factors, rather than indicative of a disagreement concerning the relative importance of moral values. In the article 'How to Argue about Disagreement', Doris and Plakias note that such so called 'defusing explanations' of moral disagreement are often deployed by those moral realists who depend upon the theoretical possibility of a universal convergence on moral judgements to justify their position, and can take several broad forms.[69] Here I will relate the most relevant of those offered by David Brink in his *Moral Realism and*

[67] Scheidel, W. 'Brother-sister and parent child marriage outside royal families in ancient Egypt and Iran: a challenge to the sociobiological view of incest avoidance?' *Ethology and Sociobiology* 17: (1996) pp.319-340.
[68] Thornhill, N. W. 'An Evolutionary Analysis of Rules Regulating Human Inbreeding and Marriage' *Behavioural and Brain Sciences*, 14, (1991) pp.247-293.
[69] Doris, J. and Plakias, A. 'How to Argue about Disagreement: Evaluative Diversity and Moral Realism' in *Moral Psychology Volume 2* (2008) p.319

the Foundations of Ethics.[70] Then, following Doris and Plakias, I will draw attention to a number of examples of moral disagreement which do not seem to be accountable in terms of such explanations.

Firstly, it might be said that whilst different cultural groups in actuality possess the same weighting of moral values, different moral norms are required in order to realise them depending on the environmental circumstances which they live under. In the words of Brink, "people who live in different social, economic and environmental conditions might apply the same moral principle to justify quite different policies".[71] As alluded to earlier, this sort of explanation might be offered for the seeming absence of moral concern for one's kin and tribesmen exhibited by the Ik and other groups which are reported to leave their relatives to die, or even kill them, if they become infirm. The lack of care they show towards their kin does not necessarily indicate that they in fact regard kinship relations as of negligible moral value. Rather, they might place a high value on one's kin, and in different circumstances would regard caring for them as morally required, but are simply forced by extreme scarcity to prioritise their own survival over that of their relatives. If one were to expose any cultural group to such life threatening scarcity for an extended period of time, it might be supposed that they would react in a similar fashion and adapt their moral norms accordingly. Thus the apparent discrepancy in the moral values of the Ik compared to most cultural groups could actually reflect an underlying agreement concerning the importance of self-preservation.[72] Although we might reject self-preservation as a moral value which can legitimately override concerns for one's family in our current context this might be only because we are lucky enough to live in a society where security and adequate sustenance are taken for granted. I refer to these sorts of defusing explanation as *Varying Circumstances* explanations.

Secondly, cultural groups might live under the same environmental conditions relevant in determining whether the norm under dispute best realises the proper weighting of values, but hold different relevant non-moral beliefs or reasoning capacities. For instance one might suggest that some of those cultural groups which share moral norms prohibiting incest do so

[70] Brink, D.O. 'Moral Disagreement' in his *Moral Realism and the Foundations of Ethics,* Cambridge, Cambridge University Press (1989)

[71] Ibid, p.200

[72] As a side note, the Ik apparently did not regard their behaviour as a necessary evil given the lack of adequate sustenance; they were sometimes observed to find the suffering of the weak as amusing rather than a tragic necessity. This might cast doubt on the suggestion that they did, in fact, highly value altruism but held self -preservation to be of greater importance given the circumstances. However, one yet could suppose that their attitude towards their own behaviour did not accurately reflect their underlying moral concern.

chiefly as a consequence of their empirical understanding that the children of incestuous partnerships are prone to genetic defects, whilst those which do not moralise against incest lack such knowledge. Whilst these groups may agree that couplings between partners which are more likely to produce children who suffer from medical conditions is wrong, they simply disagree over which couplings are prone to produce such offspring. Thus, their moral disagreement might be merely apparent; if both groups shared the same non-moral understanding and reasoning regarding the relevant issue, they would potentially cease to disagree on the moral value of the prohibition on incest and so resolve their differences. This might seem fairly implausible in explaining the moral norms against incest within many small scale societies which have no understanding of genetic mutation. Nonetheless, in some such cultures non-moral understanding is directly relevant in determining which kind of inter-familial sexual relations are regarded as morally wrong and which are not. For instance, Kwayne Antony Appiah points out that within the Akan tribe, where his father was born, whilst engaging in sexual relations with one's mother's sister's child is regarded as morally abhorrent, one is actively encouraged to have sex with one's father's sister's offspring.[73] This may seem an inexplicable moral distinction until you understand the Akan's background beliefs with respect to human composition. The Akan hold that humans inherit their blood from their mothers, whilst a sort of spiritual essence, *sunsum,* is inherited from their fathers. Thus, whilst sex with one's father's sister's children is not considered sex with a 'blood' relative, sex with one's mother's sister's children is. The Akan's non-moral understanding here goes some way towards explaining why their moral judgements concerning sex with one's cousins differ from other cultural groups. Indeed, as Brink points out, this sort of explanation could potentially account for a whole range of moral disagreements which turn in some way on non-moral facts, concerning such matters as the relative malleability of human nature, the economic impact of various government policies and the truth or falsity of various religious claims.[74] From here on, these sorts of defusing explanations shall be termed *Non-Moral Disagreement* explanations.

[73] Appiah, K.A. 'More Experiments in Ethics', *Neuroethics,* (2010) pp.236-237

[74] Brink, D.O, *Moral Realism and the Foundations of Ethics,* p.203 Whether or not two parties holding different religious beliefs constitutes non-moral disagreement is a contentious issue. Many religious beliefs can be conceived of in terms of non-moral beliefs. For instance, the belief that God created the world in seven days seems to be non-moral. However, others are less obviously so, and might be regarded as in some respects as fundamentally moral – for instance, beliefs about the dictates of God, the nature of sin and the teleological structure of human nature. This chapter does not have the scope to discuss this issue in depth. For now, I will hold that particular religious beliefs can fall into either category depending on their nature, and thus one cannot simply cite differences in religious belief between parties as a form of defusing explanation for any moral disagreement.

Thirdly, it might be argued that many divergences in ethical norms between communities are attributable to self-serving biases, prejudice, or conflicts of interest, leading to a distortion of moral reasoning. Each cultural group might deep down recognise the same goods and evils, and regard them all as normatively salient to an identical extent, and yet be variable in the degree to which they are willing or able to translate these universally recognised values into their moral codes and judgements due to these factors. For instance, high status members of a cultural group might recognise that social equality is an important moral value and yet endorse a system which perpetuates extreme inequality and degradation, even slavery, because such a system benefits those in power. One might argue that at least some of the ruling classes who lived in the ancient world, for example, on some level recognised that the slavery which their civilisation was dependent upon was a morally unjustified breach of equality and human dignity. Nonetheless, most chose to wilfully ignore the wrongs that they were complicit in and cite ad hoc justifications of the practice, such as Aristotle's claim that many individuals could be deemed 'natural slaves' who ultimately benefitted from their condition[75], in order to reap the benefits which such a system afforded them. It was perhaps not the case that they truly believed that slavery was morally justified, but that it was simply convenient for them to exercise a form of self-deception over the issue. These types of explanations of apparent moral disagreement will go by the term *Wilful Ignorance* explanations. It is important to note that *Wilful Ignorance* explanations can only be legitimately posited when there is an obvious reason why an individual or cultural group would be biased or prejudiced with regards to the moral issue under dispute. Without this restriction, it would be too easy for one to attribute *Wilful Ignorance* to either or both parties engaged in a moral disagreement without adequate justification.

3. Brandt vs. Moody-Adams on the Hopi

The ethical theorist Richard Brandt, who wrote on the implications of moral disagreement for ethical theory, recognised that the anthropological evidence as it stood did not discount such explanations of differences in moral norms between cultural groups.[76] He held that prior

[75] Aristotle, *Politics,* Book 1, Chapters 4-7, trans. Reeve, C. D. C. Indianapolis: Hackett (1998) Incidentally, one might conceive Aristotle's own argument in favour of slavery being based on *Non-Moral Disagreement,* given that his conception of 'natural slaves' is derived from his teleological taxonomy of human nature. However, as mentioned in the previous footnote, it is not clear that disagreements concerning teleology are in fact non-moral.

[76] Brandt, R.B. 'The Significance of Differences of Ethical Opinion for Ethical Rationalism' in *Philosophy and Phenomenological Research* 4 (1944) pp.469-495.

to his efforts anthropology had failed to provide "an adequate account of a single case, clearly showing that there is ultimate disagreement in ethical principal."[77] This is because such disagreements as were found could not be said to have taken place under *ideal conditions*: conditions where those disagreeing could be conclusively said to possess similar levels of non-moral knowledge and reasoning capacities, be equally impartial, and to be situated within relevantly similar conditions. However, in arguably the first example of experimental philosophy, he conducted fieldwork which he claimed proved that moral disagreement can be sometimes in fact 'ultimate' or, in the terminology used here, fundamental. Brandt conducted a study to discern the source of some of the supposedly opposing moral norms of contemporary westerners and native Hopi peoples of the American southwest. He noted that within Hopi culture, children would regularly capture and keep small animals and birds as pets. However, they would treat these creatures in a way which, he claimed, most people within contemporary western society would find morally abhorrent. The children would routinely 'play' with their pets in a way which obviously caused them great suffering, broke their bones and eventually killed them. He asked the elders of the village whether they were aware of this practice, who confirmed that they were. Yet they did not seem to regard it as wrong in any way and were perplexed by Brandt's concern about it.

Brandt searched for evidence that this apparent moral disagreement concerning the treatment of animals could be explained by differences in non-moral understanding or reasoning; whether they could be explained away through *Non-Moral Disagreement* type explanations. He asked the Hopi questions such as whether they believed that the animals were capable of pain and suffering, that the animals would receive rewards in the afterlife for the entertainment they gave humans and for any other relevant non-moral beliefs which might seem to justify their ill-treatment. Yet the Hopi did not express any such beliefs: they seemingly had exactly the same non-moral understanding of the practice as contemporary westerners but simply did not attach the same moral significance to it. As a consequence of this work, Brandt concluded that there was a "basic difference of attitude," [78] between the two cultural groups that could not possibly be attributed to non-moral disagreement since "groups do sometimes make divergent appraisals when they have identical beliefs about the objects." [79]

Brandt's fieldwork would seem to rule out *Non-Moral Disagreement* type explanations of the differences in attitude towards cruelty to animals amongst the Hopi and members of

[77] Brandt, R.B. *Ethical Theory* Englewood Cliffs, N.J.: Prentice-Hall. (1959) p.102
[78] Brandt, R. B. *Hopi Ethics: A Theoretical Analysis*. Chicago: University of Chicago Press (1954) p.245
[79] Ibid, p.284

contemporary western cultural groups, as he conceived of them. Moreover, neither *Varying Circumstances* nor *Wilful Ignorance* explanations would seem to be able to plausibly account for the Hopi's toleration of animal cruelty. The Hopi's environmental circumstances are certainly dissimilar to that of contemporary westerners. However, there doesn't seem to be anything obviously relevant about their different living conditions which would explain why animal suffering would become less morally salient relative to the entertainment of children in the light of our own moral principles. Furthermore, to argue that the Hopi deep down recognise animal cruelty to be as wrong as contemporary westerners supposedly believe it to be but are simply biased or selfish enough to practice self-deception on the issue is similarly far-fetched. The elders who tolerated the practice could only be regarded as benefitting from it to the extent to that it kept their children occupied which could have been easily achieved through many other means. Moreover, there is no good justification for attributing errors in reasoning to them in a manner that does not simply beg the question. It seems plausible to presume that the Hopi simply do not value the prevention of animal cruelty as much as Brandt thinks contemporary westerners do; as he claims, their disagreement is best conceived of as fundamental.

However, in her book *Fieldwork in Familiar Places* Michele Moody Adams argues against the prospect of descriptive cultural relativism and, in particular, claims that Brandt is wrong to maintain that his work provides a conclusive example of fundamental moral disagreement between the Hopi and contemporary westerners.[80] Firstly she suggests that in general it is in fact impossible to establish that actual as opposed to merely apparent moral disagreement is occurring between members of distinct cultural groups. This, she claims, is because no matter how well one considers oneself to understand the outlook and beliefs of a different group, one cannot ever be certain whether one is ascribing the same 'situational meaning' to the same event/act. If they do not in fact attribute the same situational meaning to an event, then this alone might be the source of the disagreement. Moody-Adams takes this potential interpretation of apparent moral disagreement from Gestalt psychology theorists, such as Karl Duncker and Solomon Asch, who held there to be culturally invariant laws of ethical valuation and thus deduced that descriptive cultural relativism must be a myth. As Moody-Adams explains, "The situational meaning of any practice...would typically include a complex set of non-moral beliefs about both the "objective" (especially causal) properties and the

[80] Moody-Adams, M. *Fieldwork in Familiar Places – Morality, Culture & Philosophy*, Cambridge, MA: Harvard University Press (1997)

"subjective" features (or affective associations)."[81] To illustrate, she relates Duncker's own example of a culture wherein the elderly and infirm members are routinely killed by their own children. If this practice is interpreted by the members of such a cultural group as a means of sparing their parents a slow and painful death, or as increasing their chances of a happy afterlife, then the act is simply not the same as if the killing is interpreted as the unnecessary murder of an innocent person. Thus, when members of two cultural groups make different moral judgements about the same behaviour, one can never discount the possibility that each party is in fact assigning it a different situational meaning. If this is the case, then their disagreement might be conceived as purely one of interpretation, and does not constitute fundamental moral disagreement after all.

This might simply be taken to be a nuanced version of the *Non-Moral Disagreement* defusing explanation, yet it presents itself as more difficult to rule out through the sort of interviewing techniques which Brandt employed. For even if one establishes that a member of a different cultural group reports to hold identical factual beliefs which we might take to be relevant in making their moral judgement, on this explanation their disagreement could yet be limited to a differing interpretation of the act/event itself rather than indicate a strictly evaluative divergence. We cannot ever be sure that the same act is interpreted against an identical background of situational meaning by members of two different cultural groups. Therefore, we cannot conclusively determine that they are disagreeing about the moral judgement of the same practice, as they interpret it.

Now, it might possibly be the case that one needs to be entirely immersed within a particular cultural group in order to fully understand the *precise* situational meaning which an individual belonging to said culture ascribes to an action or event. The extent to which enculturation shapes our understanding and perception of reality is far-reaching, and it is probably impossible to regard the world in an identical manner as those brought up in unfamiliar environmental and social conditions. Nonetheless, it is a big step to claim that as a consequence of this there is a principled and impervious barrier to identifying any form of cross-cultural disagreement as being distinctly evaluative. For although we might differ in the situational meaning that we ascribe to various practices depending on the particular cultural conditioning we have been exposed to, this does not necessarily entail that we cannot attempt to evaluate the act from the standpoint of one who interprets the practice in a different manner. If one was barred from at least roughly comprehending and imaginatively adopting

[81] Ibid, p.35

the different situational interpretations of those from a different cultural group to this extent, then we could expect to encounter far more difficulties in intercultural understanding and communication than we do at present. Although we might experience some difficulty in discerning how a member of an unfamiliar cultural group is interpreting something or other, humans have historically managed to translate the differing understandings of others well enough to facilitate cross-cultural interactions, such as, for instance, trade and diplomacy. Anthropologists and ethnographers, those who are most acutely aware of just how problematic cross-cultural interpretation can be and most familiar with tackling it, are themselves generally confident that inferences concerning genuinely evaluative diversity can be made given enough training, exposure and empathic sensitivity. We should not so easily dismiss their confidence as entirely misplaced on the basis of the supposed implications of an unproven and outmoded psychological theory.

Moreover, as Jesse Prinz argues against this very point, it would be strange if it turned out that, whilst there could be the sort of extreme cultural relativism concerning non-moral understanding and interpretation that Moody-Adams envisions, there could be no descriptive cultural relativism in the field of value whatsoever. In his words, "Unless we have independent reasons for thinking values are fixed and immune to cultural permeation, we should take divergence in non-moral beliefs as evidence for the possibility of moral divergence."[82] Since Moody-Adams does not give us any such independent reasons to believe that non-moral and moral cultural relativism is different in this regard, the conclusions she draws are very much underdetermined.

However, even if this interpretation of what is going on when different cultural groups apparently morally disagree isn't taken as plausible, Moody-Adams denies that Brandt's study would provide a good example of a fundamental moral disagreement in any case. She does not claim this need be on the grounds that a defusing explanation does in fact apply in this particular case, but for the far simpler reason that Brandt misrepresents the actual ethical commitments of such a diverse group as 'contemporary westerners'. She argues that Brandt simply assumes both that 'the reader' will regard the way in which the Hopi are described as treating animals as abhorrent, and that such a disapproving response reflects a general cultural sensitivity towards animal suffering. It might indeed be true that most readers of Brandt's work on ethics would morally disapprove of allowing one's children to maim and kill animals for fun. However, Moody-Adams asks, "why should it be assumed that there will be some

[82] Prinz, J. *The Emotional Construction of Morals*, Oxford: Oxford University Press (2007) p.193

one kind of response from all imaginable readers of Brandt's *Ethical Theory*, and that such a response could legitimately be taken to represent a monolithic moral concern for animals in those reader's culture(s)?"[83] As she points out, sometimes parents in the western world also give pets to their children which they cannot possibly care for, leading to their inevitable neglect, abandonment and/or death. Moreover, whilst Brandt claims that the readers' letters column of the *New York Times* often features letters complaining of the amount of suffering which animals endure in slaughterhouses and factory farms, this is hardly representative of the whole western culture's moral outlook; many contemporary westerners remain unfazed by such practices. Meanwhile, it remains potentially the case that a minority of the Hopi do in fact disapprove of animal cruelty; Brandt's work only shows that those of the small sample he interviewed did not regard it as morally wrong.

On this point, one may respond that Brandt needn't prove that every individual westerner places a high moral value on animal welfare whilst every single Hopi disregards it as unimportant in order to make his point. He need only show that, as a general rule of thumb, the Hopi tend to place a lower value on animal welfare than members of western cultural groups for reasons which cannot be explained by simple *Non-Moral Disagreement, Wilful Ignorance* or *Contingent Circumstance,* which is consistent with some westerners valuing it lower than some Hopi. This, one might argue, would be enough to establish the existence of at least some general fundamental moral disagreement between most members of different cultural groups. I return to this point and specifically the relation between moral disagreement between and within cultural groups in my next chapter.

4. Doris and Plakias's Case for Intercultural Moral Disagreement

As Doris and Plakias note in their aforementioned article, since the work of Brandt more studies have been conducted which more conclusively establish decisive and prevalent moral disagreements between members of different cultural groups. Inspired by Richard Nisbett, who demonstrated that members of East Asian cultures tend to perceive and evaluate the world in more collectivist terms than westerners,[84] Peng, Doris, Nichols and Stich set out to determine whether this translated to similar differences in moral judgements. To do this, they presented their participants – Americans of predominantly European descent and Chinese

[83] Moody-Adams, M. *Fieldwork in Familiar Places* p.40
[84] Nisbett, R.E. *The Geography of Thought – How Asians and Westerners Think Differently*, London: Nicholas Brealey Publishing (2003)

living in China – with a vignette which described a variation of the classic 'Magistrate and the Mob' hypothetical moral conflict. It asked the reader to imagine a town where an unidentified member of an ethnic minority group is known to be responsible for a murder. The police chief and judge of the town, which already has a history of ethnic conflict, know for certain that unless they quickly find and punish the culprit, rioting against the ethnic group will swiftly follow. In order to avoid the considerable amount of property damage, injuries and deaths which will inevitably be the consequence of such rioting, they decide to arrest, convict and imprison Mr Smith, an entirely innocent member of the minority ethnic group.

Since, as I discussed in my previous chapter, ethical theories are supposed to deliver answers to such cases of moral conflict which are widely held to be intuitively satisfactory, this example is often marshalled in attacks on utilitarianism. For it is widely assumed that the putative utilitarian response to this thought experiment – that the police chief and judge were right to frame Mr Smith, as it would maximise utility – would be intuitively regarded as wrong by the vast majority of people. Indeed, the prominent moral philosopher Elizabeth Anscombe famously poured scorn on any who would even consider framing the innocent as a viable option, declaring "I do not want to argue with him; he shows a corrupt mind."[85] However, whilst this presumption proved to be correct with regard to the American participants, with the majority judging that the police chief and judge had acted wrongly and should be punished, the same did not prove true of the Chinese. In fact, Chinese participants were significantly more likely to judge that the police chief and judge did not act morally wrongly in doing what they did, and furthermore that the potential rioters were primarily responsible for the scapegoating. Tellingly however, they did *not* report that American participants universally condemned the police chief and judge as having acted wrongly, or that Chinese participants unanimously judged their action to have been permissible. A minority of Americans judged the scapegoating as morally permissible, whilst some Chinese judged it to be impermissible. The salience of this point shall be discussed in my next chapter. Yet for now, it remains significant that a general trend was discerned which reveals a moral disagreement between *most* Chinese and *most* American participants.

Peng et al. also probed the participants on some of their non-moral beliefs relevant to the hypothetical situation, such as whether they believed the scapegoat would suffer from false imprisonment, or whether the riots would cause the members of the ethnic group to suffer. They found no significant differences in the answers to such questions between the Chinese

[85] Anscombe, E. 'Modern Moral Philosophy' *Philosophy* Vol.33, no.124 (1958) p.18

and American participants, ruling out any simple form of *Non Moral Disagreement* defusing explanations. It would therefore seem that the best explanation for the differences in moral judgements between the two groups is that, at least in this hypothetical situation, the Chinese participants generally valued community stability over justice, broadly conceived, whilst the North Americans generally ranked these values in the reverse order.

Even between sub-cultural groups which both inhabit the same wider societal culture, recent studies have investigated and found general differences in moral judgements which resist defusing explanations. In an earlier work, Nisbett and Cohen set out to explore the relative attitudes towards violence between those raised in the northern and southern states of the US. Using a wide variety of evidence, such as results from various psychological experiments, survey data, crime statistics and legal practices, Nisbett and Cohen argued that southern US citizens are significantly more tolerant of the use of violence in response to affronts than their northern counterparts.[86] This is, they argue, because southerners are brought up within a 'culture of honor', whereby individuals are expected to respond to affronts with aggression in order to maintain their reputation. For instance, they found through surveys that southerners were more likely to judge that violence was 'extremely justified' in response to many different forms of offence, and to regard those who would not react violently to such offences in a negative light. Moreover, utilising an imaginative methodology, they sent hundreds of identical letters of inquiry to US employers, purporting to be from ex-convict who had been convicted of manslaughter. The letter explained that he had accidently killed a man whilst in a fight, which came about after the victim publicly gloated that he was sleeping with his fiancé and challenged him to step outside "if he was man enough". After analysing over 100 letters of response, they concluded that the southern employers tended to reply in a fashion which indicated sympathy and even respect for what he did, whilst northern employers were significantly less likely to express any such tolerance. Thus, it seems that there is a general disagreement between northern and southern US citizens over the extent to which violence is a morally appropriate response to offence.

Nisbett and Cohen's explanation as to why such a 'culture of honor' exists within the southern as opposed to the northern US states is that the former's economic base used to primarily consist of livestock herding, whilst the latter relied chiefly on agrarian agriculture. Herd animals can easily be stolen in a way crops cannot, and thus the honor culture developed

[86] Nisbett, R.E. and Cohen, D. *Culture of Honor: The Psychology of Violence in the South,* Boulder, CO: Westview Press. (1996)

as a means to deter would-be thieves at a time when the state's weak law enforcement capabilities failed to offer an effective deterrent. This might suggest a potential *Varying Circumstances* defusing explanation for the original disagreement over the appropriate exercise of violence: one might claim that although the southerners are just as morally averse to violence as northerners, their particular circumstances force them to react violently to offences in order to prevent greater moral harms. However, given that the economies of the north and south of the US are no longer so focused on agriculture nor relevantly dissimilar in any other respect, this can only possibly be offered as an etiological explanation. It can no longer function as an adequate defusing explanation of the two different cultural groups' attitudes towards violence, which persist despite the circumstances having evolved to be relevantly similar.

Doris and Plakias cite the Nisbett and Peng studies in particular as providing sufficient empirical basis for verifying fundamental moral disagreement.[87] They argue that none of the various defusing explanations offered by moral realists could possibly apply in either case. They go on to claim that forms of moral realism that rely upon the theoretical possibility of a universal convergence of moral beliefs based on shared non-moral understanding – 'convergentist' moral realisms – are undermined in light of this empirical truth. As previously stated, I need not take a firm stance on the latter point. Nonetheless, Doris and Plakias do seem to be right that such studies are compelling evidence of fundamental moral disagreement, and of the sort which would support my suggestion that different people place a different relative priority on different moral values.

Of course this is not entirely uncontroversial, with some arguing that these studies are not enough to vindicate the existence of fundamental moral disagreement. For instance, Brian Leiter contends that much of the data garnered from Nisbett's study only conclusively proves that southerners are more likely to regard violence as more *permissible* in certain contexts than northerners, and thus excusable, which does not necessarily imply that they regard it as *morally justified* as such.[88] On his view, this does not count as full blown moral disagreement. Ben Fraser and Marc Hauser agree with Leiter on this point, but go further in their critique.[89] They argue that even the cited data from surveys which directly ask the participants whether

[87] Doris, J. and Plakias, A. 'How to Argue about Disagreement' in *Moral Psychology Volume 2* (2008) pp.303-331

[88] Leiter, B. 'Against Convergent Moral Realism: The Respective Roles of Philosophical Argument and Empirical Evidence' in *Moral Psychology Volume 2* (2008) pp.333-337

[89] Fraser, B. and Hauser, M. 'The Argument from Disagreement and the Role of Cross-Cultural Empirical Data', *Mind and Language*, 25, 5 (2010) pp.541-560

they regard a hypothetical act of violence as morally justified or not does not necessarily indicate any concrete disagreement as such. For whilst Nisbett reports that southerners are more likely than northerners to *strongly* agree with statements such as 'A man has the right to kill to defend his family' (80% vs. 65%), and regard acts of violence in response to various insults as *extremely* justified (19% vs. 13%), they do not report on how the remainder of the participants answered in either case. Fraser and Hauser suggest that this data merely shows there to be a difference in degree of agreement, and does not constitute evidence of there being actual disagreement concerning the matter. For assuming that the remaining northern participants generally agreed to some extent that a man has the right to kill to defend his family, and that acts of violence in response to insults are at least somewhat justified, then the members of each cultural groups could not necessarily be said to regard each other as being in error. The northerners and southerners are still making the same moral judgements, only with the latter being more emphatic in them than the former.

Nonetheless, even if we were to accept this interpretation, on my account this yet may count as a fundamental moral disagreement. For although members of the two groups might typically come to the same broad judgement concerning whether an action is permissible or justified or not, if they attribute a different degree of permissibility or justification to an action then this could indicate fundamental disagreement concerning the relative weight of values. In this case, southerners regard violence as generally more permissible across contexts and specifically more justified when inflicted in response to insults, as well as more strongly affirming that men have the right to defend their families than do northerners. This is indicative of northerners generally placing a higher weight on the values of peace and nonviolence than southerners, and southerners taking personal honour and self/other defence to be more weighty values than do northerners. Even if the difference in degree of weighting was not enough for the two groups to offer opposing moral judgements in the cases raised by Nisbett's study, they still have value systems which are at odds with one another, if only slightly. If one were to ask more specific questions (such as 'Do people have the right to shoot unarmed burglars on their property in the head?' or 'Is it justified/permissible to punch a total stranger who has called you an asshole in a bar, unprovoked, with a closed fist?') then the underlying disagreement would no doubt make itself more clear and result in members of the two groups tending to offer different moral judgements.

Moreover, despite these objections neither Leiter nor Fraser and Hauser argue against the prospect of fundamental moral disagreement as such. On the contrary, they present no

objection to the inferences made from the cross cultural data on the Magistrate and the Mob moral conflict, and even suggest more fruitful sources of evidence for its existence. As I discuss in my next chapter, Leiter contends that the long running debates found within western ethical philosophy is evidence enough of fundamental moral disagreement. Meanwhile, Fraser and Hauser point to a study which suggests that certain rural Mayan populations do not recognise the act/omission distinction as morally salient, in contrast with the vast majority of cultural groups, highlighting this evidence as a good starting point for empirically buttressing the case for fundamental moral disagreement.[90] Thus, even those who have challenged Doris and Plakias' analysis of some of the data they use generally agree that there is sufficient reason for accepting their general point.

Conclusion

In this chapter, I have argued that fundamental moral disagreements exist between members of distinct cultural groups. Some, notably those who subscribe to convergent moral realism, have offered various defusing explanations, such as *Non-Moral Disagreement, Varying Circumstances* or *Wilful Ignorance,* for the apparent moral disagreement that is observed between different cultures. Although such explanations can potentially account for many of the moral disagreements described by anthropologists and ethnographers, more recent cross-cultural studies conducted by experimental philosophers have provided examples which are more resistant to them. I thus propose that at least some intercultural moral disagreements are best explained by appealing to the phenomenological account of moral judgement that I argued for in my former chapter, whereby individuals settle conflicts between distinct values through reference to one's own relative weighting of such values. When members of distinct cultural groups morally disagree with one another, it is often because each group tends to foster a different weighting of the relevant values amongst its members.

I now wish to develop a more controversial premise which the aforementioned Nisbett studies hint at – the possibility of intracultural moral disagreement. In the next chapter, I will make the case that fundamental moral disagreement also exists between members of the same cultural group, albeit to a lesser extent.

[90] Abarbanell, L. and Hauser, M.D. 'Mayan Morality: An Exploration of Permissible Harms' *Cognition,* 115 (2010) pp.207-224

Chapter 3 – Intracultural Fundamental Moral Disagreement

In the last chapter I argued that given recent findings in moral anthropology, there are good grounds for taking members of distinct cultural groups to sometimes be engaged in fundamental moral disagreement. I discussed various potential defusing explanations of moral disagreement as only apparent, based on either *Non-Moral Disagreement, Varying Circumstances* or *Wilful Ignorance.* I proceeded to identify some instances of intercultural moral disagreement where these defusing explanations do not seem to apply. I suggested that such instances can be more fruitfully explained by appealing to the notion that individuals of different cultural groups tend to assign different relative weight to distinct values.

Here, I will go beyond arguing for there being fundamental moral disagreement *between* distinct cultural groups and develop a case for it also existing *within* cultural groups. In section 1 I suggest that although there are some problems involved in identifying when fundamental moral disagreement is truly intracultural, depending on one's conception of what constitutes a distinct cultural group, one can still make a strong case for it by establishing its existence within both broadly and narrowly defined groups. Section 2 moves on to discuss the sorts of intracultural moral disagreements that we encounter in everyday life. I argue that although *Non-moral Disagreement* defusing explanations may sometimes apply in such cases, there are reasons to judge that this is not always the case and that it is often fundamental. Next, section 3 explores the instantiation of moral disagreement within more narrowly defined cultural groups. I acknowledge that, whilst moral disagreement is less apparent here, there is at least anecdotal evidence of it, and that there are plausible explanations as to why its true extent might remain obscured by other factors. In section 4 I present the diversity amongst the views of ethical theorists as a good candidate for fundamental intracultural moral disagreement. Finally, section 5 discusses evidence for intracultural moral disagreement from the same sort of moral anthropology studies discussed in my previous chapter.

I will conclude that although the empirical case for intracultural fundamental moral disagreement is weaker than it is for intercultural fundamental moral disagreement, we are still nonetheless justified in inferring that it exists, albeit to a lesser extent. I further suggest

that to better understand the root causes of our value pluralistic moral phenomenology as well as intercultural and intracultural moral disagreement, we must develop an account of moral psychology which is best supported by the available empirical evidence and which explains these phenomena. A presentation of such an account forms the basis of my next two chapters.

1. Identifying Intracultural vs. Intercultural Moral Disagreement

The distinction between my claims that there exists fundamental moral disagreement between members of different cultural groups, and that such disagreement also exists between members of the same cultural group, may be seen as an ambiguous and problematic one. This is for two reasons. Firstly, it might be said that there is a tension between the two claims. By emphasising the extent to which members of different cultural groups tend to morally disagree with one another, one might suggest that I have implied that cultural groups are internally homogeneous in terms of the amount of moral concern that they ascribe to values. For instance, one might interpret my previous discussion of the Tiv as implying that every individual within the group places greater moral weight on family loyalty and group stability than every individual within contemporary western cultural groups. However, to the extent that this is the case it was but an unfortunate consequence of trying to emphasise the extent of intercultural moral disagreement. As I have tried to stipulate, my claims regarding moral disagreement between cultural groups are not intended to portray members of such groups as morally homogeneous. Whilst I do suggest that individuals within cultural groups are typically more likely to place the same sort of weight on values as each other relative to those from different cultural groups, this is not supposed to imply that all individuals within the same cultural group will weight values in the same manner. The claim is merely that the average member of the Tiv values group stability and family loyalty relative to impartiality more than the average westerner. This is entirely consistent with some Tiv weighting these values in a manner more similar to the average westerner in contrast with the majority of their community and vice versa, which would constitute fundamental moral disagreement within a cultural group.

Secondly, it is not always clear where one cultural group begins and another ends. For instance, although 'North Americans' and even 'westerners' in general are sometimes spoken of as if inhabiting a single, relatively homogeneous cultural group, one can meaningfully categorise such groups as consisting of two or more distinct sub cultures, as Nisbett's work

on the differences in cultural attitudes between northern and southern US citizens highlights. Moreover, people can be said to be members of multiple overlapping cultural groups, especially in contemporary pluralistic societies. For instance, one might belong to a particular ethnic, religious and socio-economic cultural group as well as a national or even regional one. Given this, it seems that one could draw the boundaries of cultural groups as loosely or tightly as one likes. We could then speak of 'North Americans' as a single cultural group, or insist that we speak only of such sharply delineated groups such as 'Anglo-Saxon, working class, protestant West Virginians'. This makes the project of establishing the existence of fundamental moral disagreement within 'cultural groups' a potentially vague one. For whilst it might be relatively easy to make the case that such disagreement exists within broad cultural groups such as 'North Americans' (indeed, if one accepts Doris and Plakias' analysis of Nisbett's data which I discussed in the previous chapter, then the case has already been made), it is a lot trickier to establish it within more narrowly defined cultural groups. As a consequence of this ambiguity, it may seem impossible to conclusively determine whether fundamental moral disagreement within cultural groups has ever truly been established or not. Nonetheless, I will attempt to make the strongest case possible for it given this limitation.

To clarify my aims, my thesis as a whole argues for the TEA model of moral psychology, whereby although socio-cultural influences are strong determinants of one's particular interpretation and weighting of moral values, it is not the only one. In later chapters, I will argue that emotion plays a crucial role in moral judgement and that humans are predisposed to recognise a similar range of moral values as a consequence of our widely shared innate emotional repertoire. Because not only our moral concepts but also our emotional tendencies are subject to change through environmental influences, the relative weight that we assign to values is shaped by our cultural environment. However, given that humans possess innately variable emotional capacities, to some extent those subject to the same sort of culturally shared environmental influences might yet differ in the relative weight that they ascribe to values. If this is indeed the case, then whilst we should expect moral disagreement to be far more ubiquitous and of a stronger degree between those brought up within the same cultural groups, we should also see some signs of it within them, however tightly defined they may be. There may be more moral agreement within any particular cultural group, but we should not expect to see a perfect moral consensus. It is therefore incumbent on me to establish the existence of fundamental moral disagreement both between and, to a lesser extent, within cultural groups.

2. Everyday Intracultural Moral Disagreement

The first source of evidence for fundamental moral disagreement within cultural groups can be derived from simply observing the kind of ethical debates which occur in everyday discourse. Whilst there may be a general moral consensus on many issues within most cultures, there are yet ethical controversies which provoke strong disagreement between members of the same cultural group, broadly defined. In the western world, these include such issues as abortion, capital punishment, euthanasia and same sex marriage. Given that the parties in disagreement over such issues reside within the same cultural group, and thus share similar environmental circumstances, *Varying Circumstances* defusing explanations can be largely ruled out here. Moreover, although each party might charge the other of *Wilful Ignorance*, there usually does not seem to be any obvious reason why either party would have a vested interest in maintaining a certain moral position with regard to many of these issues. Someone might suggest that pregnant young women who want abortions might have purely self-interested reasons to hold that abortion can be morally justified, or that career criminals might have non-moral reasons for morally opposing capital punishment. Yet, these sorts of cases are a rarity. Most of those who take a moral position in these sorts of issues have no obvious personal stake in the matter at hand. For instance, many of those who either morally condemn or support euthanasia are healthy individuals with no sick friends or relatives, and many heterosexuals support same sex marriage. This leaves it difficult to explain their disagreement in terms of the distorting influence of self-interest.

However, as mentioned earlier, one could claim that these sorts of disagreements are best explained by *Non-Moral Disagreement* defusing explanations. There certainly are disagreements over non-moral facts which relate to these cases. For instance, there are debates concerning whether capital punishment does in fact constitute an effective deterrent, and whether foetuses become sentient earlier than 28 weeks after conception or not. Such non-moral considerations do certainly indeed bear on these issues, in terms of contributing to one's understanding of the nature of the trade-off of distinct values that is at stake. For instance, if it is true that capital punishment does not effectively deter crime, then one who holds the capacity to deter as a morally salient consideration in the application of criminal punishment has less reason to endorse it than if it is does, in fact, provide an effective deterrence. Moreover, if one takes sentience to be a salient factor in determining the moral worth of a being, then the age at which a foetus achieves it is relevant in deciding up to what point after conception abortion can be justified. Nonetheless, I contend that at the core of intracultural

moral disagreements one can often identify a difference in how each party prioritises such values which cannot be resolved via such non-moral factors.

For instance, in the case of capital punishment, some individuals accept that the evidence indicates that it does not constitute an effective deterrent, and yet still maintain that it is nonetheless morally justified. As evidence for this, a survey of police chiefs across the US found that although the participants rated the efficacy of a capital punishment as a deterrent as extremely low, they also were far more likely to support it than the average population.[91] This is most plausibly because, although they might take deterrence to be morally valuable, they ultimately believe that the value of retribution alone is more important than the value of the lives of those found guilty. In contrast, one who opposes capital punishment might concede that it effectively deters potential criminals, but that the right to life of criminals is stringent enough to override the moral benefits of its implementation. As a consequence, judgements concerning the moral status of capital punishment cannot necessarily be held to be entirely dependent on one's views regarding its capacity for deterrence. Similarly, two parties might agree, at least for the sake of argument, that foetuses are sentient before they reach 28 weeks old. However, whether or not they believe that abortion can be morally justified still hinges on matters which cannot be resolved empirically, such as whether the value of women's control over their bodies trumps that of the foetuses' life. For instance, Judith Thomson famously argued that even if it were uncontroversial that foetuses possessed all the capabilities of an adult human, the right to abortion would still be justified.[92] As a consequence, I maintain that the best explanation of such moral disagreements is that those who dissent simply weight the conflicting values at stake in a different manner, rather than their moral disagreement being grounded in non-moral disagreement.

More generally, one could maintain that a significant proportion of political disagreements within all cultural groups, past and present, can be attributed to such differences in value weighting, and not just the more obviously ethically contentious examples cited above. When individuals differ, for instance, on the extent to which we should redistribute wealth, restrict free speech or engage in humanitarian intervention, they do not necessarily disagree on the costs and benefits of taking such measures. They often seem to merely disagree over whether the distinct benefits are worth the distinct costs, all things considered, in each instance. In other words, such political disputes as those I refer to above are best characterised in terms

[91] Deiter, R. "The Death Penalty is not an Effective Law Enforcement Tool," in Schonebaum, S.E (ed.): *Does Capital Punishment Deter Crime?* San Diego: Greenhaven Press, (1998) pp.23-27
[92] Thomson, J.J. 'A Defense of Abortion' *Philosophy and Public Affairs,* Vol.1 No.1 (1971) pp.47-66

of each party to the dispute disagreeing about not only how best to go about achieving certain ends, but on what ends are most morally important.

However, can we ever be entirely sure that such fundamental moral disagreements of these sorts actually exist? One might claim that there is no conclusive example of a moral disagreement taking place within a cultural group that cannot potentially be attributed to some kind of non-moral disagreement(s), if one looks hard enough. To a certain extent, this is a fair point. Although philosophers and psychologists such as Brandt, Peng and Stich have attempted to rule out *Non-Moral Disagreement* defusing explanations of moral disagreements between members of different cultures through empirical investigation, there have been no equivalent studies of disagreements between members of the same culture. Secondly, as mentioned in the previous chapter when discussing claims that intercultural moral disagreements are only apparent, it is impossible to know for certain that each party involved in moral disagreement are in possession of the same situational meaning of an act or event. We might point to instances which seem to indicate disagreement over the relative importance of values rather than the most effective means of realising that which we all value to an equal extent. Nonetheless, we might yet be missing some background interpretational difference which is in some way relevant in determining the differing moral judgements.

On the second point, whilst we might not be able to prove conclusively that those within the same cultural group are attributing the same situational meaning to the same practice when they morally disagree, this is less of a problematic assumption than it is between those of different cultural groups. As I argued in the previous chapter, there is sufficient reason to doubt the case that the situational meaning of the same acts varies enough between members of different cultural groups to fully explain their moral disagreements. There is even more justification for discarding this defusing explanation as applicable between members of the same cultural group. The chief reason Duncker and, later, Moody-Adams posit a difference in situational meaning as a possible explanation for apparent moral disagreement is that they believed that one's pattern of enculturation in large part determined the meaning that one attributed to a practice. When two individuals have been subject to a set of similar socialising processes they are therefore likely to interpret the same practice in the same way. Thus, even if one accepts that a differing interpretation of the situational meaning of a practice sometimes explains a differing moral judgement towards it, this is rarely applicable in terms of explaining disagreements within cultural groups.

However the first point is the more important objection, and must be considered carefully. It is especially worrying given that, anecdotally, it seems that an individual's moral judgement that a certain practice is justified or unjustified typically correlates with the relevant non-moral beliefs that would lend credence to their attitude. To take a case study, despite my suggestion above, most of those who support capital punishment do in fact tend also to report a belief that it is an effective deterrent, whilst those who oppose it generally deny that it effectively deters. Nonetheless, I contend that such common correlations do not necessarily entail that the moral judgements of such individuals are being driven by their non-moral beliefs. In actuality, the path of causation often runs in the opposite direction. This is because of the influence of a well-documented psychological tendency of humans called the 'confirmation bias' and related 'attitude polarisation' effect. The confirmation bias causes individuals to unintentionally gather, remember, evaluate and interpret evidence selectively, in a manner which supports their prior commitments and beliefs. Where the evidence is ambiguous they tend to interpret it as confirming their position as correct. This results in attitude polarisation, whereby disagreement becomes more rather than less extreme as different parties consider evidence relevant to the issue under dispute. This bias has been informally observed as an aspect of human reasoning since as far back as Ancient Greece: the historian Thucydides noted that "it is a habit of mankind ... to use sovereign reason to thrust aside what they do not fancy."[93] However, in the 1960s Peter Wason conducted a series of empirical psychological studies on human reasoning which highlighted the effect as particularly pervasive, coining the term confirmation bias and using it as an explanation of his results.[94]

Furthermore, although this psychological quirk has been noted as affecting all aspects of information processing, it has been discerned to be particularly prominent when it comes to the interpretation of evidence bearing on one's political and ethical opinions. In 1979, a study led by psychologist Charles Lord was conducted to research the impact of the attitude polarisation effect upon those who held strong views on capital punishment.[95] In the experiment, they first gathered data on participants' attitudes towards capital punishment and their views on its efficacy as a deterrent, then selected those who were particularly strongly opposed or in favour of the practice. These participants were divided into small groups and

[93] Thucydides, Crawley, Richard (trans) The History of the Peloponnesian War, (431 BCE) The Internet Classics Archive, http://classics.mit.edu/Thucydides/pelopwar.mb.txt

[94] Wason, P.C. "On the failure to eliminate hypotheses in a conceptual task", Quarterly Journal of Experimental Psychology (Psychology Press) (1960) 12 (3) pp.129–140

[95] Lord, C. Ross, L. and Lepper, M. 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence', Journal of Personality and Social Psychology, (1979). 37 (11) pp.2098-2109

presented with a short statement of empirical evidence from a purported piece of research, which either supported or disputed the notion that capital punishment is an effective deterrent. They were then given more information about the research projects which produced the evidence they were offered, including critiques and counter-critiques concerning their validity. The participants were finally subjected to the same procedure, this time presenting them with research which supported the opposite conclusion to that of the first piece of evidence they were given.

The results showed that participants tended to hold their original position on capital punishment more strongly when presented with evidence which supported their initial beliefs. However, they were far more critical of the research which disputed their prior beliefs, typically judging it as poorly designed and conducted, and thus invalid, whilst taking the research which confirmed their views on face value. As the experimenters note, this is not necessarily an irrational response, depending on how substantiated one's initial belief is: "When an "objective truth" is known or strongly assumed, then studies whose outcomes reflect that truth may reasonably be given greater credence than studies whose outcomes fail to reflect that truth."[96] Yet the participants were willing to interpret evidence in a biased manner to support the very same belief that justified the interpretational bias. They thus ran the risk of ensuring that their beliefs are unfalsifiable, and based upon an initial, unsubstantiated conviction. This conviction was itself not based upon a fair and balanced weighing up of the non-moral evidence, but rather driven by a prior moral commitment. Ultimately, the attitude polarisation effect led to the participants being more likely to hold their initial attitude towards capital punishment more strongly than before, after being presented with both sets of evidence. The results of Lord's study therefore suggest that providing people with an identical range of ambiguous non-moral information relevant to making a moral judgement actually results in stronger disagreement, rather than convergence, in cases where both parties already hold existing moral beliefs about the matter at hand.

From this we have good reason for doubting that the moral disagreement which we find within cultural groups is entirely the consequence of each party having a different non-moral understanding of the relevant issues. For it seems that disputing parties at least some of the time tend to assess and acquire non-moral knowledge in order to support their moral judgements, rather than basing their moral judgements on an even-handed evaluation of the evidence. This is not to say that the latter is impossible. There are certainly cases where one

[96] Ibid, p.2106

might genuinely change one's moral judgements when they are introduced to new information which suggests that their previous understanding of the nature of the trade-off of values was incorrect. However, given that those who live in a broadly similar cultural environment generally have access to the same range of non-moral information, this does not seem to be an adequate explanation of much of the moral disagreement within societies. The disagreements are often fundamentally moral; they are concerned with the relative weight that each party assigns to different values. The different non-moral knowledge which each party tends to report is often a red herring which does not play the chief role in causing the disagreement at hand, but is rather a mere consequence of their prior moral commitments.

3. Moral Disagreement in Narrowly Defined Cultural Groups

Moral disagreement, then, is evident within large scale cultural groups such as the USA and Britain, and we have good reason to believe that in some cases it is fundamental. However, in contrast to the level of disagreement we typically find when looking at national cultures as a whole, more sharply defined sub-cultural groups tend to share a stronger consensus on such ethically contentious issues. For instance, the beliefs that capital punishment is justified and that abortion, same sex marriage and euthanasia are morally wrong are widespread amongst southern, rural, working class and deeply religious communities in the US, whilst they are rarer within richer, urban and more secular east/west coast cultural groups. This is undoubtedly in large part due to the aforementioned impact of enculturation in shaping the weight one assigns to values, and thus determining one's moral views. According to my model of moral psychology, which I will elaborate in later chapters, the more similar the environmental influences individuals are subjected to, the more likely that such individuals will weight values in the same manner and thus agree on particular issues. Given that the sort of environmental influences individuals within tightly defined sub cultural groups are exposed to are more likely to be similar, we should expect there to be more moral consensus within narrowly defined sub-cultural groups than within wider, societal cultural groups. Still, as noted earlier, on my view we should still expect there to be some fundamental moral disagreement within those raised in even strictly defined cultural groups.

Yet this relative moral homogeneity within sub-cultural groups might also be partly explained by other factors. For a start, even if we did accept that sub-cultural groups are in fact generally morally homogeneous, this would not necessarily indicate that moral disagreement cannot emerge between those subjected to the same cultural influences. This is for the reason that

any moral disagreement which does arise might be swiftly resolved through the dissenting minority party leaving the cultural group. Iconoclast individuals, or groups of individuals, may voluntarily choose to leave the sub-cultural group in which they were raised and to join or form new sub-cultures whose value weighting more closely reflects that of their own, at least in more socially mobile societies. For instance, take someone born into a cultural group in which moral norms which maintain values associated with conservatism, such as order, tradition and security, were publicly endorsed. If such an individual rejected the importance ascribed to such values in their community, and regarded tolerance, liberty and equality as of greater moral salience, they might attempt to leave their cultural group and instead embed themselves within one of a more liberal bearing. Given the relative ease with which one may relocate and immerse oneself in a different community in, for instance, contemporary, pluralistic western societies, this is always an option for people who feel that their own moral views do not cohere with those of their native cultural group.

This being the case, the fact that there is less moral disagreement within sub-cultural groups than within societal cultural groups might, in part, reflect individuals culturally migrating into groups which endorse the set of values that they most relate to. This is particularly true of those cultural groups which define themselves primarily in terms of the values that they adhere to, and whose membership mostly consists of those who voluntarily join precisely because of the values that they endorse. Thus, the fact that members of the hippie counter-culture are almost unanimous in their endorsement of liberal as opposed to conservative values does not threaten the underlying point that I am trying to make when citing moral disagreement within cultural groups. As I note earlier, ultimately I am trying to show that cultural influences are not the only determinant of an individual's values. If some individuals who have been exposed to differing cultural influences reject their native cultural group and come together to form a new cultural group in light of their shared values, this only goes to provide further evidence for my argument. All I need claim is that at least some of those who are born and raised within the context of this new group will sometimes morally disagree.

Moreover, there are other ways in which one might account for apparent moral consensus within groups of those who have been exposed to similar cultural influences, and most especially within narrowly defined sub-cultures. One important factor is the probability that at least some of those who profess to share the same moral judgements of the majority of their sub-cultural group are merely reluctantly conforming, and in actual fact disagree with them. This suggestion was raised in my previous chapter's discussion of the Tiv, where I posited the possibility that some members of this cultural group might only be endorsing the

prevalent moral norms on a public level, whilst privately rejecting them. There I made the point that an individual publicly endorsing a particular moral norm or judgement which is dominant within the cultural group is not necessarily indication that the individual is deep down in agreement with the ranking of values that the norm or judgement entails. They may in reality weight values differently and thus fundamentally disagree with the moral outlook endorsed by their cultural group, yet not have the courage or conviction to register their disagreement openly. I noted that we still have reason to accept that, in general, most individuals within the cultural group who openly endorse the common moral judgements of the cultural group are doing so authentically. It is too easy to suggest that those who express a prioritisation of values in contrast to one's own are only doing so under social duress and deep down agree with our own. Nonetheless, given the extent to which humans are prone to social conformism, one certainly can't discount the likelihood that a minority of those who claim to support an apparent moral consensus are in actuality silently dissenting.

Yet even within narrowly defined cultural groups there still remains open moral disagreement which can be interpreted as being fundamental. Anecdotally at least, many admit to encountering at least some moral disagreements over certain issues with individuals from an almost identical cultural background. Although we might tend to morally agree with those from a similar background more often than with those hailing from a less familiar sub-cultural group, often we find ourselves in fierce ethical debates with them. Of course, one might dismiss this as a peculiarity of the pluralistic, liberal and tolerant cultural groups that we happen to inhabits. Whilst moral disagreement might exist in such groups, it is certainly not the norm universally, or so the claim would go. However, the tendency for humans to wrongly perceive cultural groups of which they are not a part to be more homogeneous than their own is also a known socio-psychological bias.[97] This is called the 'out-group homogeneity bias', and has been proven in various studies to lead individuals to erroneously judge that members of different ethnic and cultural groups share similar traits, whilst attributing a higher level of individual variability within their own group. It is therefore not surprising that we might assume that, whilst our own cultural group is morally diverse, moral disagreement is non-existent within other cultural groups, given our tendencies to stereotype out-groups. Given all this, we should be wary of dismissing the notion that fundamental

[97] See, for example, Halsam, S. A.; Oakes, P. J.; Turner, J. C.; McGarty, C. "Social categorization and group homogeneity: changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness" British Journal of Social Psychology 34 (2) (1995) pp.139-160.

moral disagreements are non-existent within cultural groups, even narrowly defined ones, on the basis of mere social perception.

4. Moral Disagreement Amongst Ethicists

Let us move now from moral disagreements found within general society to those between individuals who specialise in the study of ethics. As alluded to in my previous chapter, Leiter highlights the history of western philosophy as an alternative source of evidence for fundamental moral disagreement, which provides many prominent examples of ethical controversies. He points out that "Even the last hundred years of intensive systematic theorising about ethics has done essentially nothing to resolve fundamental disagreements between, for example, deontological and consequentialist moral theories. For this observation, we do not need empirical studies; we just need to know the history of philosophy."[98] These sorts of disagreements might be conceived as particularly good candidates for being fundamental, given the onus on specialised ethical theorists to acquaint themselves with all the relevant non-moral facts, avoid self-interest, bias and other errors of reasoning and focus primarily on the formulation of universal, generalizable ethical principles rather than their context-dependent applications. Therefore, at least in principle, they should not be easily explainable through *Non-Moral Disagreement, Wilful Ignorance* or *Varying Circumstances* defusing explanations. Although ethical theorists sometimes accuse each other of the former two factors to explain why their opponents have erred in identifying the uniquely correct ethical principle(s), they often do so with no real basis other than the persuasion of their own opposing intuitions. What's more, these disagreements often take place within what one might regard as prime examples of relatively homogeneous sub-cultural groups. Since classical times, philosophers have been accused of being cut from the same cultural cloth of the isolated intellectual elites of the time. Today, the lack of cultural diversity reflected within university philosophy departments remains a constant worry – the majority of academic philosophers are still upper/middle class males of western upbringing and liberal cultural backgrounds. However, this does nothing to prevent the sort of disagreements over fundamental principles which have occurred among similarly situated ethical theorists since long before utilitarians and Kantians drew their lines in the sand.

[98] Leiter, B. 'Against Convergent Moral Realism' in *Moral Psychology Volume 2* (2008) p.336

One might object here that these sorts of disagreements, although concerning the nature of ethics, are not of the kind that I am looking for. Up to now I have been considering disagreements concerning whether certain acts, behaviours and practices are morally justified or not. Yet the disagreements between moral philosophers which Leiter highlights are primarily theoretical rather than applied. The different parties do not necessarily disagree about what is or is not justified, but upon what set of moral principles we should adopt to determine the rightness and wrongness of particular acts. For instance, a utilitarian and Kantian could both agree that the rule of law must be upheld impartially, yet the former might do so on the ultimate basis that this general rule maximises happiness in society, whilst the latter on the basis that it respects the dignity of persons. In fact one might claim that there is very little disagreement amongst contemporary ethicists upon what is right and wrong in particular instances, only about how to go about determining right and wrongness. This is a legitimate worry, as it is certainly true that there is more agreement with regard to, for instance, the permissibility of abortion and impermissibility of capital punishment amongst philosophers than there is concerning the theoretical justification of such stances.

However there certainly are areas of disagreement within the realm of applied ethics which are not purely theoretical. To offer a few examples, amongst philosophers who discuss the ethics of war, there is deep disagreement over such issues as under what conditions, if any, employing torture, exercising self-defence or engaging in humanitarian intervention is morally justified. Even when the vast majority of philosophers do agree that a certain action is right or wrong, there are those who attempt to defend a minority position. For instance, in a famous article Judith Jarvis Thompson assesses the moral justification of acting in various so called 'trolley problems'. These are situations whereby saving a group of individuals standing on a train track who are about to be killed by a runaway trolley necessitates causing the death of a single individual in some manner or other.[99] In the standard variation, one could flip a switch to redirect a runaway trolley onto another track with only one person there in order to save the lives of five individuals. In another, one could push a heavy man off a footbridge above the track, killing him in order to stop the runaway trolley and save the five. Although most who discussed the article agreed with Thompson's normative analysis of the first two formulations of the problem, there were those who dissented, arguing that her judgements regarding the moral permissibility acting in the two variations were wrong.[100]

[99] Thomson, J.J. 'Killing, Letting Die, and the Trolley Problem', *The Monist*, 59, (2), (1976) pp.204-217

[100] For instance, both Peter Singer and Peter Unger have argued on different grounds that there is no morally relevant difference between the two chief variants of the trolley problem. (Singer, P. 'Ethics and

Moreover, there was more significant dissent over the more complex variations of the trolley problem she discussed. For instance, in one variation of the trolley problem, the trolley will hit and kill five individuals unless you flip a switch which will redirect the trolley onto a looping side track. Stood upon this side track is a heavy man, who will stop the trolley from re-joining the main track at the cost of his life, and thereby prevent the death of the five. Some philosophers have used this as suggestive that the Kantian prohibition on using people as a means to an end can't quite explain the differences in our responses to the original two problems, since they assume we unanimously and intuitively believe it's right to use the heavy man as a means to saving the five in this case. Yet Michael Otsuka simply rejects the notion that flipping the switch in the looping track case is permissible, arguing, in contrast, that such a judgement is counterintuitive.[101] Thus, although philosophers tend to spend much of their time debating the potential implications of widely shared ethical intuitions about particular cases with regard to ethical theory, they do not always shy away from straight forwardly rejecting their opponents intuitions as mistaken.

5. Evidence from Moral Judgement Surveys

Moreover one can also detect signs of moral disagreement within cultural groups by analysing the data collected from various experimental studies in the field of moral psychology. The experimental designs of moral psychologists often involve testing folk responses to hypothetical cases of moral conflicts through surveying participants on whether they judge a potential action in a given situation to be right or wrong, as in the Magistrate and the Mob study conducted by Stich et al. Such studies attempt to discern general trends in the moral judgements of various populations and use these trends to inform their theories of the moral psychology of humans. To take a paradigm example of this methodology in action, the studies of John Mikhail ask participants to make judgements on which actions are morally permissible in the different variations of Thompson's aforementioned trolley problems. On the back of the data gathered, Mikhail argued that humans are 'intuitive lawyers' when it comes to making moral judgements, generating them via an innate, unconscious "moral grammar" that is analogous in some respects to Noam Chomsky's model of the innate

Intuitions', *The Journal of Ethics* (2005) pp.347-349, Unger, P. 'Causing and Preventing Serious Harm.' *Philosophical Studies* 65 (1992) pp.227–255)

[101] Otsuka, M. 'Double Effect, Triple Effect and the Trolley Problem: Squaring the Circle in Looping Cases', *Utilitas*, 20 (1) (2008) pp.92-110

linguistic grammar which enables language development.[102] Mikhail stressed the lack of variation in responses to his survey from members of very different nationalities, ages, levels of education and gender, arguing that moral principles such as the act/omission bias are not subject to cross cultural moral disagreement.[103]

In a later chapter I will critically discuss Mikhail's model of moral psychology in greater detail. For now, what is more relevant in analysing Mikhail's own data is that whilst there is a general consensus amongst the majority of participants on which actions are morally permissible and which impermissible, supposedly regardless of their cultural background, there is a statistically significant minority of respondents who dissent in each case. For instance, in the standard case whereby a bystander notices a train about to hit and kill five individuals and has to choose whether or not to flip a switch that will redirect it onto a line to kill a single individual, Mikhail reports that 90% of those tested thought that it was morally permissible for the bystander to flip the switch. Meanwhile, in the 'footbridge' variation, only 10% thought that it was permissible to push a heavy man off a bridge onto the path of the oncoming train, killing him but stopping the train and saving the five. It was thus found that Thompson's judgements were generally reflected by popular opinion: she had argued that it would be intuitively right to kill the one to save the five in the first case but wrong in the second case. However, just as a minority of philosophers have disagreed with Thompson here so too did some of the participants. In each case the 10% minority is betraying signs of a fundamental moral disagreement with the majority; there was no perfect consensus on what would be morally permissible. Meanwhile, other variations of the hypothetical moral conflict had far slimmer majorities – in the looping track case, a majority of only 52% of those who responded judged that it would be morally impermissible to kill the one to save the five given the particular circumstances. Moreover, as in the disagreements found by Peng et al. between respondents to the Magistrate and the Mob moral dilemma, there does not seem to be any obvious way in which the various defusing explanations could apply. The participants were all presented with identical relevant non-moral information and environmental circumstances in the vignettes, and there was no plausible reason why they would exercise wilful ignorance in making their judgements.

[102] Mikhail, J. 'Universal Moral Grammar: Theory, Evidence, and the Future', *Trends in Cognitive Sciences*, 11, (2007) pp.143-152

[103] Hauser's study on rural Mayan's moral judgements in trolley problems discussed in the previous chapter might suggest that this is empirically contentious.

Although Mikhail is right to suggest that his work demonstrates the moral importance that humans generally attach to principles such as the act/omission distinction, the evidence suggests that not everyone attaches the same amount of significance to them in determining overall moral judgements such as permissibility. Thus, even data which is supposed to prove the universality of certain moral principles also indicates that there is still disagreement over the relative importance of such principles, even amongst members of the same cultural group. Mikhail's study is just one example of this, but most psychological studies which reveal general trends in human moral judgement also show signs of such minority dissent; very rarely is there 100% agreement on any moral matter surveyed. A meta-analysis of a wide range of moral psychology studies would be required in order to reveal precisely how pervasive and robust this observation is. Nonetheless, this example alone goes a long way towards substantiating the existence of some fundamental moral disagreement, even within cultural groups.

Of course one might maintain that despite the above evidence of ethical controversies within cultural groups, the suggestion that such disagreement is fundamental is yet empirically underdetermined. One could object that the mere observation of moral disagreement within contemporary western society is still too indeterminate to count as proof that such disagreement is truly fundamental. Perhaps, they might say, such disagreement could in fact be explained by a sophisticated range of defusing explanations, or at least attributed to intercultural disagreement, if one takes a sufficiently fine grained definition of a cultural group. It might be more difficult for one to similarly dismiss the ethical debates found within the history of philosophy as either not being fundamental or not occurring between those of the same cultural group, yet one might charge that there is no conclusive evidence that strictly prevents such an interpretation. Finally, one could dispute the claim that the evidence of minority dissent within cultural groups from the data collected by moral psychological studies, such as Mikhail's, really indicates fundamental moral disagreement amongst the participants. One might attempt to explain away such apparent dissent as reflecting insignificant 'random noise' in the results, the fact that some of the participants do not understand the questions asked, or that the survey methodology utilised in such studies does little to reveal the actual moral intuitions of those surveyed, as some have argued.[104] None of these potential objections is entirely without basis, and it has to be admitted that each individual source of evidence in favour of fundamental moral disagreement within cultural

[104] See for example Cullen, S. 'Survey-Driven Romanticism', *Review of Philosophy and Psychology,* (2010) pp.275-296

groups can't be deemed conclusive when considered in isolation. However, I contend that their combined force gives us sufficient reason to accept the existence of fundamental moral disagreement within cultural groups as an inference to the best explanation. Until more evidence or sophisticated argument arises challenging the notion that fundamental moral disagreement does in fact exist, or that such disagreement can only possibly be explained by cultural influences, we can provisionally assume that it exists within, as well as between, cultural groups.

Conclusion

In this chapter I have provided a robust case for fundamental intracultural moral disagreement. It may indeed be difficult to provide a conclusive proof, given the complexities involved in defining cultural groups and ruling out various defusing explanations. Yet although there tends to be more moral consensus within cultures than between them, especially the more narrowly we define a cultural group, there are nonetheless good reasons for taking there to be intracultural fundamental moral disagreement. Some of the moral disagreement within cultural groups could be masked by conformism, or through individuals leaving groups which don't share their weighting of values to join those that do. Moreover, anecdotal evidence suggests that moral disagreement does indeed occur between those of very similar cultural backgrounds. Although we might explain this as stemming from *Non-Moral Disagreement* given the different non-moral beliefs disputants in such disagreements tend to hold, this might only indicate the influence of confirmation bias. Amongst the relatively culturally homogeneous field of ethicists there is not only disagreement concerning the theoretical basis of value, but normative disagreement which indicates variation in the weighting of values. Finally, whilst the results of moral judgement surveys typically establish a degree of commonality between the judgements of participants, they also invariably indicate some level moral disagreement, even between those of the same cultural group. Taking all this evidence together, I hope to have convinced the reader that we have a strong presumptive basis for taking moral disagreement within cultural groups as sometimes fundamental.

In the next two chapters, I intend to provide a sketch of an account of human moral psychology which has the capacity to explain these conclusions and is independently supported by empirical evidence. More specifically, I will attempt to account for the following facts: 1) Agents experience a widely shared range of values as distinct across cultural groups,

2) There are vast disagreements over the relative importance of such values between cultural groups and 3) Some such disagreement emerges and persists even within cultural groups. My general claim will be that humans possess a psychology which is predisposed to recognise and internalise the moral norms and value ranking of the environment in which they are socialised. However, I also posit an important role for innate affect mechanisms in shaping our moral values, help account for both the general cross-cultural themes in morality and for the moral disagreement found within cultural groups.

Chapter 4: The TEA Model of Moral Psychology

The latter two chapters argued that at least some of the moral disagreements which we encounter both between and within cultural groups are best understood as fundamental. This is to say that they are not easily conceived of as subject to defusing explanations, such as *Non-Moral Disagreement, Varying Circumstances* or *Wilful Ignorance*. Rather, the most plausible explanation is that such disagreements are the consequence of different individuals ascribing different relative weights to the distinct considerations which they consider to be morally salient. In making my argument, I reviewed and analysed a range of empirical evidence garnered from psychological and anthropological studies. I further suggested that in order to better understand the root causes of fundamental moral disagreement, it would be useful to further explore the conclusions that moral psychologists have come to regarding how the human mind comes to attribute moral significance to various considerations.

One thing to emphasise from the outset is that my approach in this regard is guided by my previous discussion on the scope of moral disagreement. There I claimed that there is some broad similarity in the moral values of all cultural groups, but that there is also fundamental moral disagreement, as a consequence of diversity in the weighting of values. This disagreement occurs most obviously between members of different groups but is also evident between members of the same cultural group. This motivates the general thought that our moral psychology is a product of some kind of interaction between innate psychological factors and cultural forces. There may be other explanations for the fact that certain values are widespread across many distinct cultural groups, but that such values stem from innate psychological features is most plausibly the case. Moreover, the fact that intercultural moral disagreement is more evident than intracultural moral disagreement indicates that some, but not all, of our value weighting is shaped by enculturation. I thereby take there to be a presumption in favour of a psycho-cultural interactionism account of moral psychology. The challenge is to elaborate on this, and determine just what kind of interaction is involved.

In this chapter, I will attempt this task. Section 1 briefly discusses Elliot Turiel's moral/conventional distinction. In section 2 I move on to explicate Shaun Nichols position on the psychology of moral judgement. In the course of doing so, I sympathetically highlight his contention that emotion plays a necessary and prominent role in individual moral

judgement and cite various sources of evidence in support of this claim. Section 3 introduces Nichols' 'Affective Resonance' hypothesis, which holds that emotion also has an important part to play in the cultural evolution of moral norms. Whilst I agree with Nichols' general framework, in section 4 I argue that it leaves more to be said about the precise relation between emotions, values and norms. In order to address this explanatory deficiency, in section 5 I turn to Jon Haidt's account of the emotional foundations of moral concern. Section 6 explains how the fundamental insights of both Nichols' and Haidt's models of moral psychology can be combined to help furnish my own proposed TEA (Two-stage Enculturated Affect) model of moral psychology. This model explains both individual moral judgement and the cultural construction of moral concepts.

I conclude by suggesting that the sketch of moral psychology which I present has the potential to help in our understanding of moral conflict and fundamental moral disagreement. It is well supported by a wide range of empirical evidence and, as such, we have good grounds for accepting it as a working model of moral psychology to adopt in order to explain the conclusions I drew in my previous chapters. In the next chapter, I will go on to discuss some competing models of moral psychology. I will argue that the account presented here is to be favoured for independent reasons, but also gains credence over the alternatives insofar as it is better situated to account for the phenomenology of moral conflict, along with intercultural and intracultural moral disagreement.

1. The Moral/Conventional Distinction

After the work of Kohlberg mentioned in the introduction, empirical investigation into the psychology of moral development continued with the work of Elliot Turiel, who designed and conducted the moral/conventional task to record the emergence of moral judgement in children.[105] He found that when presented with an example of a prototypical moral transgression, children as young as three displayed a signature pattern of responses which tended to cluster together in a law-like fashion. These response patterns systematically differed from those which they offered towards transgressions of prototypical conventional rules. Participants consistently judged that moral transgressions, which typically involved harm, injustice or violation of rights, were, in contrast to violations of conventional rules,

[105] See Turiel, E. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press. (1983)

universally wrong independently of whether any authority permitted it or not, or whether it occurred in a different locality or period in history. For instance, participants typically judged that it would be okay to talk in class if the teacher allowed it, or within a different community which had no rule against talking in class. However, they denied that it would ever be okay to hit or pull another child's hair, regardless of the locality or if the teacher permitted it. Furthermore, moral transgressions were judged by participants to be more serious than violations of conventional rules. Finally, the justifications participants offered as to why moral transgressions were wrong invoked the harm or injustice suffered by the victim, whereas conventional transgressions were explained as wrong by virtue of their potential to upset the social order.

The moral/conventional task was replicated many times across a wide variety of cultural groups and similar response patterns were reliably elicited from many diverse sets of participants. However, in some variations, transgressions which did not involve harm, injustice or rights violations were found to evoke a signature moral response pattern amongst members of certain cultural groups. For instance, Jon Haidt found that members of low socio-economic status groups in Brazil and the USA judged washing a toilet bowl with the national flag and masturbating with and then cooking and eating a dead chicken to be serious, authority independent transgressions which were universally impermissible.[106] Yet although members of some cultural groups might treat transgressions involving disgusting or disrespectful acts as prototypically moral rather than conventional, this only suggests that individuals exposed to certain environmental influences take a wider range of considerations to be morally salient than those exposed to different conditions. It only expands the range of norms which people may treat as prototypically moral, depending on their cultural group, rather than directly threatening the existence of the moral/conventional distinction as psychologically real.[107]

---

[106] Haidt, J. Koller, S. and Dias, M. 'Affect, culture, and morality, or is it wrong to eat your dog?' *Journal of Personality and Social Psychology, 65,* (1993), 613-628.

[107] Some recent studies have produced evidence which does purport to challenge the existence of the moral/conventional distinction – see Kelly, D et al. 'Harm, Affect and the Moral Conventional Distinction' in *Mind & Language*, Vol. 22 No. 2 (2007), pp. 117–131. I will not discuss this counter-evidence in detail, except to note that although Kelly et al. show that certain instances of harm do not elicit a pattern of moral response amongst participants, the instances of harm they tested for do not count, in the context presented, as obviously unjustified transgressions as such. The moral conventional task is not aiming to show that a moral response pattern will be necessarily be elicited when a participant is confronted with any instance of harm, only instances of clearly unjustified harm which transgress a social norm. See Sousa, P. 'On Testing the 'Moral Law'', *Mind and Language* 24 (2009) pp.209-234. Thus, there is still a solid basis for us to conclude that the distinction between moral and conventional transgressions is one which people are typically psychologically disposed to make.

2. Nichols' Account of Affect and Moral Judgement

Shaun Nichols takes the aforementioned studies on the moral/conventional distinction as the starting point for articulating his thesis that emotional responses are heavily implicated in both individual moral judgement and the cultural development of moral norms. He promotes his account within a range of articles, but the most thorough-going account can be found in his book *Sentimental Rules – On the Natural Foundations of Moral Judgement.*[108]

The evidence collected from the moral/conventional distinction task suggests that humans are almost universally psychologically disposed to treat certain norms as more morally salient than others. Nonetheless, James Blair found that individuals diagnosed with psychopathy and children with psychopathic tendencies, in contrast with individuals diagnosed with other types of mental disorder (notably autism), consistently fail to draw the moral/conventional distinction.[109] As Nichols notes, it would seem that this evidence undermines those accounts of moral psychology which ground moral judgement primarily upon perspective-taking and mindreading, or 'mentalizing'.[110] Such psychological abilities are used in attributing, understanding and imaginatively adopting the mental states of other individuals. Since some of the major characteristics associated with autistic individuals are impaired mindreading and perspective-taking abilities, the proven ability of autistic individuals to successfully draw the moral/conventional distinction implies that these abilities are not necessarily the most important psychological capacities required for making characteristically moral judgements. Rather, the evidence suggests that the capacity for moral judgement is primarily facilitated by psychological features which are impaired in psychopathic individuals in particular. Such individuals typically suffer no impairment in their cognitive mindreading or perspective-taking abilities and yet exhibit a reduced empathic response towards others.[111]

---

[108] Nichols, S. *Sentimental Rules – On the Natural Foundations of Moral Judgement,* Oxford: Oxford University Press, 2004

[109] See Blair, R. 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath.' *Cognition*, 57. (1995) and Blair, R. 'Moral Reasoning and the Child with Psychopathic Tendencies' *Personality and Individual Differences*, 26. (1997)

[110] Ibid. pp.10-11 Nichols later holds that the fact that autistic individuals can successfully draw the moral conventional distinction is insufficient evidence for completely disassociating mindreading ability from moral judgement. He highlights evidence which suggests that although autistic individuals have impaired mindreading abilities, they can represent at least some mental states of others, given that they possess the ability to attribute simple emotions and desires to others. As such, drawing the moral/conventional distinction might yet actually require some minimal mindreading ability that autistic individuals do possess. Yet this does not challenge the central point, which is that mindreading ability is not the only, or even most the important, psychological capacity required for forming moral judgements.

[111] Blair, R. et al. 'Theory of Mind in the Psychopath.' *Journal of Forensic Psychiatry,* 7 (1996) pp.15-25

Drawing on a range of evidence, Blair maintains that a major psychological feature which is impaired in psychopathic individuals is the Violence Inhibition Mechanism (VIM).[112] The VIM is hypothesised as a feature of the psychologies of a range of social mammals and activates in response to the representation of distress cues or action types associated with distress cues. This elicits a withdrawal response, and through what Blair refers to as 'meaning analysis', is experienced by the subject as affectively aversive in nature. On Blair's account, this aversion is what leads to the drawing of the moral/conventional distinction and represents the key component in making characteristically moral judgements. Those norm transgressions which activate the VIM generate an aversive response on the part of individuals, who then go on to identify such transgressions as seriously and universally wrong independent of authority, in contrast to conventional transgressions which do not activate the VIM. Psychopaths and children with psychopathic tendencies, however, have a defect in this mechanism, as evidenced by their abnormally low physiological response to distress cues, whilst autistic individuals show no such deficit in their distress response.[113] Thus, psychopaths fail to make the moral/conventional distinction as a consequence of their tendency not to experience the aversive affective response associated with the activation of the VIM. Meanwhile, autistic individuals can typically draw the distinction since whilst they may have impaired perspective taking abilities, their VIM generally remains intact.

Nichols argues that Blair's research does much to highlight the crucial role that affective mechanisms play in moral judgement. Most accounts of psychopathy agree that the core symptoms are explainable in terms of emotional deficiencies. Moreover, given the evidence that Blair cites, the impairment of something like the VIM is plausibly an important feature of psychopathy (although most accounts suggest that this is not the only affective impairment characteristic of the disorder).[114] Yet, as Nichols explains, psychopathic individuals do not seem to possess any characteristic impairment of their general reasoning capacities.[115] In contrast psychopathic individuals are often highly adept at deliberation and manipulation, which they utilise in order to advance their own self-interest; they simply lack the inclination

[112] Blair, R. 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath.' *Cognition*, 57. (1995)
[113] See Blair, R. et al. 'The Psychopathic Individual: A Lack of Responsiveness to Distress Cues?' *Psychophysiology* 34 (1997) pp.192-98, Blair, R. 'Reponsiveness to Distress Cues in the Child with Psychopathic Tendencies.' *Personality and Individual Differences* 27 (1999) pp.135-45 and Blair, R. 'Psychophysiological responsiveness to the Distress of Others in Children with Autism' *Personality and Individual Differences* 26 (1999) pp.477-85
[114] See Blair, R. Mitchell, D. and Blair, K. *The Psychopath: Emotion and the Brain*, MA: Blackwell Publishing (2005) for an account of the broader range of affective impairments involved in psychopathy.
[115] Nichols, S. *Sentimental Rules,* pp.77-82

to moderate their behaviour out of concern for others. It would thus seem that their inability to draw the moral/conventional distinction, as well as their associated lack of moral regard exemplified by their behaviour, stems chiefly from some sort of affective impairment.

As we have seen from Haidt's study on disgust based transgressions, members of some cultural groups typically offer a prototypically moral pattern of responses to transgressions which we would not expect to activate the VIM. It therefore seems likely that prototypically moral judgements can sometimes be motivated by affective aversions other than those generated through distress cues specifically. This will be discussed further in section 3, where I articulate Haidt's account of the emotional foundations of moral judgement. In any case, the evidence from the moral/conventional distinction task points to the conclusion that our affective responses have a strong causal role in facilitating the moral judgements of individuals. Before moving on, though, I will first highlight some further evidence for this proposition.

Haidt is another moral psychologist who emphasises the extent to which moral judgement is governed by affective psychological mechanisms which form our moral intuitions. In contrast to Kohlberg, he suggests that our explicit moral reasoning has only a very weak role in shaping our moral judgements. To the extent that reasoning does have a causal role in moral judgement, Haidt maintains that it lies mainly in the moral reasoning which others within the individual's social group offer for their judgements. For the most part, private, conscious moral reasoning only serves to provide post-hoc rationalisations and justifications for the moral judgements that the individual has already intuitively arrived at. In his famous and influential 2001 paper, 'The Emotional Dog and its Rational Tail', Haidt presents this 'Social Intuitionist' model of moral judgement, which refers to an impressive range of empirical evidence to support his claims.[116]

Haidt's major source of evidence for the automatic, unconscious nature of moral judgement comes from his own research into 'moral dumbfounding'. Dumbfounding occurs when an individual experiences a strong moral response towards something, and yet is unable to provide adequate reasons to justify the judgement that they deliver. Haidt observed this phenomenon upon conducting a study where participants were presented with the following story:

[116] Haidt, J. 'The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' *Psychological Review. 108*, (2001) pp.814-834.

> 'Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other.'

Participants tended to immediately judge that Julie and Mark acted wrongly, before being asked to offer reasons for their condemnation. Yet the reasons typically offered were not applicable to this specific example. For instance, they would claim that it was wrong because of the deleterious genetic effects of inbreeding, or because it would cause emotional harm. Yet the story specifically stipulates that Mark and Julie used two forms of birth control and were not in any way traumatised by the event. Confronted with the inapplicability of their objections, participants would often, stutter, laugh and concede that they could not provide any satisfactory justification for their moral condemnation. Nonetheless, they would continue to forcefully maintain their original conclusion that Mark and Julie's actions were impermissible. On the basis of this, Haidt claims that moral judgements cannot always be the product of conscious deliberation, but often stem from gut responses which we later try to rationally justify to ourselves and others, sometimes unsuccessfully.

The fact that we often do not have conscious access to the processes which determine our moral intuitions alone does not prove that they are a product of emotion rather than reason. Nonetheless, Haidt maintains that these intuitions are affective rather than cognitive in nature. He highlights a range of evidence to support this claim, but the most important comes from studies which he himself conducted. In one such study Haidt and his colleague Thalia Wheatley manipulated participants through hypnotic suggestion to feel a pang of disgust when confronted with affectively neutral words, such as 'take' or 'often'. They were then asked to read stories which contained the words 'take' or 'often', then make moral judgements regarding them. Participants made higher ratings of both disgust and moral condemnation towards those stories which contained the word that they had been primed to feel disgust upon encountering. This suggests both that the hypnotic suggestion technique did successfully manage to manipulate the disgust of the participants, and that disgust had at least some causal role in moderating the strength of the moral judgement they arrived at.[117] The

[117] Wheatley, T. Haidt, J. 'Hypnotic Disgust Makes Moral Judgements More Severe' *Psychological Science* Vol.16 No.10 (2005) pp.780-784

evidence for this claim has been replicated through experiments which employed alternative means of priming for disgust. These studies found that participants who had been exposed to disgusting smells and environments whilst making moral judgements tended to make more severe judgements, especially towards transgressions of norms regulating certain behaviours (such as drug consumption and sexual promiscuity), than those who hadn't been so exposed to disgusting stimulus.[118] Moreover, although most of the research on the causal influence of emotion on moral judgement has concentrated on disgust, priming for other emotional states such as anger, anxiety, sadness and amusement has also been found to have an influence.[119]

Further evidence for the role of affective responses being causally involved in moral judgement comes from the work of Joshua Greene. Greene conducted experiments which subjected participants to functional neuroimaging (fMRI) whilst engaging with the sort of trolley problems mentioned in chapter three, along with other moral conflicts. He found that trolley problems which involved personal harm violations (such as the footbridge variation mentioned in the previous chapter) activated areas of the brain associated with emotional response. [120] Moreover, areas of the brain associated with cognitive control and working memory were activated amongst participants who gave characteristically utilitarian judgements (such as pushing the heavy man off the bridge to stop the trolley) in such scenarios, and their reaction time was slower. Greene argues that this supports a dual-processing model of moral judgement, whereby characteristically deontological judgements are produced by an automatic, emotional system, whilst characteristically utilitarian judgements are driven by more controlled cognitive processes. This neatly explains the differences in the neural activity and reaction time between individuals who came to opposing moral judgements. According to Greene, those who made deontological judgements went with their emotional gut instinct, whilst those who made utilitarian judgements experienced an emotional response, but through exercising cognitive control managed to override it and deliver a utilitarian judgement.

---

[118] Schnall, S., Haidt, J., Clore, G.L., Jordan, H. 'Disgust as Embodied Moral Judgment.' *Personality and Social Psychology Bulletin*, 34 8, (2008) pp.1096-1109.

[119] See Horberg, E.J, Oveis, C, Keltner, D. 'Emotions as Moral Amplifiers: An Appraisal Tendency Approach to the Influences of Distinct Emotions upon Moral Judgment' *Emotion Review* Vol.3 No.3 (2011) pp.239-240, Perkins, A. M. et al. 'A Dose of Ruthlessness: Interpersonal Moral judgment is Hardened by the Anti-anxiety Drug Lorazepam' *Journal of Experimental Psychology: General*, 142, (2013) pp.612–620 and Youssef, F. et al. 'Stress Alters Personal Moral Decision Making', *Psychoneuroendocrinology*, 37, (2012) pp.491–498.

[120] Greene, J.D. et al. 'An fMRI Investigation of Emotional Engagement in Moral Judgment'. *Science* 293, (2001) pp.2105–2108

This position gains further support from evidence that brain injury patients with emotion-related damage in the ventromedial prefrontal cortex (VMPFC) are prone to make highly utilitarian judgements when faced with moral dilemmas. On Greene's account, these patients experienced no emotional response when faced with personal harm violation. The resulting lack of conflict between their emotional and cognitive systems then led them to arrive at the default utilitarian judgement. This entails that emotional response does indeed have a causal role in moral judgement, at least insofar as it causes some individuals to come to deontological moral judgements. In fact, this represents just one example of a wide range of evidence from various brain injury patients which all implicate the importance of the VMPFC and other areas of the brain associated with emotion shaping interpersonal and social reasoning. [121]

Finally, there is also evidence that the particular emotional dispositions of individuals can predict their moral judgements in particular areas. In a study by Inbar et al., it was found that participants who ranked highly on a measure of general disgust sensitivity were more likely to make judgements which indicated an intuitive moral disapproval of gay men kissing in public. This effect did not occur when participants were asked about public kissing between heterosexual couples. The more disgust sensitive participants were, the stronger the effect was, despite most participants not offering explicit moral condemnation of homosexuality.[122] This suggests that one's particular affective dispositions have the capacity to influence one's moral judgements, at least on the implicit level.

Considered collectively, this evidence strongly supports Blair's contention that affective responses play an important role in shaping moral judgement. Nonetheless, Nichols points out that there is an important explanatory gap within Blair's account; its failure to account for the distinction between judgements of badness and judgements of moral wrongness. Badness and wrongness are no doubt related, insofar as an action is often considered wrong to the extent that it involves badness. However, although we might judge certain events which would activate the VIM or other negative affective responses as bad, we might not necessarily judge them to be wrong.[123] We do not, for instance, consider pains such as toothache or natural disasters as 'wrong', nor make moral judgements when we witness accidental injuries. So although something like the VIM and other affective dispositions might help explain how

[121] Koenigs, M. et al. 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements' *Nature* 446, (2007) pp.908–911

[122] Inbar, Y. et al. 'Disgust sensitivity predicts intuitive disapproval of gays.' *Emotion* 9 (3) (2009) pp.435-439

[123] Nichols, S. "Norms with Feeling: Towards a Psychological Account of Moral Judgment." *Cognition,* 84 (2002) pp.221-236.

we come to judge something as bad, it is not sufficient for explaining why we judge some of these things as also being wrong. Nichols therefore suggests that there is a further psychological mechanism underpinning the capacity to make characteristically moral judgements; the ability to retain a body of internally represented rules prohibiting certain behaviours, which constitutes a 'normative theory'. Thus, one can fail to make a characteristically moral judgement by either lacking the relevant component of one's normative theory or the affective mechanism which facilitates the distinction between conventional norms and moral norms. Although psychopathic individuals are capable of developing such a normative theory through social transmission, their lack of an affective response to the transgression of certain norms explains their inability to draw the moral/conventional distinction.

Nichols thereby conjectures two main features of human moral psychology; the normative theory, and the affective dispositions which facilitate our tendency to treat some norms as moral and others as merely conventional. Although we might evaluate various behaviours and events as good or bad on the basis of our affective responses to them, we also require a normative theory which prohibits and recommends certain behaviours in order to judge them as right or wrong. Where, then, do the contents of our normative theory come from? Although our capacity to develop a normative theory is presumably innate, Nichols conceives the content of each individual's normative theory itself as for the most part shaped by the cultural environment in which they are raised. Nonetheless, the norms which happen to originate and persist within cultural environments are not conceived of as arbitrary. Rather, according to Nichols 'Affective Resonance' hypothesis, the affective dispositions of individuals also influences the sorts of norms which are liable to originate and persist within a particular cultural environment, and thus end up as part of our normative theory. The next section will elaborate on this proposed account of the cultural evolution of moral norms.

3. Nichols' Affective Resonance Hypothesis

In articulating his Affective Resonance hypothesis, Nichols builds on the work of the anthropologist Dan Sperber. Utilising an epidemiological approach to cultural evolution, Sperber argues that our innate biases, which are a product of evolutionary processes, have the potential to significantly impact the development of cultural norms and beliefs.[124] He suggests that much of the innate structure of our minds is composed of domain specific learning modules which shape the learning process, making some outcomes more likely than others. During the process of social learning, the learner, or so called 'cultural child', does not absorb and retain all norms and beliefs of the 'cultural parent' indiscriminately. Rather, the process of cultural transmission of information is moderated through the innate biases generated by the cultural child's learning modules, or so called 'attractors'. These attractors render certain norms and belief more appealing and easier to detect, infer, remember and store than others. To couch it in terms of cultural evolution, those norms and beliefs which are more compatible with the innate biases arising from our learning modules gain a fitness advantage over those that aren't, and are more likely to survive and spread across the generations than those that are less congruent with our psychological predispositions.

However, Sperber is not an adherent of the meme theory of culture, which he holds takes the analogy between cultural and biological evolution too far. In contrast, he argues that "...there is much greater slack between descent and similarity in the case of cultural transmission than there is in the biological case. Most cultural descendants are transformations, not replicas.'[125] For he maintains that attractors not only influence which norms and beliefs survive the process of cultural transmission but also have the potential to modify the content of what is culturally transmitted. For instance, if a cultural parent tells the cultural child a story, that child's psychological attractors might lead them to misremember the precise details of the plot and characters, then go on to retell her own, slightly modified version which is more attuned to their attractors. Although such attractors might only have a weak effect on an individual level, over the span of generations they have the potential to significantly influence the path which cultural development follows. This provides a potential explanation as to why there are many common themes in the norms and beliefs of distinct cultural groups, despite the evident diversity between them.

---

[124] Sperber, D. *Explaining Culture: A Naturalistic Approach.* Oxford: Blackwell (1996)
[125] Ibid, p.108

Sperber, along with his followers, generally concentrate on innate cognitive psychological features such as folk biology, physics and psychology, as the most important bases of the attractors which influence cultural development. For instance, Pascal Boyer applies the Sperberian, epidemiological account of cultural evolution to explain the cross-cultural commonalities found in religious beliefs. In doing so he invokes learning biases which are the result of our innate tendencies to attribute mental states such as intentions, beliefs and desires towards sentient beings, even those which are conceived of as supernatural. [126] However, Nichols argues that our innate emotional tendencies are also prime candidates for taking the role of Sperberian learning biases in the cultural evolution of moral norms. It seems intuitive that a stimulus which elicits an affective response strikes us as more important and distinctive than a stimulus which doesn't, and Nichols suggests that this common-sense notion is to some extent vindicated from psychological studies on memory. He points to evidence which indicates that descriptions and images which elicit affective responses are more easily retained and recalled in the long term by participants than affectively neutral stimuli. For instance, in one study, participants were given lists of affectively-charged words (such as 'rape' or 'vomit') and affectively-neutral words.[127] It was found that participants were far better at recalling the affectively-charged words than the neutral words after a week. These sorts of findings have been replicated in a range of further studies, which collectively suggest that long term retention is generally enhanced when an emotionally salient stimulus is invoked.[128] The potential for long term retention of a norm or belief is one of the most important determinants of its cultural fitness. It follows that those norms and beliefs which concern behaviour that we are innately prepared to take as emotionally salient will be more likely to survive the process of cultural evolution than those concerning behaviour which we are less disposed to be emotionally aroused by.

However, Nichols does not take this for granted, but offers independent evidence for his Affective Resonance hypothesis in the form of a systematic review of the genealogy of etiquette norms. [129] Noting that historians of etiquette have emphasised the role that disgust has played in the cultural development of table manners, Nichols decided to test whether norms prohibiting actions that we are innately disposed to experience a disgust response

[126] See Boyer, P. *Religion Explained,* New York: Basic Books (2001),
[127] Kleinsmith, L., Kaplan, S. 'Paired-associate Learning as a Function of Arousal and Interpolated Interval' *Journal of Experimental Psychology* (1963) pp.190-193
[128] See Heuer, F., Reisberg, D. 'Emotion, Arousal and Memory for Detail' in ed. Christianson, S. *The Handbook of Emotion and Memory,* Hillsdale, NJ: Lawrence Erlbaum (1992) for a review.
[129] Nichols, S. 'On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution' *Philosophy of Science,* 69, (2009) pp.18-30

towards are, in fact, more likely to survive the process of cultural evolution. To achieve this he performed a statistical comparison of the norms detailed in one of the first influential etiquette manuals in Western Europe (Desiderius Erasmus's *Good Manners For Boys* of 1530*)* with contemporary etiquette norms. He employed individuals who had no knowledge of his hypothesis to code for those norms in the manual which prohibit actions which elicit core disgust (namely those involving bodily fluids, such as norms prohibiting spitting, vomiting and urinating) and those which didn't, then determine whether each norm remained part of contemporary etiquette standards. Upon performing his analysis from the independent encoder data, he concluded that those norms mentioned which prohibit disgusting actions did indeed exemplify the increased cultural fitness which he refers to. Not only were these rules more likely to survive the process of cultural evolution, but they now are so strongly embedded within our culture that it appears odd to our contemporary minds that they needed to be formally prohibited in the form of written norms at all. In his words, "While the norms prohibiting core-disgusting actions have gained in normative strength, the non-core-disgust norms have often simply disappeared from the culture."[130]

Etiquette norms concerning the particulars of table manners which do not elicit core disgust, such as those regulating the proper use and positioning of various items of cutlery, are found within Erasmus' work, and a few do indeed remain recognisable as part of current-day Western etiquette: "For instance, Erasmus tells us that: "The cup and small eating knife, duly cleaned, should be on the right-hand side" (281). This remains part of our tradition of etiquette today."[131] However, Nichols' analysis shows that such norms were far less likely to survive as a feature of contemporary etiquette standards. In all, whilst more than 90% of actions which elicit core disgust have survived the cultural transmission of etiquette norms to the present day, only 30% of those which do not elicit core disgust have survived. However Nichols does not maintain that emotion is the only force in the cultural evolution of norms and that we should therefore expect *all* affectively backed norms to survive and all affectively neutral norms to perish. He thus claims that "…this in no way threatens our hypothesis, which is probabilistic, not categorical: norms prohibiting actions that elicit negative emotions are more likely to survive than affectively neutral norms."[132]

On the basis of this evidence, Nichols suggests that other innate affective dispositions which have 'core stimuli' (such as, in the case of disgust, the sight and smell of potential

[130] Nichols, S. 'On the Genealogy of Norms', p.29
[131] Ibid, p.29
[132] Ibid, p.29

contaminants like bodily fluids or rotting meat) will also have had an impact on the cultural evolution of moral norms. Specifically, Nichols emphasises the degree to which norms regulating the infliction of harm have been shaped by the human innate disposition to have an aversive affective response upon perceiving or imagining distress cues, perhaps arising from something akin to what Blair would call the VIM. Recall that Blair hypothesised the VIM as a psychological mechanism which causes us to experience an aversive affective response when prompted by distress cues. This is a particularly good candidate for an innate affective response to certain eliciting conditions which would have the sort of influence on the cultural development of moral norms which Nichols has in mind.[133] There is much evidence indicating that such a disposition for an aversive response to distress cues develops in most humans at a very early age and is perhaps present at birth, thus representing a cross culturally universal feature of the human emotional repertoire. [134] Therefore, if the Affective Resonance hypothesis is correct, we can expect norms which prohibit actions which trigger something like the VIM to be cross-culturally ubiquitous. Given their enhanced cultural fitness, stemming from the affective backing that they enjoy, such norms are more likely to have survived and gained greater prominence in the cultural evolution of the moral systems of cultural groups.

As it turns out, we do in fact find that harm norms are extremely prominent, although not universal, amongst the wide range of moral norms that we find when we examine the anthropological record. Of course, some cultural groups do seem to accord relatively little moral weight to the prevention of harm to others. For instance, amongst the Yanomani of South America there are relatively few norms prohibiting the causing of harm. Moreover, the use of violence to settle disputes, claim resources or display dominance is commonplace and

[133] Fiery Cushman and Ryan Miller have made a compelling case for there being two distinct forms of harm aversion – 'Outcome Aversion', an aversive emotional response which derives from imagining or witnessing the consequences of harmful acts, and 'Action Aversion', a similarly aversive response which instead derives from imaginative engagement with what it would be like to inflict the harm oneself. See Miller, R. and Cushman, F. 'Aversive for Me, Wrong for You: First-person Behavioral Aversions Underlie the Moral Condemnation of Harm' *Social and Personality Psychology Compass* (2013) pp.707-718 Additionally, it might also be suggested that there is a sense in which humans are innately disposed to experience a *positive* affective response towards harm, given that many people of various cultural groups seemingly enjoy witnessing or causing the infliction of violence, especially towards those they deem as members of an out-group or those perceived to be guilty of wrongdoing. However, this possibility is not ruled out by hypothesising the influence of something like the VIM on the cultural evolution of harm norms. The claim is only that *most* individuals are disposed to experience a *predominantly* negative affective response towards harm, although this may in some instances be countervailed by perceived justifications of the harm, which might cause them to experience no negative affect or even enjoy witnessing or inflicting it.

[134] See Simner, M. 'Newborn's Response to the Cry of Another Infant', *Developmental Psychology*, 5 (1971) pp.136-150.

regarded as praiseworthy rather than morally condemned.[135] Yet this is perfectly consistent with the Affective Resonance Hypothesis. Again, according to Nichols, it is not simply the case that we are innately hard wired to morally object to the infliction of harm, and that this is reflected in the cross-cultural abundance of harm norms. Although we are innately disposed to have affective responses towards harm, and judge instances of harm as bad, these judgements of badness do not in themselves constitute moral judgements. The story is more complex; our emotional dispositions lead norms prohibiting behaviour which harms others to be particularly memorable and seem distinctive to us, improving their cultural fitness, but not *ensuring* that they become entrenched in the moral norms of every cultural group.

However, although Nichols himself neglects to make the point, we can attribute a greater role to emotional dispositions in the development of moral norms than merely that of Sperberian biases which improve the cultural fitness of affectively-backed norms. Although Nichols generally emphasises the role of affect in influencing the cultural *transmission* of norms, we might also say that affect will also influence which norms happen to *emerge* within a cultural group. For if norms which are associated with emotionally salient stimuli are more compelling to the recipients of cultural transmission, it seems extremely plausible that they will also be more psychologically appealing to the originators of cultural norms, and thus more likely to emerge as a product of cultural invention. For norms do not merely randomly occur amongst cultural groups, but at some point were constructed by individuals or groups of individuals, and are thus from the outset likely to have been shaped by the innate psychological dispositions of such individuals. This is a point which Chandra Stripada notes, suggesting that such 'origination biases' are equally important in shaping the moral norms of cultural groups as the Sperberian biases which Nichols focuses on.[136]

To sum up so far, on Shaun Nichols account our affective dispositions play two major roles in influencing our moral judgements. In the first instance, and on the individual level, the extent to which we experience witnessed or imagined events or patterns of behaviour as good or bad is determined by our affective response towards the stimuli. Whether or not the behaviour which constitutes any given norm violation elicits an affective response or not leads individuals to treat particular internalised norms differently. If the behaviour comprising the violation of a norm does cause them to experience an independent aversive affective

[135] See Chagnon, N.A. *Yanamono: The Last Days of Eden,* San Diego, CA: Harcourt Brace Javanovich (1992)

[136] Stripada, C. 'Nativism and Moral Psychology: Three models of the innate structures which shape the content of moral norms' in *Moral Psychology Volume 2* pp.332-334

response, then they take it to be a moral norm violation, and react as such. Meanwhile a norm violation which does not involve behaviour that elicits an independent aversive affective response is judged to be an instance of a conventional norm violation. Secondly, on a population level, if a norm is concerned with regulating behaviour which we are disposed to have an independent affective response towards, then it is more likely to originate and survive the process of cumulative cultural evolution over time.

We may also conceive of another sense in which affective responses have an important role in moral judgement, and this is in terms of their capacity as norm compliance enforcers. When someone judges that an individual has acted wrongfully they are typically motivated to punish the transgressor in some way, and make a further moral judgement that it is right that they are punished. The level of punishment that individuals are motivated to administer and which they judge as appropriate is proportionate to the perceived wrongness of the behaviour.[137] All this may seem entirely intuitive and obvious, but it is worth noting that recent experimental evidence in the field of behavioural economics indicates that such punitive motivations are best interpreted as at least partly *intrinsic*, rather than as purely instrumental in satisfying another end, such as deterrence. This is a point which Stich and Stripada emphasise in their account of norm acquisition and enforcement.[138] Put briefly, the evidence suggests that individuals are motivated to punish norm transgressors when it serves no further end, when such punishment costs them personally to administer, and even when they are not personally harmed by the transgression of the norm. Stich and Stripada also cite evidence that such an intrinsic motivation to punish norm violators is culturally universal, does not need to be taught and, importantly, derives its felt normative force from affective responses. This, again, is intuitive – both our other-regarding and self-regarding reactive attitudes towards wrongful behaviour are most plausibly to be conceived as stemming from emotional responses, such as anger and guilt respectively.

Note that such punitive attitudes do not exclusively apply in cases where we judge a person to have acted wrongly in terms of the prototypical moral response pattern found in the studies relating to the moral/conventional distinction. In some cases, we might judge an individual to have transgressed a norm that we take to be merely conventional (that is, authority

[137] A similar account of our disposition to praise and reward those who act in a manner we judge as morally right can be given, with regard to an intrinsic motivation ultimately deriving from different kinds of affective response.

[138] Stich, S. and Stripada, C. 'A Framework for the Psychology of Norms' in Carruthers, P. Laurence, S and Stich, S. (eds.) *The Innate Mind Volume 2: Culture and Cognition*, Oxford: Oxford University Press (2007) pp.280-301

dependent, culturally relative, taken as less important than moral norms and not regulating a behaviour we are independently disposed to emotionally respond to in itself), and yet still be intrinsically motivated to punish that individual. The very fact that a norm has been violated is often enough to provoke an affective response which motivates punishment, even if the behaviour doesn't provoke an independent affective response in itself. Therefore being intrinsically motivated to punish a person who violates a norm is not necessarily an indication that such a norm is being treated as prototypically moral. Nonetheless, given that affective dispositions which motivate punishment are typically far stronger in the case of moral norm violations, and further impinge upon when we think punishment to be morally justified, we can hypothesise that they are an important way in which affective responses play a role in the overall psychology of moral judgement.

4. The Missing Links in Nichols' Account

Nichols' account is impressive in terms of both its evidential support and explanatory power. It is grounded by a wide range of empirical evidence and provides a compelling account of why we encounter, for instance, a ubiquity of harm and disgust based norms across a wide range of cultural groups, and why these norms are typically treated as moral rather than merely conventional. Nonetheless, his articulation of the relation between emotion and moral judgement needs further elaboration if we are to fully understand how human moral psychology yields the similarity and diversity in moral judgement which I noted in previous chapters. I will now highlight three respects in which Nichols' presentation is insufficient to explain adequately the psychology behind moral disagreement. The first concerns the *extent* to which affective dispositions are in fact innate, rather than a product of cultural influences. The second concerns the *range* of dispositions which should be conceived as innate. The third concerns the relation between moral norms and values.

Firstly, Nichols focuses specifically on the role of *innate* affective dispositions in particular in shaping the development of individual moral judgement and population level moral norms. Yet he says little on how and why these dispositions are to be conceived of as innate. Are they entirely fixed from birth, or are they somewhat malleable, in that the particular environment that one is exposed to can have an influence upon their development? Whilst Nichols speaks as though it is a given that the VIM is a purely innate psychological faculty, the extent to which an individual experiences harm aversion and other affective responses towards stimuli could itself be dependent on the sort of enculturation that they have been

exposed to. If this is the case, then the story of how affective dispositions influence the cultural evolution of moral norms is far more complex than Nichols suggests. If his account is to explain the similarities and variations between the moralities of different cultures, we need a clearer understanding of the extent to which such dispositions are themselves shaped by cultural influences.

In terms of the second insufficiency, aside from his focus on harm aversion, Nichols does not offer much insight into the range of additional affective dispositions which might also influence moral judgement and cultural evolution. Yet many culturally ubiquitous moral norms are concerned with the regulation of behaviours which have little to do with the infliction of harm, such as norms regulating fairness and sexual conduct. In fact, in some cultural groups such norms are considered more morally salient than harm norms. Are these norms similarly ubiquitous, and widely treated as moral, due to the impact of affective resonance deriving from different affective dispositions? Although Nichols maintains that disgust also influences the cultural evolution of etiquette norms, he does not refer to any other affective dispositions besides this. If Nichols is to provide a comprehensive explanation of the common themes that crop up in the moral codes of cultural groups, then he ought to appeal to more than harm aversion alone.

The third lacuna which needs to be addressed regards the role that emotional dispositions have in determining the sort of moral *values* humans are disposed to internalise, as well as moral norms. In my current chapter, I have been exploring how emotion relates to moral judgement and *norms*. However, in my previous chapters, I suggested that fundamental moral disagreement, both between and within cultural groups, is best cashed out in terms of disagreement concerning the relative moral weight of values. I take a moral value to be a cluster of considerations that an agent takes to be genuinely and independently reason giving. Moral norms and values are intricately linked – a moral norm is typically advocated and followed in service of a moral value. For instance, moral norms prohibiting lying and promise keeping are designed so as to preserve the value of honesty and trust. Thus, values are more basic than norms, and a variation in moral norms can be taken as an indication of disagreement over the relative moral weight of the values that they serve. To take an example, one who denies the force of moral norms prohibiting harm is indicating less of a concern for the value of compassion towards others than one who upholds the importance of harm norms.

Nichols does imply something akin to this in his discussion of the difference between 'bads' and 'wrongs', which I highlighted earlier. On his view, emotional responses inform judgements of goodness and badness, which relate to the underlying standards of evaluation which roughly constitute what I mean by values. Norms regulating actions and behaviour which individuals judge good or bad are treated as distinctly moral norms by such individuals, as opposed to being merely conventional. A norm thus constitutes a moral norm if it regulates behaviour which our underlying values mark out as morally salient. Yet Nichols doesn't explain much about how individuals and cultural groups come to have such values. Although he clearly regards values as intimately connected with emotional dispositions, given their role as determinants of what is good and bad, it would be too simplistic to hold that a moral value *just is* a disposition to experience an emotional response in relation to obedience or violation of a norm. More needs to be said about the relation between emotional responses and values. Yet his account is chiefly concerned with the role of emotional dispositions in the development of moral norms rather than how they relate to the values which underlie them.

In order to help plug these explanatory gaps, I will now turn to Haidt's 'Moral Foundations Theory'. This offers a more thorough explanation of how emotional dispositions are shaped within human psychology, and helps flesh out the range of emotional dispositions which influence the cultural construction of moral norms. Moreover, given its focus, it allows us to better understand the link between emotion and moral values as an intermediary between our affective dispositions and moral norms.

5. Haidt's Moral Foundations Theory

As discussed earlier, Haidt, like Nichols, holds that emotion plays an important role in moral judgement. More specifically, he holds that our emotional responses to stimuli provide us with instinctive moral intuitions, which are the primary source of moral judgement. This forms part of his 'Social Intuitionist' theory of moral judgement. However, Haidt also provides a categorisation of those innate emotional dispositions that are particularly related to moral judgement, and how each maps onto a distinct range of moral considerations. He suggests that certain clusters of innate emotional dispositions lead to the cultural construction and recognition of moral values.

In the first articulation of this idea, he built upon the work of the cultural anthropologist Richard Shweder, who suggested that the major ethical concerns found across all cultural groups generally fall into one of the 'Big three' ethics of community, autonomy or divinity.[139] Haidt's hypothesis, based on empirical studies performed by Paul Rozin and himself, was that the emotions of contempt, anger and disgust map onto concerns of community, autonomy and divinity, referring to this model through the acronym CAD.[140] He held that we are innately disposed to emotionally respond with contempt to violations of community, anger towards violations of autonomy and disgust towards violations of divinity, and that as such, these emotions are responsible for our moral treatment of such concerns. However, in developing this model further, Haidt, along with Craig Joseph, later further delineated his list of emotionally-induced values. He identified those clusters of considerations which are most widely recognised as morally salient across cultural groups, along with the emotional precursors which can be observed amongst non-human primates, to inform his categorisation. This led him to divide the ethic of autonomy into the concerns of harm/care and fairness/reciprocity, whilst the ethic of community was divided between the concerns of ingroup/loyalty and authority/respect. The ethic of divinity was rephrased in terms of concern for purity/sanctity.[141]

Haidt holds that each of these five foundations of morality is constituted by a cluster of emotional dispositions, which represent an innate product of our evolved psychology. This is consistent with the psychological theory of 'basic emotions'. To elaborate: although some anthropologists, such as Franz Boas and his follower Margaret Mead, historically emphasised the extent to which emotions are socially constructed and a product of cultural conditioning,[142] evidence from the work of Paul Ekman has convinced most psychologists

[139] Shweder, R. A., Much, N. C., Mahapatra, M. and Park, L. The "Big Three" of Morality (Autonomy, Community, and Divinity), and the "Big Three" Explanations of Suffering. In A. Brandt and P. Rozin (Eds.), *Morality and health* New York: Routledge, (1997) pp.119–169.

[140] Rozin, P., Lowery, L., Imada, S. and Haidt, J. The Moral-Emotion Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Ethics (Community, Autonomy, Divinity) *Journal of Personality and Social Psychology, 76*, (1999) pp.574-586.

[141] Haidt, J. Craig, J. 'The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-Specific Virtues, and perhaps even Modules.' In Carruthers, P. Laurence, S. and Stich, S. (eds.) *The Innate Mind, Volume 3: Foundations and Future*. New York: Oxford, (2008) Note that in his later works Haidt accepts that there are probably additional foundations which give rise to moral concerns which the five listed do not cover, such as liberty/oppression, waste/efficiency and honesty/deception. He limits his focus to the five foundations because he holds that these are the ones we have the best case for, but is open to the possibility that we will eventually be able to substantiate more than these.

[142] See Mead, M. *Coming of age in Samoa: A psychological study of primitive youth for western civilization*. New York: Morrow (1961)

that there are at least some emotions which are 'basic' or 'primary'. Basic emotions are biological in origin; the genes which encode for their development have reached fixation within the human population and are thus universal across cultural groups. In 1969, Ekman investigated the extent to which emotions are basic through cross-cultural studies on the interpretation of facial expressions, building on the insights of Charles Darwin on the universal expression of emotions and their precursors in non-human animals.[143] Ekman discerned that even individuals from extremely isolated cultural groups were able to reliably identify the facial expressions of certain emotions in photographs of strangers. Moreover, they were able to match these expressions with descriptions of situations and stimuli which would normally elicit such an emotion. For instance, they would associate a facial expression depicting sadness with a situation describing the loss of a child, and the expression of disgust with encountering rotten food.[144] Given this universal recognition of such emotions there is a strong case that they are also universally instantiated in some form in human nature.

From his initial research, Ekman concluded that there are at least 6 basic emotions – anger, fear, disgust, happiness, sadness and surprise. More recently he has expanded the list of emotions he holds to be basic up to 17. It now includes emotions such as contempt, guilt, shame, contentment, excitement and amusement.[145] The suggestion is that humans are innately disposed to develop such emotions, or 'affect programs', and that they are typically elicited by the same core stimuli universally. Haidt maintains that this range of basic emotions help psychologically dispose us to attribute moral weight to certain social concerns, and these psychological dispositions constitute the five moral foundations. Similar to Nichols, he argues that the foundations are best conceived of as Sperberian 'learning modules'; that is, they are innate learning biases which act so as to constrain and shape the development of our moral values. However, he insists that the account is not necessarily tied to such a modular approach to learning capacities, adding that

[143] Darwin C, Ekman P and Rodger P. "The Expression of the Emotions in Man and Animals" Oxford: Oxford University Press. (1998)

[144] Ekman, P., Sorenson, E. R. and Friesen. W. V. 'Pan-cultural elements in facial displays of emotions.' *Science,* 164, (1969), pp. 86-88

[145] Ekman, P. 'Basic emotions' in T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*. New York: John Wiley & Sons. (1999) pp. 45-60 See also, Haidt, J. Keltner, D. 'Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition' *Cognition and Emotion*, 13 3 (1999) pp.225-266, which reports that the facial expressions of an extended range of emotions are recognised cross culturally, although the level of recognition of some of these emotions varied depending on the cultural group. This suggests that some emotions are more basic than others, so to speak, to the extent that their expressions are more readily identifiable cross culturally.

> "…readers who do not like modularity theories can think of each one as an evolutionary preparedness to link certain patterns of social appraisal to specific emotional and motivational reactions. All we insist upon is that the moral mind is partially structured in advance of experience so that five (or more) classes of social concerns are likely to become moralized during development. Social issues that cannot be related to one of the foundations are much harder to teach, or to inspire people to care about."[146]

Table 1 below provides a useful representation of Moral Foundations Theory, illustrating the adaptive challenge, proper domain, actual domain, characteristic emotions and relevant virtues for each of the proposed moral foundations. According to Haidt, each foundation evolved to solve a particular adaptive challenge, such as the emotional concern for harm/care evolving to motivate individuals to care for the vulnerable and first and foremost their offspring. Moreover, whilst each evolved to respond to a 'proper domain' of eliciting stimuli, a by-product of the evolution of such foundations is that they lead us to be disposed to emotionally respond to a wider range of stimuli than they were originally adapted to respond to. For instance, cheating individuals and inequitable resource distributions form part of the proper domain of the fairness/reciprocity foundation. In terms of their etiological function, we are disposed to morally respond to such stimuli precisely because of the adaptive benefits of doing so. Nonetheless, such an evolved emotional disposition also serves to trigger a moral response towards stimuli like broken vending machines, despite this not serving any adaptive purpose and thus not constituting part of the proper domain of the disposition. All stimuli which are, in fact, associated with the moral foundation represent the 'actual domain'; the proper domain represents the subset of these stimuli which the psychological disposition is specifically adapted to trigger a moral response towards. Furthermore, each distinct moral foundation has its characteristic emotions associated with it. Whilst there is not a clear mapping of a single emotion to each foundational value, as Haidt's earlier CAD model, each foundation nonetheless has a cluster of widely cross-culturally recognisable emotional dispositions motivating the particular concern.

[146] Haidt, J. and Joseph, C. 'The Moral Mind.' in *The Innate Mind, Vol. 3* p.383

	Harm/Care	**Fairness/ Reciprocity**	**Ingroup/ Loyalty**	**Authority/ Respect**	**Purity/ Sanctity**
Adaptive challenge	Protect and care for young, vulnerable, or injured kin	Reap benefits of dyadic cooperation with non-kin	Reap benefits of group cooperation	Negotiate hierarchy, defer selectively	Avoid microbes and parasites
Proper domain (adaptive triggers)	Suffering, distress, or threat to one's kin	Cheating, cooperation, deception	Threat or challenge to group	Signs of dominance and submission	Waste products, diseased people
Actual domain (the set of all triggers)	Baby seals, cartoon characters	Marital fidelity, broken vending machines	Sports teams one roots for	Bosses, respected professionals	Taboo ideas (communism, racism]
Characteristic emotions	Compassion	Anger, gratitude, guilt	Group pride, belongingness; rage at traitors	Respect, fear	Disgust
Relevant virtues [and vices]	Caring, kindness, [cruelty]	Fairness, justice, honesty, trustworthiness [dishonesty]	Loyalty, patriotism, self-sacrifice [treason, cowardice]	Obedience, deference [disobedience, uppitiness]	Temperance, chastity, piety, cleanliness [lust, intemperance]

Table 1 – 'The Five Foundations of Intuitive Ethics'[147]

As the final row of the table demonstrates, Haidt also holds that each foundation provides the basis for particular groups of virtues and vices. He argues that conceptions of what constitutes ethically virtuous and vicious behaviours and character traits are often culturally specific, and thus partly a product of social construction. Nonetheless, such constructions ultimately derive from one or more of the five innate affective bases for our moral intuitions.

[147] Taken from Haidt, J. Joseph, C. 'The Moral Mind.' pp. 367-391

Such bases represent the 'first draft of morality', which predispose us to experience automatic flashes of approval or disapproval when we encounter certain patterns of social behaviour, but like Nichols, he admits that these flashes do not by themselves constitute what we might call moral judgements. As Haidt puts it, "Mature moral functioning does not consist only, or even primarily, of simple affective or intuitive reactions to social stimuli. Disgust felt towards dog feces, or even towards an act of homosexual intercourse, is not in itself a moral judgment."[148] Rather, in order for us to possess full-blown morality, we must develop moral concepts such as the virtues he speaks of through the process of enculturation. My contention is that along with the virtues Haidt focuses his discussion on, we could also represent moral *values* as falling under the heading of such socially constructed moral concepts.

Haidt holds that each cultural group draws upon the five foundations of morality in a different fashion, emphasising each to a greater or lesser degree in the development of their moral concepts. For instance, in contemporary western liberal cultural groups, moral concepts are to a large extent constructed from the two foundations of harm/care and fairness/reciprocity. The values of autonomy, equality and impartial justice are given precedence, whilst honesty, kindness and fair-mindedness are taken to be amongst the chief virtues. However, in other cultural groups both past and present, the moral concepts internalised and employed by individuals stem more from the affective bases which make up the ingroup/loyalty, authority/respect and purity/sanctity building blocks of morality. In this way, different cultural groups selectively cultivate and foster the moralisation of different intuitive flashes of affect amongst their members, fine-tuning their initial first draft of morality to construct more particular and conceptually intricate value systems.

To help explain, Haidt sometimes invokes a rough analogy between cultural diversity in morality and gastronomy.[149] He compares the five foundations of morality he proposes to different types of taste buds, each acting as a receptor of a basic flavour. Such flavours provide adaptive perceptual information on the chemical composition of potential foodstuffs in the form of automatic, affectively valenced experiences which motivate us to either avoid or consume what we taste. Generally we choose to avoid eating anything with an overwhelmingly bitter or sour flavour, and this is adaptive insofar as strong bitterness

[148] Ibid, p.387

[149] See, for instance, Haidt, J. Bjorklund, F. 'Social Intuitionists Answer 6 Questions about Moral Psychology' in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press. (2008) p.202

typically indicates something being nutritionally unsuitable for humans. Food which is moderately sweet, salty or fatty, however, generally provides us with a positive taste experience which indicates the nutritional suitability of what is sampled, and motivates us to eat more of it. In a similar fashion, the five foundations act as receptors of social information which provide either positive or negative affective experiences that help regulate our motivations towards others. Humans come hard-wired with both taste buds and the affective dispositions of the moral foundations, but this is not to say that our sense of morality or gastronomy is entirely innate. Rather, through the process of enculturation, our morality is shaped in a similar way as our palate develops in response to cultural feedback. As a consequence, whilst individuals brought up in a culinary environment which emphasises the heavy use of spicing and meat might experience a dish like macaroni cheese as bland and unappetising, one who has been raised within a cultural group whose gastronomy concentrates on the use of fatty dairy and pasta products is likely to find it delicious. Similarly, whilst an individual whose cultural background cultivated impartial compassion might take an act like killing a member of an out-group as morally wrong, one encultured in such a way as to foster in-group loyalty might judge it to be morally praiseworthy.

However, beyond invoking this analogy and briefly highlighting the role of cultural narratives, Haidt is vague on what form this 'fine-tuning' through enculturation takes precisely, and more needs to be said. For understanding how, exactly, cultural forces shape the moral concepts of individuals is important when it comes to understanding why people weight values differently. I will now propose two ways in which an individual's cultural environment can influence their development of moral concepts. Firstly, enculturation can influence the development of our affective dispositions themselves. Secondly, it can help determine which affective responses we take to be morally salient, and which we consciously reject as morally irrelevant.

Let us begin by exploring the first point. For one thing, enculturation can moderate the intensity of one's affective responses and the range of eliciting stimuli. If this influences an affective disposition which forms part of a particular moral foundation, then this will go on to shape an individual's set of moral concepts. There may well be core stimuli and a similar phenomenology associated with the basic emotions cross-culturally. For instance, danger is the core stimuli of fear. Still, it is widely accepted that there remains some cultural variation

in affective dispositions. It thus seems likely that affective dispositions are malleable in the face of cultural influences: I will now elaborate three ways in which I take this to be the case.

Firstly, the cultural influences that an individual is exposed to can adjust the intensity to which the core stimuli of a basic emotion will in fact elicit an affective response. For example, if one is raised within a hierarchical society, one is likely to develop a strong disposition to experience contempt when observing another failing to demonstrate respect towards a higher status individual. Meanwhile, being raised in a more egalitarian community will dampen one's disposition towards responding with contempt in such a situation. Thus, the extent to which the core stimuli of contempt (instances of disrespect) elicit this affective response on the part of group members depends upon the cultural practices which such group members have been influenced by.

Secondly, cultural influences can expand upon the range of stimuli that will elicit an affective response beyond that of the core stimuli. This is often achieved through cultural mechanisms which act so as to associate a particular practice, institution or object with the core stimuli of a particular emotion. For instance, within many cultural groups, members of low status classes are strongly associated with core disgust stimuli, so as to justify and maintain the hierarchical structure of the group by entrenching it within the group's affective dispositions.[150]

Thirdly, there may be scope for some cultural mechanisms to influence the particular emotion that a stimulus elicits. For instance, in cultural anthropology it is generally agreed that whilst wrongdoing primarily elicits guilt on the part of the wrongdoer in Western societies, in more collectivist societies one's own wrongdoing elicits the phenomenologically and conceptually distinct emotion of shame.[151] This is even more so the case when it comes to the higher cognitive, non-basic emotions, which can be cultural specific in terms of their phenomenology and eliciting stimuli. Many of these emotions are widely recognised within a certain cultural group yet do not have an obvious correlate in others. For instance, take the Japanese emotion of 'amae', which represents a pleasurable feeling of dependency akin to that of a child towards a parent, but is experienced by an adult towards an institution.

[150] For instance, within the Indian caste system, the low status 'untouchable' caste is associated with faeces and other disgusting stimuli, whereas the high status Brahma caste is associated with cleanliess and purity.

[151] See Hiebert, P.G., *Anthropological Insights for Missionaries*, Grand Rapids: Baker Book House, (1985)

Although those from another cultural group might be able to vaguely understand and experience such an emotion, it is certainly not clearly defined or strongly felt in the way in which it generally is amongst the Japanese.[152]

It would then seem that despite the case for basic emotions representing an innate, evolved feature of our psychology, to at least some extent affective dispositions can be moulded by enculturation. However this point must not be overstated. Although our emotional repertoire is certainly not fixed from birth, neither is it purely the product of socialisation. As Haidt suggests, there is a level of innate preparedness when it comes to the development of our affective dispositions, which constrains the range of potential outcomes and ensures that some are far more likely than others. This point will be expanded upon in the next chapter.

Moving on to the second level, there is also the more cognitive reinforcement which determines whether or not the implicated affective responses are taken by the subject as legitimately morally salient. When we experience an affective response to a stimulus, we sometimes endorse it on a second order level and judge it as normatively relevant, and sometimes not. For instance, within western liberal cultural groups the experience of affective harm and inequity aversion is typically automatically endorsed as an indicator of the genuine wrongness of the stimulus. On the other hand, the experience of disgust towards, for instance, certain sexual behaviours is not necessarily taken to be appropriate, nor a good reason for judging it as morally problematic. This is part of the explanation as to why western liberal cultural groups are less likely to possess moral concepts relating to the purity domain, or weight them especially highly. Although individuals of such groups might experience the disgust responses associated with the domain, they reject its normative import.[153]

This is not to say that if a cultural group does not collectively conceive of a particular affective response as legitimately normative then its members do not let it implicitly influence their moral judgements. Recall Haidt's dumbfounding experiments, whereby participants condemned a hypothetical act of incest whilst offering inapplicable justifications for their disapproval. A plausible interpretation is that the participant's condemnation was motivated by an intense disgust response. However, since disgust-based purity is not recognised as a

[152] For more on Amae, along with an account of many other culturally-specific emotions, see Prinz, J. *Gut Reactions: A Perceptual Theory of Emotion*, Oxford: Oxford University Press (2004) pp.131-157

morally salient concern in the western world, they attempted to rationalise their in terms of a widely accepted concern, such as harm. Yet this does not entail that our conscious beliefs regarding the relative normative salience of affective responses have no capacity to influence our moral judgements. It only establishes that, in some case, these beliefs are not enough to override the moralising tendencies of our affective responses. In fact, the FMRI studies conducted by Greene attest to the ability of some individuals to override their initial emotional response to stimuli in arriving at moral judgements. Participants who judged that it would be permissible to kill one to save five lives in the 'footbridge' variation of the moral dilemma ultimately dismissed their aversive affective response as morally irrelevant, deciding to privilege utilitarian concerns as paramount in this instance.[154]

In any case, enculturation generally takes the form of both direct and indirect feedback from elders and peers within a child's everyday experience concerning appropriate affective responses and what is and what is not morally salient, but Haidt also emphasises the impact of storytelling in shaping the moral concepts of individuals. Since different cultural environments provide different sorts of feedback and stories which emphasise the moral value of different intuitive foundations, the developed moral concepts of individuals from different cultural groups will tend to differ, leading to conflicting moral judgements, values and virtues. This, Haidt contends, is the primary source of the moral diversity that we encounter between cultural groups: In his words, "Moral diversity, on our account, results from differences in moral education and enculturation."[155] I will elaborate and expand upon this explanation of the root cause of moral disagreement in my next chapter.

6. The TEA Model of Moral Psychology

I hold that the accounts of Haidt and Nichols offer compatible explanations of the general pattern of similarity in moral norms and values that we observe across cultural groups. They both highlight the role of universal innate affective dispositions along with more culturally

[154] Greene, J.D. et al. 'An fMRI Investigation of Emotional Engagement in Moral Judgment'. pp.2105–2108 In additional support of this, several studies suggest that some of the moral disagreement between liberals and conservatives results from liberals experiencing similar affective responses to social stimuli, but liberals consciously rejecting their normative import, whilst the conservatives endorse them. See Skitka, L. J., et al. (2002). 'Dispositions, Scripts, or Motivated Correction?: Understanding Ideological Differences in Explanations for Social Problems.' *Journal of Personality and Social Psychology*, 83(2), p.470

[155] Haidt, J. Joseph, C. 'Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues", *Daedalus,* (2004) p.65

specific influences in shaping the moral psychology of individuals. These accounts thus tie together the evidence I referred to earlier, which indicates both that moral judgement is intimately bound up with emotional responses and that humans are born innately prepared to learn to respond with certain emotions to certain stimuli. Nonetheless, they both leave room for the impact of enculturation as an important factor in determining our moral judgements.

Yet whilst Nichols offers a useful framework for understanding the cultural evolution of moral norms, Haidt's account provides a more developed picture of the universal affective dispositions which prompt the moralisation of certain social concerns across cultural groups. The moral concerns which derive from the proposed five foundations of morality go beyond the harm norms which Nichols focuses his discussion on, and Haidt's discussion helps to illuminate how exactly cultural forces interact with these foundations in the development of moral concepts. Moreover, Haidt's evolutionary account provides an account of how and why humans came to be universally psychologically predisposed to develop the particular affective dispositions which make up the foundations. These aspects of Haidt's theory help to plug the explanatory deficiencies which render Nichols' account problematic when considered in isolation.

On the other hand it would not be correct to say that Haidt's account is simply an expanded and improved version of Nichols'. For each focuses on different, but related, aspects of morality, leaving space for each to benefit from the other's insights, and thus combine to offer a fuller picture of moral psychology as a whole. Nichols' epidemiological account of the cultural evolution of moral norms offers a historical explanation as to why norms regulating certain types of conduct are more common across cultural groups than others. Such behaviours, he holds, invoke affective responses which make the norms associated with them more distinct and memorable for the individuals who learn them, increasing the chances that they will both emerge and persist over time. Meanwhile, Haidt's account focuses more on the cultivation of the underlying values and virtues which individuals are disposed to regard as morally salient across cultural groups, which give rise to the moral appraisals from which moral norms derive their distinct normative force. On his view, moral concepts such as values and virtues are culturally constructed through building upon the intuitive, emotional foundations, which represent an innate feature of our evolved psychology.

Now, insofar as these two accounts deal with different aspects of morality, I hold that the central points of both can be fruitfully combined to furnish a single model of moral psychology. To this end, we should conceive the cultural construction of morality as working in a two stage process. The first stage in this process is covered by my developed version of Haidt's account: some of our innate emotional dispositions are selectively cultivated and moralised, whereas others are suppressed and represented as normatively redundant by cultural influences. This leads to the cultural construction and internalisation of moral values such as compassion, equality and fairness by members of the particular cultural group. Humans are innately disposed towards developing and moralising certain affective dispositions rather than others as a consequence of evolutionary influences. Thus, most values which are culturally constructed, or at least those which persist for any length of time, could be categorised under one or more of Haidt's foundations.

It is at this point that Nichols' account can be applied to help explain the cultural development of moral norms in the second stage. For the particular moral concepts that are propounded within a cultural group go on to further influence the cultural fitness of moral norms, over and above the influence of innate affective dispositions alone, helping to determine which moral norms are most likely to both originate and persist over time. Thus, the strength of the selection advantage conferred on a particular norm is partly dependent upon the extent to which the affective disposition which gives the norm its perceived normative force has been cultivated and moralised by the particular cultural group.

Together, then, the fundamental insights from Nichols' and Haidt's accounts can be combined to provide a deeper explanation for the basis of cultural similarity in both moral values and norms than when taken in isolation.[156] I will refer to this account as the Two-stage Enculturated Affect (TEA) model of moral psychology. This name derives from the model explaining our internalisation and particular weighting of values as emerging from a two-stage enculturating process, with affective dispositions being the central mechanism through which group member's values and norms are shaped.

[156] Jessy Giroux also attempts to combine these models in a different way, arguing for a combination of Nichols 'Input' model of the role of innate emotional tendencies in shaping the evolution moral norms with Haidt's 'Output' model of how these same tendencies naturally give rise to certain moral concepts in individuals to constitute a 'Moderate Nativism' model. See Giroux, J. 'The Origin of Moral Norms: A Moderate Nativism Account' Dialogue 50 (2011) pp.281-306

To clarify how this model can explain how the moral values and norms of individuals are shaped, I will now illustrate how it is conceived to work through a pair of relatively extreme examples of groups which regulate harm in radically different ways. Let us first take a cultural group which enshrines peace as a particularly weighty moral value, and has norms which heavily restrict the infliction of violence. The Moriori people of the Chatham Islands, close to the New Zealand archipelago, were famously pacifist. They lived by a code of non-violence called 'Nunuku's law', after Nunuku-whenua, a prominent chief who established it following a series of inter-tribal conflicts.[157] This code came to be so strongly recognised that when Maori invaded the Chatham Islands in 1835, the Moriori gathered a council where the elders determined that Nunuku's law forbade any armed resistance, and were subsequently conquered and enslaved.[158]

On the TEA model, this extreme respect for peace became enshrined in the following manner. Firstly, the Moriori adopted a set of norms, Nunuku's law, which over time led to the cultural construction of peace as a moral concept within the group. This shaped the cultural environment and practices of the group, and this in turn went on to influence the affective dispositions of those raised within it.[159] Since humans are innately disposed towards developing and moralising something like the VIM (Violence Inhibition Mechanism), this bare affective response was culturally elaborated and moralised relatively easily. The innate aversion of violence of members of the group was cultivated to produce a response which was stronger than usual and elicited under a wider range of conditions, and those who experienced this aversion grew to take it to have legitimate reason-giving status. This amounts to the first stage in the TEA model of moral psychology.

Once the value of peace came to be an important part of the morality of the Moriori in this manner, the second stage kicked in. Whereas those who were first exposed to Nunuku's law might not have taken it to be that normatively weighty, it had increased affective resonance for those later generations who had internalised the value of peace from an earlier age. This

[157] Davis, D and Solomon, M 'Moriori - The migrations from Hawaiki', Te Ara - The Encyclopedia of New Zealand, found at http://www.teara.govt.nz/en/moriori/page-2 updated 13-Jul-12
[158] Ibid, 'Moirori – Impact of new arrivals'
[159] The TEA model is not strongly wedded to any particular theory regarding how such initial changes come about. As I note in my concluding remarks, one plausible mechanism by which changes in moral values occur is through high status individuals directly or indirectly influencing the rest of their cultural group to adopt the norms and values which are more in line with their particular affective dispositions. This particular case seems to be a prime example of this.

ensured that those norms which governed the regulation of violence were more memorable, and taken to have increased normative weight. This eventually reached the point where the Moriori, who had initially been a warlike people, had so much respect for the value of peace that they chose to sacrifice their lives and liberty rather than defend themselves.

This example can be contrasted with a cultural group which readily embraces violence. As previously mentioned, the Yanomani are reported to have adopted norms and values which foster warfare between tribes and violence against one another at the slightest provocation. The TEA model again explains such a situation as being perpetuated by the social environment suppressing the development and narrowing the eliciting conditions of the VIM amongst its members. Whilst this would likely leave at least some members of the group still experiencing an aversion to violence, they would be encouraged to reject the normative import of this affective disposition. With non-violence thereby not having been culturally constructed as a value of the group, norms aiming to regulate the infliction of violence are less likely to originate and persist amongst them. If such a set of norms were to be introduced, say by an authoritative leader like Nunuku, the current generation would likely only grudgingly accept its normative import. Given that the VIM is something we are innately disposed to develop and attach normative significance to, over time the change in cultural environment may shift the group's value system. Nonetheless, in the short term norms regulating violence would suffer from poor cultural fitness, and it would require some generations before the values of the group were adequately in sync with the norms in such a way that they would reliably persist over time.

Note that the TEA model presented here is only a sketch, and is it is not by any means intended as an exhaustive account of moral psychology. Many other psychological features are clearly complicit in the myriad different aspects of moral thought, and despite my concentration on affective dispositions, this is not to dismiss the impact of more cognitive processes on moral judgement. Nonetheless, it is meant to capture the major mechanisms through which moral values and norms originate, are internalised and persist across generations. As I will show in my next chapter, this is enough to serve my purpose of explaining the phenomena of moral conflict and disagreement.

Conclusion

In this chapter, I have advocated an affect based account of moral judgement, which I call the Two-stage Enculturated Affect (TEA) model. I suggested that a range of evidence from studies in moral psychology, psychopathology and neuroscience implies that affect has a strong causal role in shaping moral judgement. Such findings, I have argued, are best captured by Nichols and Haidt's complementary accounts of moral psychology. Although neither is sufficient when considered on its own, the fundamental insights of each can be combined to give a deeper overall picture of how innate dispositions and cultural forces interact to mould our moral values and norms. On this picture, our innate affective dispositions are to a certain extent malleable in terms of intensity, range of eliciting stimuli and perceived normative weight in the face of cultural influences. This enculturating process gives rise to the members of cultural groups possessing moral concepts such as virtues and values, and the presence of such concepts within a cultural group partly determines the extent to which particular moral norms have cultural fitness.

However, I accept that these considerations alone might not be taken by everyone to constitute a firm enough basis to ground the TEA model of moral psychology. Even taking on board the independent evidence from psychological studies which support such an account, it could be said that the case for it is still far from conclusive. Moreover, I am still yet to show how all of this relates to our understanding of moral conflict and disagreement. In the next chapter, I hope to remedy this by articulating the two major competing accounts of moral psychology and offering arguments why the TEA model is more plausible. In the course of doing so, I will illustrate how the TEA model makes sense of moral conflict and moral disagreement, and suggest that it possesses further explanatory power with regard to moral conflict and moral disagreement than these competing accounts.

Chapter 5 – Explaining Moral Conflict and Disagreement

In this chapter, I will set out how the conclusions regarding moral conflict and moral disagreement which I established in Chapters 1, 2 and 3 can be best explained by appealing to my proposed Two Stage Enculturated Affect (TEA) model of moral psychology argued for in Chapter 4. That is, this model provides the most compelling explanation of the phenomenology of moral conflict, as well as the evidence regarding the fundamental nature of intercultural and intracultural moral disagreement, when compared with competing accounts of moral psychology. Establishing this will allow us to get a better understanding of the underlying nature of moral conflict and disagreement. Moreover, insofar as it is the case, it reveals an additional layer of explanatory power for the TEA model and thus lends it some extra credence in its own right.

In section 1 I explicate two rival accounts of moral psychology, Jesse Prinz's Emotional Constructivism and John Mikhail's Universal Moral Grammar, which offer an understanding of the origins of moral norms and values which differ from the TEA model. In my articulation I offer reasons for preferring the TEA model over these alternative models of moral psychology independent of the facts concerning moral conflict and moral disagreement. Section 2 goes on to compare how the accounts might explain the phenomenology of moral conflict. I argue that the TEA model can offer a neat explanation of this phenomenon by appealing to the distinctness of the affective dispositions which underlie our moral values, whilst Universal Moral Grammar cannot. However, Emotional Constructionism could account for the phenomenology of moral conflict in a similar manner to the TEA model. In section 3 I will go on to explain how the various accounts might explain the pattern of intercultural similarity and variation in moral values and norms. I suggest that whilst both alternative accounts have explanatory resources to bring to bear here, the TEA model offers a better explanation than either. Finally, section 4 will evaluate how each account could potentially explain intracultural moral disagreement. On the TEA model genetic differences and differences in upbringing both contribute to intracultural differences in emotional dispositions, which help explain intracultural moral disagreement. On the other hand, moral grammarians have no plausible explanation to hand, whilst Emotional

constructivists must explain intercultural moral disagreement purely in terms of differences in upbringing, which does not cohere with evidence from behavioural genetics.

I will conclude by clarifying how the explanation of moral conflict and disagreement drawn from the TEA model help to deepen our understanding of the phenomena. My next chapter will go on to highlight how this improved understanding can have indirect normative implications when considered in relation to political liberalism.

1. Prinz's Emotional Constructivism

To recap, on the TEA model of moral psychology which I endorse, our affective dispositions play the role of the raw material of moral judgement. Such dispositions heavily inform our evaluations of particular stimuli as good or bad. Moreover, they play a crucial role in the cultural construction of moral concepts (values and virtues) and norms; such cultural constructions originate from and are taken as morally salient by individuals insofar as they appeal to our affective dispositions. Although such dispositions are to some extent malleable in the face of cultural influences, our innate learning biases tend to shape them in certain directions, such that we will more readily experience certain emotional responses when encountering certain stimuli than others as a consequence of their evolutionary adaptive value. Those affective dispositions which we are innately prepared to develop and which lend themselves to the cultural development of moral concepts have been clustered by Haidt into five foundations of moral judgement; harm/care, reciprocity/fairness, group/loyalty, authority/respect and purity/divinity. These foundations lead cultural groups to construct values based around these concerns, which further influences the cultural fitness of norms regulating behaviour relevant to such concerns.

Thus, the TEA model of moral psychology explains moral values and norms as stemming from an interaction between innate psychological factors and cultural forces. Almost all contemporary accounts of moral psychology would agree with this to some extent. Where the differences lie is in the relative importance that each account attributes to innate and cultural influences in the development of our moral values and norms. To take one extreme, one could take the position that humans are passive receptacles who simply internalise the moral concepts of their surrounding culture uncritically, either during a critical socialisation period or throughout the entirety of their lives. This is not to necessarily rule out the role of an innate psychological capacity in the internalisation process. For instance, Chandra Sripada

and Stephen Stich suggest that we need to posit an innate psychological mechanism in order to explain and shed light on the universal tendency of humans to culturally acquire and be intrinsically motivated by norms. In their paper *A Framework for the Psychology of Norms,* they do just this, and whilst they emphasise that there are many open questions as to what kind of mechanism is responsible for norm acquisition, they shed some light on what they think the basic framework involves.[160] In the sketch that they set out, they point out that such a mechanism must typically emerge early in psychological development, be activated automatically, and respond to proximal cues in the environment. Such cues can take the form of both inferences from the behaviour of others within the social environment and explicit verbal instruction, which allow us to identify and internalise the norms of our particular social group.

This model suggests that the ability to acquire norms at the very least stems from an innate capacity. However, it says nothing about the extent to which other innate factors might influence *which* norms are acquired. Stich and Sripada remain neutral on this issue. Yet they do consider what they call the 'Pac Man thesis' (a reference to the old arcade game character which consumes all that it touches) as a null hypothesis. According to the Pac Man thesis, the norm acquisition mechanism exhibits no constraints or biases in terms of which norms will be acquired and internalised – it simply detects and absorbs all norms which are present in the individual's social environment in an unselective manner. Therefore, proponents of the Pac Man thesis would have it that norms are to be conceived of as chiefly the product of cultural forces – what set of norms we happen to have is entirely determined by the norms which are present within our social environment.[161] To reiterate, the TEA model agrees here, insofar as it holds that our moral values and norms are culturally constructed. But as I have emphasised, the account of norm acquisition it offers is importantly different, insofar as it conceives of it as subject to biases and constraints in terms of the values and norms which are likely to be internalised and treated as moral rather than conventional.

The best contemporary example of such an anti-nativist conception of our internalisation of moral norms and values is Jesse Prinz's Emotional Constructivism. Prinz most obviously marks himself out as a critic of moral nativism in terms of his contention that morality represents an accidental by-product of general features of human psychology rather than a

[160] Stripada, C.S and Stich, S. 'A Framework for the Psychology of Norms' pp.280-301
[161] Stripada, C.S and Stich, S. 'A Framework for the Psychology of Norms' pp.298-301

functional adaptation.[162] This aspect of his anti-nativism is irrelevant to the question at hand: what is at issue is the extent to which innate psychological dispositions shape the content of our moral concepts and norms rather than whether these dispositions evolved to specifically serve that purpose. Nonetheless, his account of the unconstrained and relatively unbiased means by which we internalise the moral concepts of our social environment also represents the closest approximation to the Pac Man thesis. Like Nichols and Haidt, Prinz heavily emphasises the relation between emotion and moral judgement. In fact, Prinz advocates an extreme form of sentimentalism, holding that emotional responses not only motivate moral judgements, but that certain emotions *constitute* our moral concepts; specifically, the emotions of guilt and blame.[163] He further argues that humans are innately endowed with a set of basic emotions, each of which evolved in order to fulfil the functional role of representing concerns which influence our survival chances. Moreover, these emotional dispositions facilitate the development of moral concepts amongst cultural groups. As he concedes, "Natural selection has probably furnished us with a variety of behavioural and affective dispositions that contribute to the emergence of moral values."[164]

So far so good. Nonetheless, Prinz ultimately holds that those innate dispositions which we possess are not fixed or strong enough to be a deciding factor in shaping the content of our moral norms and concepts, arguing that such dispositions are most significant insofar as they provide us with a bare capacity for morality itself. He accepts that humans are somewhat more likely to socially construct certain moral concepts over others due to dispositions such as harm aversion, referring to such influences as 'biocultural interactions'. Yet he suggests that these are far too weak to form a significant basis of an explanation for the pattern of moral variation and similarity that can be observed across cultural groups.[165] Different cultures might draw upon the same range of innate psychological dispositions to inculcate the moral concepts which they construct, and such dispositions might have some small role in directing the content of such norms and values, but we should not conceive of them in terms of the moral foundations which Haidt speaks of. Rather, Prinz suggests that our affective natures are far more malleable, susceptible as they are to cultural calibration, and do not tend us strongly towards any particular intuitive moral responses over others. In his

[162] See, for example, Prinz, J. 'Is Morality Innate?' in Sinnott-Armstrong, W. (ed.) *Moral Psychology Volume 1* pp.367-406
[163] Prinz, J. 'Can Moral Obligations be Empirically Discovered?' *Midwest Studies in Philosophy* 31 (2007) pp.271-291
[164] Prinz, J. *The Emotional Construction of Morals* Oxford: Oxford University Press (2007) p.255
[165] Ibid, pp.274-287

words, "I propose that biologically based behaviors pertaining to kindness, fairness, and reciprocity are culturally malleable and insufficient to guide our behavior without cultural elaboration."[166] On his view, then, the moral norms and concepts of cultural groups are relatively independent of our innate psychological features

The main issue with Prinz's account is that it does not seem to give sufficient weight to the fact that some emotional associations are far easier to learn, and moralise, than others. I argued in my previous chapter that the elicitors and relative strength of our affective dispositions are somewhat pliable in the face of cultural influences. Nonetheless, innate learning biases relating to emotional development ensure that the extent of their malleability is constrained. To elaborate: I have previously highlighted evidence which suggests that some emotions are basic; that is, cross culturally recognised as having the same facial expression and core stimuli. Moreover, there is evidence that some great apes and other primates, our closest evolutionary ancestors, exhibit homologous bodily changes, facial expressions and typical behaviour when faced with the core stimuli of some of these basic emotions.[167] This indicates that to a large extent our emotional dispositions are, as the TEA model contends, an innate feature of our psychologies, whose presence predates the emergence of *Homo sapiens* and which evolved in order to solve various adaptive challenges.

This much Prinz agrees with. However in addition he holds that our affective dispositions can be moulded in pretty much any direction, giving rise to culturally-specific emotions, through what he refers to as *calibration*.[168] Prinz posits two mechanisms by which this calibration process takes place. Firstly, basic emotions can be combined together to produce an emotional blend – he gives the example of contempt, which he suggests is a blend of anger and disgust. Secondly, basic emotions can be linked to a particular set of eliciting conditions to produce a new emotion, such as pride representing joy which is elicited by one's own success. Recall from the previous chapter that the TEA model also maintains that affective dispositions can be calibrated in a similar manner. Nonetheless, whilst there I suggested that such malleability is constrained by our evolved nature, Prinz does not seem to take this to be the case. Although he holds that the basic emotions function so as to represent adaptive concerns, and that recalibration often works by elaborating on a subset of the core eliciting

[166] Ibid, p.277
[167] See De Waal, F. *Primates and Philosophers: How Morality Evolved* Princeton: Princeton University Press (2009) pp. 21-42
[168] Prinz, J. *The Emotional Construction of Morals,* pp.64-68

stimuli of a particular basic emotion, the process is conceived of as relatively open-ended. Given this understanding of affective dispositions as unconstrained by innate factors, so too are our moral values. Hence his statement that "I tend to think, somewhat cynically, that the range of moral rules is relatively unconstrained…I adamantly believe that we could teach people to value the recreational torture of small babies."[169]

However, although humans in particular have also evolved to be susceptible to environmental influences in the shaping of what elicits emotional responses, given their initial etiological function we should expect that this malleability will not go all the way down. For if a complex psychological characteristic has a long evolutionary heritage and specified function, we should expect that it should be fairly resistant to dramatic modification. If our emotional dispositions were utterly plastic, then their chances of in fact motivating us to behave in an adaptive manner would be entirely hostage to the influences we were exposed to in our surrounding environment.

Of course, we cannot come to firm conclusions regarding just how flexible our emotional repertoire is based purely on the logic of adaptation. As many rightfully stress, evolution is not an optimising nor predictable process, and natural selection is by no means the sole force involved.[170] Nonetheless, there is independent evidence that humans and other social mammals exhibit various learning biases when it comes to developing emotional associations. Preliminary proof for such biased learning was first demonstrated in the 1960s by researchers studying the capacity of rats to learn avoidance responses towards stimuli associated with negative bodily effects, based on environmental cues. Robert Koelling and John Garcia designed and ran an experiment whereby rats were exposed to different combinations of stimuli and negative bodily feedback.[171] They found that whilst rats quickly learned to avoid flavoured water which triggered nausea or emitters of lights and noises which triggered physical pain, they failed to learn to avoid flavoured water which triggered physical pain, or lights and noise which triggered nausea. An obvious explanation as to why the groups of rats under these conditions did or did not develop avoidances in the particular pattern that they did is because their learning mechanisms were already predisposed to associate certain

[169] Prinz, J. 'Reply to Dwyer and Tiberius." in *Moral Psychology Volume 1,* p.429
[170] For instance, see Gould, S.J. and Lewontin, R.C. 'The Spandrels of San Marco and the Panglossian Paradigm' *Proceedings of the Royal Society, (*1979) pp.581-598
[171] Garcia, K., Koelling, R. 'The Relation of Cue to Consequence in Avoidance Learning' *Psychonomic Science, 4.* (1966) pp.123-124

aversive bodily responses to certain types of environmental cues. In their natural environment, rats are more likely to experience nausea after having consumed something of a distinctive taste, given the link between flavour and poisonous food. Conversely, they are more likely to experience physical pain after contact with prominent visual and auditory danger cues. Hence, rats are innately prepared for learning certain types of connections, but not others, based on the adaptive value of making the sorts of connections which are typically found in their natural environment, and hence more likely to be reliable.

With this in mind, it is highly plausible to suppose that the psychological mechanisms which facilitate the calibration of our emotions also involve a strong degree of preparedness which biases us towards forming some associations and aversions much more easily than others. For instance, it is common knowledge that snakes and spiders often provoke intense fear amongst humans, and that phobias of these animals are common. This is despite the fact that few individuals are ever physically harmed by such creatures, at least in contemporary Western societies. Based on this, some have suggested that humans are born with an innate fear of potentially dangerous animals which were common in the environment where human evolution took place. However, recent studies suggest that rather than being born with a fully developed, innate fear of such creatures directly encoded into our psychology, humans instead possess psychological biases which lead us to very easily *learn* to fear them. Such studies demonstrated that both humans and non-human primates are easily inculcated to fear snakes, as they possess certain characteristics which we are innately disposed to associate with danger, whilst we are not similarly inclined to develop a fear of other features of the environment, such as rabbits and flowers.[172] Just as the learning mechanisms of rats are biased towards more easily developing an aversion to a taste which leads to nausea rather than physical pain, so too do humans more easily develop fear towards snakes than other organisms which would have presented less of a threat to our survival chances. Thus, our emotional malleability, such as it is, is in certain respects directed to develop in such a way that typically has adaptive value. This helps explain why although there are some generalisable differences in the emotional tendencies found within distinct cultural groups, basic emotions tend to have cross-culturally universal core stimuli. Whilst we may not be born with, for instance, an innately fixed psychological association between faeces or rotting meat and a

[172] See Öhman, A. and Mineka, S.' Fear, Phobias and Preparedness: Toward an Evolved Module of Fear and Fear Learning.' *Psychological Review,* 108, (2001) pp.483-522

disgust response, humans typically have such an association because it is one which we are so strongly predisposed to learn to make.

This all being the case, it seems that Prinz's claim that our affective dispositions are culturally malleable enough to ensure that our moral concerns are equally flexible is implausible. Again, Prinz does not deny that there is an element of innate preparedness which shapes our emotional dispositions, and as mentioned, he is even willing to grant that our innate affective tendencies facilitate the social construction of values and norms. Yet it is his insistence that these tendencies don't play a significant role in shaping the content of such values and norms is at odds with the above portrayal of our emotional learning biases. If we are strongly disposed to, for instance, learn to experience an aversive affective response when we witness or imagine distress cues, and if values and norms gain their perceived normative force from affective responses, then we will surely in turn be equally strongly disposed to endorse and internalise values and norms which dissuade behaviour which elicits distress cues rather than those which permit or promote them. Similarly, if we are innately prepared to learn to respond positively to an equitable distribution of resources and respect for group loyalty or hierarchy, it will be far easier to learn to accept values and norms which encourage such states of affairs than those which don't. Moreover, that this is in fact the case is evident from everyday experience and the lessons of history. As Haidt points out, just as it is notoriously difficult to foster a preference for vegetables over sweet and fatty foods in children, even the most persistent socialisation attempts to inculcate certain moral values (such as impartial concern for all humankind over prioritising one's family and friends) typically meet with difficulty, as in the case of various communes like the Israeli kibbutzim.[173]

In sum, Prinz's Emotional Constructivism might draw plausibility from the evidence indicating that moral judgement is strongly related to our emotional responses, and that such emotions are somewhat malleable in the face of cultural influences. However, it does not make enough room for the influence of innate emotional learning association biases in shaping the cultural construction of moral concepts. I will now move on to consider another alternative account of the origin of moral concepts, the Universal Moral Grammar account, which I argue makes the opposite mistake insofar as it places too much weight on the import of innate psychological factors.

[173] Haidt, J. and Joseph, C. 'The Moral Mind.' p.377

2. Mikhail's Universal Moral Grammar

On the other end of the moral nativist/anti-nativist spectrum is the position that some moral concepts are almost entirely the product of an innate component of our psychology, without the need for much contribution from experience. A proponent of this position denies that we necessarily require specific cultural input in order to develop moral concepts. It is true that we do not make moral judgements whenever we experience an affective response which causes us to evaluate something as good or bad, and thus there must be more to moral judgement than emotional responses alone. However this is not to say that in judging something as morally right or wrong we can only do so in virtue of its relation to a culturally acquired value or norm. Perhaps some innate principles of human moral reasoning can play the role that the TEA model ascribes to culturally supplied values and norms.

For instance, moral judgements may sometimes result simply from the evaluation of a particular event as good or bad combined with a tacit inference that such an event can be causally attributed to individual or collective human agency. For instance, if I witness Jane step down hard on Edward's foot, and Edward exhibits distress cues as a consequence, I might experience an aversive emotional response which causes me to judge the event as bad. If I infer that the action was intentional and deliberate on the part of Jane, this might be enough for me to go on to judge it as morally wrong, whilst I would refrain from making such a judgement if I were to judge the action as merely accidental, or if it were done to prevent greater badness (such as if it was the only effective means of alerting Edward to an incoming threat).

This moral judgement could stem from the perceived violation of the internalised norm 'Don't deliberately cause harm to others (without good reason)', which gains its normative force from the culturally constructed value of compassion. Nonetheless, an individual who had not internalised such an explicit norm through the process of enculturation might yet judge it as wrong purely on the basis that it constituted harm, and that it was deliberately caused by Jane for no good reason. Such a judgement of wrongness might arise automatically due to an innate principle of reasoning rather than a culturally specific moral concept or norm.

Something akin to this proposition is suggested by Universal Moral Grammar, which represents one of the strongest forms of moral nativism. This approach is favoured by a

range of theorists, such as Susan Dwyer[174] and Marc Hauser[175], but I will concentrate on John Mikhail's presentation, given his status as the leading proponent of this position. Mikhail argues that in morally evaluating behaviour we automatically and unconsciously engage in a form of action analysis, which influences our judgement in relation to the event, and that humans are innately disposed to make moral judgements towards transgressions involving harms, injustice and rights. [176] For Mikhail, this is a consequence of our species typical innate psychological mechanisms, which cause us to come to similar resolutions in certain moral conflicts.

This may sound similar to the line that the TEA model takes, which also explains cross-cultural commonalities in moral norms in terms of widely shared psychological characteristics that indirectly shape moral judgement. Nonetheless, there are two differences on Mikhail's model. Firstly, he denies that affective dispositions are the source of the similarities in moral judgements across cultures, and secondly he holds that our innate psychological characteristics influence our moral judgements more directly than the TEA model suggests. He maintains that such psychological characteristics take the form of implicit principles of reasoning which directly lead us to make certain moral judgements, rather than indirectly shape them through the medium of cultural construction.[177]

As discussed in chapter 2, Mikhail claims to have found a high level of cross-cultural agreement amongst even very young participants in their responses to trolley problem moral conflicts. He combines these findings with a 'poverty of the stimulus' argument in making the case for innate moral principles: children must be endowed with innate moral principles, since they cannot possibly have learned the complex principles that they implicitly employ in resolving these conflicts. This move is explicitly inspired by Noam Chomsky's Universal Grammar hypothesis, wherein the ease with which children acquire language without much instruction is cited as the basis for contending that humans possess an innate psychological faculty which facilitates their grasp of grammar. Similarly, Mikhail suggests that children are

[174] See Dwyer, S. 'How Good is the Linguistic Analogy?' In *The Innate Mind*, *Vol. 2* pp.237-256.
[175] See Hauser, M. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* New York: Ecco Press (2006)
[176] Mikhail, J. 'Universal Moral Grammar: Theory, Evidence and the Future' in *Trends in Cognitive Science,* Vol.11 No.4 (2007) pp.143-152
[177] Interestingly, some evidence suggests that we have conscious access to only certain principles which regulate our moral judgements, whereas others motivate them implicitly. See Cushman, F. Young, L. and Hauser, M. 'The Role of Conscious Reasoning and Intuition in Moral Judgment – Testing Three Principles of Harm', *Psychological Science,* Vol. 17, No. 12 (2006) pp.1082-1089

not exposed to the necessary proximal cues which would allow them to deduce and imitate or internalise the kinds of complex rules and principles which explain the pattern of moral judgements that they consistently and cross-culturally make. Although children are often given explicit or implicit moral instructions, such as "Don't hit" or "Share your toys", Mikhail argues that their solutions to moral conflicts like trolley problems imply a faculty enabling a sophisticated level of implicit moral reasoning which could not possibly be acquired through their limited range of experience. For when confronted with trolley problems, "children must represent and evaluate these novel fact patterns in terms of properties like ends, means, side effects, and *prima facie* wrongs such as battery, even where the stimulus contains no evidence of these properties. These concepts and the principles which underlie them are as far removed from experience as the hierarchical tree structures and recursive rules of linguistic grammars. It is implausible to think they are acquired by means of explicit verbal instruction or examples in the child's environment"[178]

Mikhail thus argues that humans possess an innate, dedicated moral faculty that leads us to make the same sorts of moral judgements from an early age across all cultural groups. He suggests that to the extent that moral concepts are required in order for us to make judgements of right and wrong, they need not be primarily the product of cultural forces. Rather, the most salient principles responsible for our moral reasoning are innately endowed, thus making it possible to develop and internalise moral values or norms without the need for any specific cultural input in shaping them.[179]

However, as noted in chapter 3, the level of agreement amongst participants which Mikhail cites, although significant, is not as impressive as he claims. His data reveals high levels of intracultural moral disagreement, especially regarding what is permissible in non-standard variations of trolley problems. Furthermore, also as previously mentioned, Hauser has

[178] Mikhail, J. "The Poverty of the Moral Stimulus." in *Moral Psychology, Vol.1* p.355

[179] Note that Mikhail need not necessarily hold that *no* cultural input is required in order to internalise moral concepts; his view is compatible with it turning out that just as so called 'feral children' who are deprived of human contact during a critical period of development will not learn language, feral children might similarly fail to develop moral norms. It is only the background principles of moral reasoning which are innate, not full blown norms and values as such. Nonetheless, these principles so strongly dispose us in the direction of certain norms that even a very limited exposure to any human culture will lead us to develop them.

produced evidence that members of certain cultural groups are not typically inclined to make the sorts of moral distinctions which Mikhail suggests we are innately disposed to make.[180]

It has further been argued that the poverty of the stimulus argument that Mikhail utilises is less plausible when applied to the learning of moral rules than it is concerning the learning of grammatical rules, as Mikhail underestimates the level of feedback related to moral rules that children are exposed to. It might be that we cannot plausibly account for the widespread and early internalisation of complex norms which children apparently make judgements in accordance with from an early age, *if* we take explicit verbal instruction to be the sole means by which children infer the local norms of their social environment. Nonetheless, there are alternative means by which children can infer, identify and go on to internalise the norms of their cultural environment which supplement such explicit socialisation processes. For instance, Kim Sterelny points out that children can also make inferences from the rich variety of moral exemplars in the various narrative devices (such as songs, stories and fables) that cultural groups typically expose them to. Further, children can also pick up moral concepts and norms from their interactions with others of their community; interactions which will be particularly emotionally charged and thus memorable when involving morally salient behaviour.[181] So, there may yet be sufficient proximal cues for children to detect and learn complex moral principles through mechanisms such as the emotional feedback they receive from parents and others, without the need for the principles behind them to be necessarily innately encoded. Moreover, remember that the TEA model does not deny that humans are innately prepared to acquire and moralise certain concepts and norms, thus partially explaining the moral precociousness of young children. It merely maintains that this range of concepts and norms is less determinate than Mikhail would have us believe. Thus, although the poverty of the stimulus argument may have some bite against more strongly anti-nativist positions, such as emotional constructivism, it does not necessarily tell against the TEA model.

Another issue with the Universal Moral Grammar account is that, insofar as Mikhail takes the innate factors which guide our moral judgements as principles of reasoning rather than affective in nature, it flies in the face of evidence which implies the causal efficacy of emotion in moral judgement. On Mikhail's model, emotion is merely a by-product of moral judgement; the rational principles of our innate, dedicated moral faculties are doing the causal

[180] Abarbanell, L. and Hauser, M.D. 'Mayan Morality: An Exploration of Permissible Harms' *Cognition,* 115 (2010) pp.207-224

[181] Sterelny, K. 'Moral Nativism: A Sceptical Response' *Mind and Language* 25 (2010) pp.290-293

work.[182] Yet recall from the previous chapter that evidence from various psychological and neuroscientific studies imply a strong causal role for affect in moral judgement. Given this, the Universal Moral Grammar account stands on shaky empirical ground.

One way in which this could be avoided is to modify the theory in such a way as to divorce it from its rationalist implications. One could easily offer a more emotivist interpretation of the moral grammar model by portraying the innate principles of action evaluation which Mikhail posits as necessary, but not sufficient for full-blown moral judgement. On this reinterpretation, we may be born with an innate faculty for analysing actions which influences our emotional responses to actions, but the emotional response itself is still ultimately responsible for delivering the moral judgement. Indeed, in a response to Nichols, Blair conjectures this sort of faculty as constituting an important part of the VIM (Violence Inhibition Mechanism).[183] The extent to which we experience an action or event as affectively charged could be partly dependent on whether we analyse the event as, for instance, intentionally caused and for what reason, and this further influences whether we judge it as wrongful or not. Upon analysing an emotionally salient event as attributable to human agency, we might automatically go on to morally evaluate it. Our tendency to engage in such an analysis and make corresponding judgements based on it could be part of our innate psychological framework, and thus facilitates the formation of moral judgements.

In any case, it must be noted that even if this alternative interpretation of strong moral nativism is taken as a plausible explanation of at least some moral judgements, it is not necessarily incompatible with the TEA model. It may be the case that we are so strongly innately disposed to endorse certain moral principles that they reliably arise independent of the specific content of cultural input. Yet such principles are only minimal in scope, and culturally constructed norms and values remain the primary source of our moral judgements. Mikhail's Universal Moral Grammar model, meanwhile, suggests that our evaluation of actions as right or wrong stems more from innate principles than culturally constructed values, which stands in contrast with the TEA model.

I have now critically discussed two alternative accounts of moral psychology, each differing in the extent to which they attribute the source of moral values and norms to cultural versus innate psychological factors. Prinz's Emotional Constructivism takes moral concepts to be

[182] Mikhail, J. 'Emotion, Neuroscience, and Law: A Comment on Darwin and Greene' in *Emotion Review* Vol. 3 No. 3 (2011) pp.293-295

[183] Blair, J. 'Normative Theory or Theory of Mind? A Response to Nichols' in *Moral Psychology, Volume 2* pp.275-278

almost entirely socially constructed, whereas Mikhail's Universal Moral Grammar conceives them as primarily the product of an innate, dedicated moral faculty. During my discussion I have offered reasons for favouring the TEA model: Emotional Constructivism does not seem to afford enough weight to the influence of innate biases in the learning of emotional associations, whereas the basis of the poverty of the stimulus argument which grounds Universal Moral Grammar is dubious. I now mean to show that the TEA model has an additional advantage over these alternatives, in that it has more explanatory power with regards to the pattern of moral conflict and intercultural and intracultural moral disagreement that I discussed in chapters 1, 2 and 3. Drawing out how exactly this is the case will not only improve the plausibility of the TEA model, but can help us make further sense of the causes of such phenomena.

3. Explaining Intrapersonal Moral Conflict

Recall that in Chapter 1 I suggested that when individuals are faced with situations of moral conflict, they are often able to come to solutions which they consider to be morally best, all things considered. Nonetheless, their resolutions are also accompanied by the experience of 'tragic remorse', a disquieting feeling of moral loss, which we endorse as an appropriate response to the situation. Such tragic remorse is more commonly associated with conflicts between distinct moral concerns rather than between two instantiations of the same value. This, I argued, suggests that individuals are implicitly committed to a range of distinct values, which are to some extent incommensurable: the realisation of one does not make up for the loss of another. I referred to this phenomena as descriptive value pluralism.

On first blush each of the accounts of moral psychology which I have previously considered can explain descriptive value pluralism. For each allows that humans possess multiple, distinct moral concepts and norms. The accounts simply disagree over the relative extent to which these concepts and norms derive from innate psychological characteristics versus cultural construction. This being the case, each can make sense of individuals recognising a range of moral concerns which might sometimes come into conflict.

Nonetheless, I contend that the TEA model can go further than this and offer a more comprehensive explanation of descriptive value pluralism. For this model maintains that our moral concerns derive their perceived normative force from a range of affective bases.

Individuals are apt to internalise moral concepts and feel moral evaluative force because of their relation to phenomenologically distinct emotional responses. For instance, whilst compassion and harm norms are grounded ultimately in our capacity to experience harm aversion, purity-based values and norms regulating sexual conduct are treated as morally significant on the basis of disgust. Assuming that this is the case, we can readily explain the characteristic phenomenology of moral conflict. When moral concerns conflict, we find it so difficult to resolve them in an ultimately satisfactory way because distinct affective responses underpin each concern. Although we often still can come to an all-things-considered best judgement in such cases, we are left with tragic remorse since, whatever we prioritise, we experience the normative force of the neglected value as phenomenologically distinct.

The potential to experience tragic remorse is, to some extent, mitigated by the enculturating process. As Haidt emphasises, cultural groups tend not to draw upon the full range of the five foundations in culturally constructing their moral concepts, in order to minimise these sorts of value conflicts within the group. A culture cannot cultivate and moralise each affective base, he argues, because this "would risk paralysis, as every action triggered multiple conflicting intuitions".[184] Rather, the affective bases which constitute some moral foundations are culturally elaborated upon at the expense of neglecting or even deliberately suppressing the others, in order to cement moral solidarity within the group. Moreover, individuals assign culturally constructed moral concepts and norms different relative weights which enables them to balance their relative importance against each other when they conflict. This ensures that individuals do not experience tragic remorse too often, since some considerations will not be regarded as morally salient, and they seldom encounter moral conflicts which they take to be dilemmas which cannot be resolved. However, when an individual *does* encounter a situation where their internalised values or norms prescribe conflicting courses of action, then the fact that this ultimately involves a conflict between distinct affective sources of normativity lends itself to a consequent experience of tragic remorse. Understanding culturally constructed moral concepts as deriving from a plurality of distinct emotions helps to explain our experience of values as partly incommensurable.

Meanwhile, Mikhail has less resources at his disposal to explain descriptive value pluralism. Universal Moral Grammar says nothing of the innate, underlying principles of morality which it hypothesises as deriving from distinct psychological sources. Indeed, the model holds that

[184] Haidt, J. 'The Emotional Dog and its Rational Tail' p.827

humans are innately endowed with dedicated moral faculty, which has the specific function of providing an integrated body of principles which are designed to work smoothly together. If this was the case, then although we might still expect individuals to encounter conflicts between *prima facie* moral rules, we would not predict that they would experience tragic remorse to the extent that they do. After all, the faculty would require some mechanism to resolve apparent moral tensions arising from situations when principles recommend incompatible judgements, such as principles having set 'lexical order'[185] relative to each other. For the system to then go on to produce an accompanying phenomenology of distinct moral loss after such resolutions would be extraneous – tragic remorse seems not to have any obvious functional application, after all. Of course, an advocate of Universal Moral Grammar might deny that their model of moral psychology need be conceived as so efficiently designed. Perhaps tragic remorse is a necessary side effect of having the capacity for moral judgement, which even a dedicated moral faculty cannot overcome. Yet even so, the phenomena of tragic remorse brings out just how distinct we experience moral values to be from one another. This is less congruent with the notion that moral intuitions are the product of a single organised faculty than the TEA model's contention that they derive from a range of distinct affective bases.

Emotional Constructivism, on the other hand, can explain the phenomenon handily. After all, like the TEA model, it conceives of emotions as the primary source of felt normativity, and Prinz could equally explain the phenomenology of moral conflict as a by-product of distinct moral concerns mapping onto distinct affective bases. Furthermore, despite Prinz rejecting Haidt's Moral Foundations Theory, he does endorse its precursor – the CAD hypothesis mentioned in the previous chapter.[186] Recall that this thesis explicitly holds that different emotions arise from different classes of moral transgressions: transgressions against community trigger contempt, autonomy based transgressions cause anger and transgressions of divinity elicit disgust. Furthermore, Prinz holds that other emotions such as guilt, shame, sympathy, admiration and gratitude are also culturally calibrated in the social construction of morality. This plurality of moral emotions could be said to contribute to the sense of

[185] Meaning that principles are ordered in terms of a set priority, whereby the higher principle in the order is always taken as more important when conflicts occur between them. The phrases comes from Rawls, where he suggests that of the principles of justice which he proposes, the liberty principle has lexical priority over the equality principle. (Rawls, J. *A Theory of Justice*, p.204)

[186] Prinz, J. *The Emotional Construction of Morals,* pp.72-76

fragmentation in our moral phenomenology.[187] Insofar as Emotional Constructivism accepts this, then it is equally well situated to explain descriptive value pluralism as is the TEA model. Whilst Universal Moral Grammar fares poorly in this regard, we will have to look elsewhere to vindicate my contention that the TEA model has greater explanatory power overall than Emotional Constructivism.

4. Explaining Intercultural Moral Disagreement

Let us move on to see how each account fares in explaining intercultural moral disagreement. I will concentrate on the cultural diversity that we observe concerning moral concepts and norms associated with harm. As previously mentioned, the anthropological record reveals that many cultural groups have norms regulating the infliction of violence. Furthermore, as we would expect given the prevalence of these norms, many cultural groups consider the caring for and defence of the weak and vulnerable as morally valuable, and praise those who act in accordance with compassion as virtuous. However, some cultural groups have wider range of harm norms than others, and the extent and range of conditions under which harm is recognised as morally wrong differs considerably from group to group.[188]

The TEA model of moral psychology has great explanatory power in this regard. This account has it that, in the first stage, those affective dispositions which constitute the harm/care foundation are likely, but not guaranteed, to be drawn upon in the cultural construction of moral concepts. Because humans are innately disposed to respond to distress cues with an aversive emotional response, and to moralise this response, cultural groups are liable to construct values and virtues which represent harm as morally wrong. The harm aversion of members of such groups will be intensified, and they will come to treat this affective response as morally salient. Within such groups, norms which regulate the infliction of harm will enjoy increased cultural fitness, and are likely to originate and persist over time.

However, again, the extent to which this happens is partly dependent on the particulars of the cultural environment in which we are raised. If the harm/care foundation is not emphasised by the moral concepts of a particular cultural group, harm might not feature as

[187] Walter Sinnott-Armstrong suggests something similar in his article 'Is Moral Phenomenology Unified?' *Phenomenology and the Cognitive Sciences,* 7 (2008) pp.85-97

[188] See Silverberg, J. Gray, P. *Aggression and Peacefulness in Humans and Other Primates* New York: Oxford University Press (1992) for a review.

much of a consideration within its values and virtues. Not only will the emotional development of the group's members thereby be shaped in such a way that harm will not elicit as strong an affective response as it will amongst members of other groups, but the response itself will not be deemed as morally relevant. This will entail that norms which regulate harm will not be as likely to originate and persist within such a group.

The Universal Moral Grammar account of Mikhail struggles more to explain such cultural variation in harm norms and moral concepts which moralise the prevention of harm. As I mention in section 1, there certainly could be a sense in which such accounts have something to them, insofar as it may well be the case that we are so strongly disposed towards some key moral principles that they do not necessarily need to be specifically culturally supplied for us to make moral judgements based on them. Nonetheless, this can at best be only a partial story regarding why humans make the range of moral judgements that they do. For if our moral psychology is comprised entirely of a dedicated moral reasoning system, complete with a range of innately encoded moral principles, then why do the values and norms of many cultural groups exhibit so much variation?

It has been suggested that nativist positions such as Universal Moral Grammar can, in fact, attribute some of the moral variation that we observe between cultural groups to the input of cultural influences. Some Moral Grammar theorists, such as Gilbert Harman argue that a so-called 'Principles and Parameters' model of moral psychology can plausibly explain the pattern of cultural variation we observe in moral norms and values.[189] Developing the analogy of Chomskian linguistic grammar, Harman argues that we could be born with an innate underlying structure of moral principles which shapes and restricts the more culturally specific moral rules. Such an innate mechanism ensures that some moral principles are genuinely culturally universal, yet leaves room for a restricted range of variability within the bounds of certain parameters. For instance, our innate psychology might contain the underlying principle 'Do not inflict harm upon X', where the parameter X can be culturally specified as narrowly as 'one's immediate family' or as widely as 'any living creature'.

[189] Harman, G. 'Moral Psychology and Linguistics' in ed. Brinkmann, K. *Proceedings of the 20th World Conference of Philosophy: Vol. 1: Ethics* Bowling Green, Ohio: Philosophy Documentation Center (1999) pp.107-115 See also Dwyer, S. 'How Good is the Linguistic Analogy?' pp.237-256

I contend that such an account by itself is less well positioned to fully explicate the pattern of moral diversity that we encounter between cultural groups. For as Stripada explains, the range of variation that exists between the harm norms of different cultural groups is more complex and subtle, going beyond that of differentiation over a few mere parameters.[190] Across cultural groups, there is variation in the moral norms concerning what kind of harms are prohibited, what class of people should be protected from such harms, under what circumstances such otherwise prohibited harms may be justified, and to what extent the level of harm inflicted may be appropriate. For instance, in some cultural groups any kind of infliction of harm to any degree towards any member of an in-group is prohibited, regardless of the circumstances. However, in others, the infliction of certain types of harm (e.g. open hand beating) is regarded as morally justified towards certain individuals (e.g. one's wife), to certain degrees (e.g. without leaving bruises) and under certain circumstances (e.g. if insulted).

One could attempt to explain this in terms of there being a particularly wide range of parameters which are culturally specified, but which all operate within the framework of innately encoded overarching moral principles. Mikhail, for one, would probably not be willing to accept this level of cultural input: if he did, it wouldn't be clear why he was appealing to cross-cultural similarities in the solution to trolley problems in order to substantiate his case. Still, if a Universal Moral Grammar theorist was willing to concede that the bulk of the content our moral concepts and norms are the product of culture rather innate psychological principles, then this explanation of intercultural moral disagreement is viable. I yet maintain that the TEA model which I advocate offers a more plausible interpretation of such variation in harm norms. Nonetheless, we cannot discount Universal Moral Grammar purely on the basis that it fails to offer a strong account of intercultural moral disagreement.

At the other end of the spectrum, I similarly contend that Emotional Constructivism may offer an explanation of the pattern of cross-cultural moral variation that we observe, but that it is less plausible. Prinz emphasises the role of cultural construction in terms of the development and inculcation of moral concepts and norms, claiming that our shared innate psychological tendencies facilitate such social constructions, but only influence their content in a very minimal sense. Of course, this entails that Prinz can very easily account for cultural variation in moral concepts and norms, but he is less well equipped to explain cross-cultural similarities in these areas. He suggests that, to the extent that cross-cultural similarities in

[190] Stripada, C. 'Nativism and Moral Psychology' pp.329-330

moral concepts exist, they are primarily to be explained through reference to the co-ordination problems that human groups all face. These are difficulties which all human groups share, leading them to construct similar moral concepts and norms on pragmatic grounds. In his words:

> "There are some social pressures that all human beings face. In living together, we need to devise rules of conduct, and we need to transmit those rules in ways which are readily internalised…Cultures need to make sure that people feel badly about harming members of their in-group and taking possessions from their neighbours….The rules are as varied as the problems, but the universal need to achieve social stability guarantees that *some* system of moral rules will be devised."[191]

Thus, Prinz would explain the preponderance of norms, values and virtues associated with the infliction of harm across cultural groups by reference to the importance of preventing violence and aggression in order to maintain social stability and cohesion within human groups. Although he accepts that such social solutions are readily facilitated by the innate disposition to experience vicarious distress, this disposition is conceived of as merely a convenient tool which is often exploited by cultural groups, rather than influencing their moral codes.[192]

Certainly, social problems and the necessity of solving them is bound to have had some importance influence in the development of the moral norms of cultural groups. Yet I take it to be a stretch to suggest that such problems are the sole factor which we should refer to when trying to explain the range of cross-culturally common themes in morality. If the moral concepts and norms that a cultural group developed were guided and constrained only by the needs of social practicality, we should expect a less consistent pattern in terms of what sort of issues are moralised across cultural groups. After all, cultures will generally instil a wide range of norms which act so as to maintain social stability amongst their members. Yet recall from the previous chapter that only those norms which regulate certain types of behaviour are typically treated as moral rather than conventional, and this has be proven to be the case across a wide range of cultural groups. It would be highly coincidental if it just so happened that each of those groups where the moral/conventional task was tested assigned exactly the same sorts of norms moral rather than conventional status, if social necessity was indeed the

[191] Prinz, J. 'Is Morality Innate?' p.405
[192] Ibid, p.404

only influencing factor in their construction. However, it must be conceded that Prinz' explanation of intercultural moral similarity is plausible enough that we cannot discount Emotional Constructivism yet.

In sum, whilst there seems to be too much moral variation across cultural groups for Universal Moral Grammar to easily explain, what variation we do observe is patterned in such a way that Emotional Constructivism equally finds it difficult to account for. Nonetheless, both theories do have some explanatory resources at hand that they can apply here. I take it that my proposed TEA model has a preferable explanation for intercultural moral disagreement, but I cannot claim a knock-down victory against either alternative on this score.

5. Explaining Intracultural Moral Disagreement

How, then, does the TEA model explain the diversity in moral judgements between members of the same cultural group, which I emphasised in chapter 3? The most obvious means by which it can explain it is through differences in the particular manner in which individuals have been socialised. Especially within broad societal cultural group, it is clear that the social environments in which individuals are raised are not uniform. As I argued in chapter 3, when we take broadly construed cultural groups such as 'North American', we typically find much moral diversity within them, especially within such contemporary pluralistic societies. I further noted that in contrast we tend to find less widespread disagreement concerning value weighting within relatively narrowly defined sub cultural groups, such as 'Anglo-Saxon, working class, protestant Virginians'. There I gave reasons to be sceptical that such a pattern could be explained purely in terms of environmental influences upon the members' moral development, but it cannot be denied that such influences do play a large role in explaining it.

The TEA model which I propose can readily accept this. After all, it hypothesises a strong degree of environmental influence in the shaping of the moral outlook of individuals. For one, the account of the development of one's affective dispositions I offered suggests that they are malleable to some degree. The extent to which we will experience an affective response towards a certain stimuli, and whether we take it to be morally significant, is in part dependent upon the sort of socio-environmental influences that we have been exposed to. Thus, for example, the strength of the aversive affective response which we are disposed to

experience when witnessing or imagining the distress cues of others will depend upon whether or not the social environment in which we were raised encouraged the cultivation and moralisation of such a response. More importantly, the particular cultural environment in which we are raised will heavily determine our moral concepts and norms. The TEA model admits that the social construction and cultural evolution of such moral concepts are to a large extent influenced by our affective dispositions. But given that such affective responses are somewhat socially malleable in the first place, and that in any case the particular content of the norms, values and virtues are underdetermined by such dispositions, they will differ from sub-cultural group to sub-cultural group. It makes perfect sense that the more similar the social environment different individuals are exposed to, the more likely they are to share the same values and norms, and regard them as having a similar degree of salience when informing their moral judgements. Thus, intracultural disagreement will of course be less significant when we are considering more narrowly defined cultural groups wherein the members are exposed to more similar social influences.

Nonetheless, as I further noted in Chapter 3, there is reason for us to believe that even amongst individuals exposed to very similar environmental conditions, moral disagreement can and does still exist. As I argued there, the fact that we observe less moral disagreement within narrowly defined cultural groups can partly be attributed to individuals joining those cultural groups whose moral values they identify with, and thus their relative moral consensus does not necessarily stem from shared environmental influences. Moreover, even if it is often latent and stifled, particularly within social groups that demand strong conformity with the moral status quo, individuals of the same narrowly defined cultural group sometimes do hold conflicting views concerning the relative importance of their culturally shared values and norms.

I want to suggest that, to a large extent, this moral disagreement is to be explained in terms of individuals within cultural groups possessing different affective dispositions from one another. Recall that the TEA model holds that our affective responses on their own do not constitute our moral outlook. Our internalisation of culturally constructed concepts such as norms and values are a necessary and important component in facilitating and shaping our individual moral judgements. Nonetheless, this is not to say that the particular affective dispositions of individuals themselves do not go a long way in influencing the extent to which we regard the culturally constructed moral concepts and norms and as morally salient. As

Blair's studies on psychopath's responses to the moral/conventional task imply, even if one is competent with the prevailing norms and values of one's social environment, if one lacks the affective response which such cultural constructions normally get their perceived normative force from, one will not regard them as morally significant. Further, the evidence from psychological and neuroscientific studies which I cited in chapter 4 buttresses the case for affective responses having a direct role in moral judgement. Thus, the TEA model suggests that the extent to which an individual experiences an affective response towards particular stimuli does in fact influence their moral judgement towards it. When individuals who have been exposed to the same set of culturally constructed moral concepts and norms morally disagree, then, the most plausible explanation is that the disagreement stems from differences between the disputants' affective dispositions.

To elaborate, given that culturally constructed moral concepts and norms gain their perceived normative force from affective dispositions, the extent to which we take a concept or norm to be morally salient is dependent on our particular set of affective dispositions. That is, the particular moral weight which we assign to a moral concept, such as a value (and the norms underpinned by that the value) will depend upon how strongly disposed we are to experience an affective response towards to the stimuli which the value relates to. Within any particular culture, there will be a plurality of values which draw their normative force from distinct affective bases. When conflicts occur between these values, one resolves them by weighting their relative importance against each other. This process, I hold, is heavily influenced by one's particular set of affective dispositions. Thus, in a conflict between compassion and equality, one who has a stronger disposition towards harm aversion is more likely to weight compassion over equality, whereas one who has a stronger inequity aversion is more likely to weight equality over compassion. Moreover, even in cases where we lack socially shared moral concepts and norms which are elaborated atop particular affective bases, affective responses on their own can nonetheless influence our moral judgements. For instance, within a cultural context which does not cultivate moral concepts explicitly drawn from the purity domain, individuals with a strong disgust response are still liable to make moral judgements which are influenced by this affective disposition. This is borne out in Inbar's study described in the previous chapter, which demonstrates that disgust-sensitive individuals are more likely to implicitly condemn gay kissing in public despite not explicitly condemning homosexuality.

So, why would individuals of the same cultural group differ in their affective dispositions? The most obvious answer is that they have been exposed to slightly different environmental influences. Even amongst individuals who share an extremely similar cultural environment,

this is not to say that they will share uniform experiences of the kind which have the capacity to mould their emotional repertoire. For every individual's life experience is unique regardless of the social environment in which they were raised, and our unique experiences will inevitably go some way in shaping our particular affective dispositions and, in turn, moral values. Thus although individuals of cultural groups might all be exposed to the same moral concepts and norms, there will be variation in the extent to which individuals will come to endorse each concept and norm, depending on the idiosyncratic environmental influences which shape their affective dispositions. Through their particular experience of the world, the affective dispositions of an individual will be shaped in a manner which will influence the normative weight they implicitly assign to the multiple socially constructed moral concepts and norms that they are exposed to.

However, I also contend that there is another important factor at play: the innate psychological variability of individuals. Although the TEA model which I propose stresses the extent to which our affective dispositions are shaped by socialisation, it also acknowledges that the psychological mechanisms responsible for affective tendencies – and the constraints and biases which influence the way in which environmental factors shape them - are ultimately a product of our genetic heritage. Therefore, our particular genetic makeup will to a large extent direct the course of the development of our affective dispositions.

Now, although it is true that many of the genes which encode for our affective dispositions will be species typical, this is not to say that they are uniform across all individuals within the human species. It may very well be the case that some are, for instance, genetically predisposed to develop stronger dispositions to experience disgust or anger than others. As many philosophers of biology maintain, an important feature of evolutionary theory is the idea that evolutionary change via natural selection is driven by what is known as 'genomic plasticity'. Genomic plasticity entails that individual organisms of the same species will inevitably exhibit variation across all traits, partly due to variation in their genetic makeup. This explains how the evolutionary process gets off the ground – Darwinian evolution depends upon there being inheritable trait variation within species for natural selection to select for. As Samir Okasha puts it:

> "...Darwinism leads us to expect variation with respect to all organismic traits, morphological, physiological, behavioural and genetic. For genetically based phenotypic variation is essential to the

> operation of natural selection. If selection is to cause a species to evolve adaptations, and eventually to evolve into different species, as Darwinian theory asserts, then there must be variation within the species for selection to operate on. Intra-specific variation with respect to all organismic traits, and thus the lack of species specific essences, is fundamental to the Darwinian explanation of organic diversity."[193]

Given this, we have reason to believe that some of the intercultural variation between the affective dispositions of individuals is partly down to genetic variation between individuals.

This notion gains further plausibility when we consider evidence from the interdisciplinary field of behavioural genetics, which studies the role of genetics in behaviour. In trying to understand the heritability of human behavioural and psychological traits in humans, behavioural geneticists generally make use of studies which compare the relative similarities between adoptive siblings, fraternal siblings, twins and identical twins, both raised in the same family environment and apart. These studies are useful insofar as they help disentangle the relative importance of environmental and genetic influences on traits such as personality variables. At one extreme, adoptive siblings share a home environment but are no more likely to share the same genes than at random. On the other end of the scale, identical twins separated at birth and raised in different adoptive homes have very similar genetic makeups, but are subject to a very different range of environmental influences. Meanwhile, other pairs such as fraternal siblings and non-identical twins raised both apart and together are subject to differing ranges of genetic and family environmental influences. By studying the extent to which such pairs possess similar behavioural dispositions, behavioural geneticists can help understand the extent to which the psychological mechanisms responsible for shaping an individual's behaviour are inheritable.

The conclusions drawn from the application of such a methodology are striking. Eric Turkheimer's article 'The Three Laws of Behavioural Genetics and What They Mean' begins by starkly stating that "The nature-nurture debate is over. The bottom line is that everything is heritable…"[194] The three laws that he cites as having been unanimously proven by the aforementioned studies are as follows:

[193] Okasha, S. 'Darwinian Metaphysics: Species and the Question of Essentialism' *Synthese* (2002) p.197
[194] Turkheimer, E. 'The Three Laws of Behavioural Genetics and What They Mean' *Psychological Science* vol. 9 no. 5 (2000) p.160

First Law. All human behavioural traits are heritable.

Second Law. The effect of being raised in the same family is smaller than the effect of genes.

Third Law. A substantial portion of the variation in complex human behavioural traits is not accounted for by the effects of genes or families.

Turkheimer goes on to argue that it would be overly simplistic to interpret these laws as entailing that environmental influences are unimportant in shaping human psychology and behaviour. On the contrary, although the evidence indicates that behavioural traits are genetically heritable, it does not indicate that they are perfectly so – roughly speaking, it is suggested that our genetic makeup accounts for about 50% of the overall influence upon our personality. Although our family environment does not have as much impact on our psychology as we might expect, the interaction between our genes and other environmental factors such as our wider cultural group, particular peer group and the non-shared influences which I mentioned earlier clearly play a substantial role. Regardless, the take home message is that the psychological mechanisms behind behavioural traits are inheritable. From this we can infer that the set of affective dispositions which an individual will ultimately develop in part depends upon their particular genetic endowment, which varies within human populations. In turn, this can help explain why individuals within the same cultural group are disposed to assign differing moral weight to conflicting moral concerns: the affective dispositions which shape their basic evaluative tendencies differ as a consequence of genetic variability.

One might object that this claim is mere conjecture, but in fact recent studies have provided more concrete evidence that our particular set of moral values are influenced by our genetic makeup. Specifically, certain twin studies have found that political orientation is remarkably inheritable. For instance, in a 2013 study involving 600 pairs of twins found that political ideology, as measured by two different indexes, was strongly inheritable.[195] This is not, of course, because genes can encode for the development of specific political ideologies. Rather, it is far more plausibly the case that more general psychological tendencies which influence an individual's political views, such as our affective dispositions, are partly determined by our

[195] Funk, C. et al. 'Genetic and Environmental Transmission of Political Orientations' *Political Psychology*, vol.34, no.6 (2013) pp.805-819

genetic endowment. Given the obvious link between one's particular weighting of moral values and one's political orientation, this represents strong evidence that intracultural moral disagreement is partly a consequence of genetic variation.

Thus, the TEA mode can explain intracultural moral disagreement as ultimately stemming from individuals of the same cultural group possessing different affective dispositions. This in turn is conceived of as a consequence of individual differences in both genetic endowment and non-shared environmental influences.

Emotional Constructivism, on the other hand, can only partially endorse this explanation. Although Prinz can still happily point to the impact of non-shared environmental influences in explaining intracultural moral disagreement, he would be less comfortable with an explanation relying on innate psychological variation. For whilst he conceives of the emotional building blocks of moral judgement as deriving from innate features of our psychology, ultimately he holds that the development of our affective dispositions is determined by our experience rather than our genes. The issue with this is that there just seems to be too much intracultural moral disagreement to explain merely through non-shared environmental influences. If, as Prinz suggests, our affective dispositions are so subject to cultural conditioning, then any chance events which happened to influence us away from the dominant values of our particular cultural group would surely be outweighed by persistent counter-veiling influences.

Another problem with going this route is that it struggles to account for the evidence from behavioural genetics. If our particular set of emotional dispositions are determined by our environment rather than innate factors, then the aforementioned twin studies would surely not provide evidence of their heritability. Prinz himself is sceptical of such evidence, claiming that insofar as twin studies typically recruit a relatively small sample of participants from a very limited range of cultural groups, no strong conclusions concerning the relative impact of environment over genetics can be drawn from them.[196] However this critique does not quite hit the mark – the participants of such studies might not have been representative of all the culturally diversity which exists, but major differences over certain characteristics were represented, political orientation being a good example. The studies consistency provide

[196] Prinz, J. *Beyond Human Nature: How Culture and Experience Shape our Lives* London: Allen Lane (2012) pp.42-53

evidence such characteristics are to a large extent shared between those of a similar genetic endowment. The precise details of how genes and environment interact for this to be the case may not be clear, but it seems intransigent to simply deny the influence of genes in the face of such evidence.

Universal Moral Grammar, meanwhile, might attempt to explain intracultural moral disagreement in a similar manner to the TEA model, relying on a combination of environmental and innate differences. However, neither factor would be as applicable here. Firstly, assuming that a proponent of the theory would be willing to adopt Harman's aforementioned principles and parameters version of the model, they too might appeal to our experiences as differentially determining what norms we internalise. However this explanation fares less well for intracultural than intercultural moral disagreement – typically, children of the same cultural group are exposed to the same cultural norms. And whilst they may be exposed to idiosyncratic environmental influences which might endow them with different affective dispositions, this is less obviously the case when it comes to variations on norms.

One might also invoke innate psychological variability between individuals as an alternative or complementary explanation of intracultural moral disagreement. But again, this suggestion is far less plausible when applied to the dedicated moral faculty which moral grammar theorists propose. For although emotional dispositions are the sort of thing which can differ on a quantitative level, thus explaining quantitative differences in the moral weight that individuals attribute to values, the set of innate principles proposed by Universal Moral Grammar are seemingly more binary. Even if we assume that genetic variation can induce different sets of inherent principles in individuals, or different relative lexical priorities, the intracultural moral variation that we observe seems to go beyond differences in such principles.

In sum, whilst the TEA model is positioned to offer a rich and compelling explanation of intracultural moral disagreement which is coherent with the evidence from behavioural genetics, Emotional Constructivism and Universal Moral Grammar again face difficulty in this regard. This is the third phenomena for which a more plausible explanation has been borne out of the TEA model than these alternative accounts, and I have also shown it to be more coherent with independent evidence. Therefore, we are in a good position to accept

both the TEA model in general and the account of moral conflict and disagreement which it offers.

Conclusion

In this chapter I have contended that, as well as being supported by an impressive range of independent empirical evidence, the TEA model of moral psychology best explains the phenomenology of intrapersonal moral conflict and the pattern of intercultural and intracultural moral similarity and diversity which I have highlighted in my previous chapters. Mikhail's Universal Moral Grammar, meanwhile, does not provide a plausible explanation of descriptive value pluralism nor the extent of the moral disagreement between and within cultural groups. Prinz's Emotional Constructivism, on the other hand, may be able to explain descriptive value pluralism equally as well as the TEA model. However, it faces difficulty explaining both the thematic similarity in what sort of considerations are treated as morally salient across cultural groups, the extent of intracultural moral disagreement and its apparent relation to genetic factors.

Thus far in my thesis, although I have been considering questions related to the phenomenology, anthropology and psychology of moral judgement, my aims have been purely descriptive. In the next chapter, I will attempt to show that the conclusions which I have come may nonetheless have indirect normative implications. Specifically, I will be exploring what my account of intrapersonal moral conflict, moral disagreement between and within cultures and the empirical explanation of these phenomena may imply for the Rawlsian project of political liberalism.

Chapter 6: Implications for Political Liberalism

Thus far in my thesis I have offered an account of moral conflict and disagreement which entails that it is both fundamental and inevitable, even within relatively small scale societies, partly as a consequence of individuals possessing different affective dispositions. But if we have good grounds for taking moral disagreement to be a permanent feature of human life then how are we ever to reach a consensus on substantive political principles? Political realists have suggested that we cannot, and that we should instead give up any ambition of a society united in its commitment to a particular conception of justice. Instead we should appeal to everyone's common interests of maintaining a peaceful and stable society in order to develop and maintain a balance of power, a *modus vivendi,* between fundamentally divided citizens.

However in *Political Liberalism* John Rawls is more optimistic about the prospect of a society regulated by a conception of justice which everyone can agree upon. He rejects *modus vivendi* political settlements whilst accepting fundamental moral disagreement as an aspect of 'reasonable pluralism'. This is the notion that reasonable people, i.e. those who are willing to cooperate with others on terms that are mutually agreeable, will forever be committed to a plurality of conflicting worldviews. Reasonable pluralism is a consequence of the 'burdens of judgement': aspects of human reasoning which preclude substantive agreement on what makes for a good life between all reasonable people. For Rawls, recognition of these burdens plays an important role in convincing citizens that any political arrangement justified purely in light of one's own comprehensive doctrine would be illegitimate, and so we must restrict ourselves to drawing on shared political values when proposing terms of cooperation. In this way we can reach an overlapping consensus on a liberal political conception of justice, despite fundamentally disagreeing about matters of value.

In this chapter I will argue that Rawls is right that recognition of the burdens of judgement promotes toleration, and is a vital criterion of reasonableness. However, the burdens of judgement are a controversial supposition regarding the nature of reasoning which Rawls fails to properly substantiate. Given this, it is important to make a stronger case than Rawls provides for the proposition that those whom one morally disagrees with are not necessarily exhibiting irrationality, wilful ignorance or selfishness in their disagreement. To this end I suggest that Rawls's account of the burdens of judgement can be reinforced and rendered

more persuasive through reference to the TEA model and its account of moral disagreement, which I have argued for in previous chapters.

I furthermore discuss the claim that the TEA moral psychology also entails that a minority of citizens will *always* be unreasonable, in that they will prioritise values incompatible with political liberalism. Conceding this, I yet suggest that recognition of the burdens of judgement also provides such people with pragmatic grounds for refraining from attempting to impose these values on the political level. As a result Rawlsians might need to accept that some citizens will only ever be committed to a liberal political conception for pragmatic rather than internal moral reasons. Given the importance of the pragmatic considerations for political theory, this is a concession which they should be willing to make.

My argument will take the following structure. Section 1 will review my previous conclusions concerning moral disagreement and how this poses a problem for liberalism, before critically discussing the political realist's proposed solution. Section 2 outlines Rawls's basic project, whilst section 3 explicates the role of reasonable pluralism and the burdens of judgement in *Political Liberalism* in more detail. In section 4 I defend the claim that, *contra* Leif Wenar, recognition of reasonable pluralism is indeed a necessary criterion for reasonableness. Section 5 will demonstrate how the TEA model of moral psychology buttresses Rawls's case for the burdens of judgement and consequent reasonable pluralism. Section 6 highlights the problem that this same account could also imply that some individuals will always be unreasonable on Rawls terms. Yet, in section 7, I respond by arguing that if we can convince such individuals of my reinforced account of the burdens of judgement, this furnishes them with the sort of pragmatic grounds for accepting the liberal political conception of justice which political realists highlight. I thus conclude that the TEA model of moral psychology has significant implications for Rawls's project of political liberalism.

1. Political Realism and the Problem of Disagreement

Let me begin by very briefly reviewing the conclusions that I have established in my previous chapters.

- Humans are disposed to invest weight in a multitude of values which are taken to have distinct normative force. Individuals can typically come to a decision about how best to reconcile conflicts between distinct values, but suffer tragic remorse upon doing so. (Chapter 1)

- Members of different cultural groups often fundamentally morally disagree; this disagreement is best explained in terms of them weighting distinct values differently. (Chapter 2)
- There is good reason to believe that such fundamental moral disagreement also occurs within even relatively homogeneous cultural groups, albeit to a lesser extent. (Chapter 3)
- Moral values are culturally constructed, but gain their perceived normative force from a range of distinct affective bases. (Chapter 4)
- Our affective dispositions are variable across individuals as a consequence of non-shared environmental influences, as well as genetic factors. This explains the phenomenology of moral conflict and the incidence of intercultural and intracultural disagreement. (Chapter 5)

For now, the most important lesson to draw is that fundamental moral disagreement is inevitable. Even if agents are ideally situated with access to all the relevant non-moral facts, and even if they exemplify impartiality and perfect instrumental rationality, disagreement about what ends are worth pursuing can, and often will, persist given the way moral psychology operates. This is a startling conclusion, and has significant implications for political philosophy. In particular, recall from my introduction that the inevitability of fundamental moral disagreement could be conceived of as a challenge to the very foundations of liberalism. This is because the dominant strand of liberalism holds that we can at least hypothetically all agree and consent to a set of laws which the state may legitimately exercise coercion in enforcing.

What, then, can the liberal theorist say in response to this? One strategy might be to weaken the consent-based account of legitimacy and move in the direction endorsed by political realists. In recent years political realism has developed as an alternative to mainstream liberal thought in political philosophy, drawing influence from the works of historical figures such as Machiavelli, Hobbes, Nietzsche and Schmitt. These theorists typically highlight the central role of conflict in politics, and take disagreement amongst citizens very seriously. Contemporary political realists follow in this tradition. This school of thought does not necessarily rely on the notion that such disagreement is moral and fundamental in the sense that I do. Some, for example Max Weber, do believe in fundamental moral disagreement,

although he claims that it stems ultimately from the ontological truth of value pluralism.[197] However others, such as Schmitt, take the Hobbesian line of attributing it to the influence of an irredeemably selfish, prejudicial and dominating human nature.[198] Nonetheless, realists do take such disagreement to be unavoidable, and stress that a deep flaw of liberal theory is that it fails to meaningfully address this inherent aspect of the political situation. As Matt Sleat puts it in his recent book on political realism:

> "The persistence of disagreement is one of the fundamental and 'stubborn facts' of political life which ensures that there is rarely any natural harmony or order in human affairs. The most basic political question, what I shall call '*the* political question', is how we are to live together in the face of such deep and persistent disagreement.'[199]

For the realist, then, liberalism's hope of reaching a universal consensus on political principles that we can all consent to is hopelessly utopian. We can never attain what Bernard Williams described as the "insatiable ideal of many a political theoretician: universal consent."[200] And even if consent were possible from an abstract hypothetical standpoint, this does not help us identify a solution as to what to do in the here and now, where disagreement is rife and shows no signs of abating.

This train of thought leads realists to emphasise the importance of negotiation and compromise in reaching a *modus vivendi;* a political settlement the content of which is determined by the particular balance of power within a society, and which all citizens can accept in the name of attaining the universal goods of peace and stability. This *acceptance* on the part of citizens is not as demanding a requirement as the consent which liberalism typically strives for. For although realists insist that a political authority must have some form of justification ready to offer each citizen in order to satisfy what Williams calls the 'basic legitimation demand'[201], this justification needn't be accepted by all citizens on the grounds of their principled agreement with the moral content of the settlement. Moreover, what determines the content of the political settlement is not based on a conception of what citizens would hypothetically consent to under imagined conditions, but the particular

---

[197] Weber, M. 'Between Two Laws' in his *Political Writings,* eds. P. Lassman and R. Speirs, Cambridge: Cambridge University Press (1999) pp.78-79 Recall that I am not committed to such an ontological account of value pluralism, only a psychological descriptive account.

[198] McCormick, J.P. *Carl Schmitt's Critique of Liberalism,* Cambridge: Cambridge University Press (1999) ch.6

[199] Sleat M. *Liberal Realism: A Realist Theory of Liberal Politics*, Manchester: Manchester University Press (2013) p.47

[200] Williams, B. In *The Beginning Was The Deed: Realism and Moralism in Political Argumen*t, Princeton: Princeton University Press (2005) p.6

[201] Ibid, p.4

preferences and interests of those who currently inhabit the society which the political settlement governs. Whilst many individuals who live under a *modus vivendi* will not be fully satisfied with the arrangement, the hope is that a large enough proportion of citizens will be placated by the compromise and understand that it is necessary in order to avoid chaos. In this way, at least some degree of stability and legitimacy can be maintained despite the persistence of disagreement.

Nonetheless many liberals would not take this response to be satisfactory. For one thing, many would not be comfortable with abandoning the liberal ideal of garnering universal consent amongst citizens. Realists are quick to point out that it would be misguided to caricature their position as boiling down to 'might makes right'. Although they might stress the salience of power dynamics in the shaping of a political settlement, they equally hold that the interests of all those governed must be taken into account and a justificatory story must be offered in order for legitimacy to be retained. Yet this account of what makes a political authority legitimate is much less demanding than that which the liberal endorses.

Another bone of contention for liberals might be that the content of a particular *modus vivendi* is not guaranteed to be of a liberal nature. Some realists do indeed stress that liberalism and realism go hand in hand, such as Michael Williams, who states that "Realism is not opposed to liberalism: it is a form of liberalism."[202] Many others suggest that given our current socio-historical situation, *modus vivendi* political settlements must have liberal content in the here and now. For instance, Bernard Williams holds that under conditions of modernity, liberalism may represent the only viable answer to the basic legitimation demand, formulating the equation 'Legitimacy + Modernity = Liberalism'.[203] Nonetheless, these proposed foundations for liberalism might be seen as worryingly contingent on the majority of citizens within a given society sharing liberal values. If the balance of power within a formerly liberal society were to change as a result of immigration or the influence of a charismatic orator who promoted illiberal values, then it would seem that the logic of *modus vivendi* would demand that the settlement compromised by shifting in an illiberal direction.

Nonetheless, if disagreement is indeed as inevitable and fundamental as my TEA model suggests, then it might be utopian to hope for anything more than a localised, contingent justification of a moderate liberal hegemony. Actual consent amongst citizens is certainly not forthcoming, and accounts of hypothetical consent fail to recognise that disagreement is

[202] Williams, M.C. *The Realist Tradition,* Cambridge: Cambridge University Press (2005) p.10
[203] Williams, B. *In The Beginning Was The Deed,* p.9

genuinely fundamental: it is not attributable to factors we can simply abstract away whilst retaining any of the voluntariness which is supposed to ground consent as normatively significant. As realists suggest, and the TEA account vindicates, it is apparently impossible to ever attain a society built on the ideal of universal consensus. With this in mind, should liberals drop their previous ambitions as over-optimistic and adopt more realistic normative aims for political theory, or would this be premature?

2. Rawlsian Political Liberalism

In this section I will provide an exegesis of the core ideas found in John Rawls's *Political Liberalism*. I focus on Rawls's account of political liberalism as he too takes moral disagreement to be fundamental and intractable. However, he specifically rejects *modus vivendi* political settlements as an appropriate solution to this problem. Rather Rawls attempts to retain the liberal ideal of consensus as necessary for legitimacy. Rawlsian political liberalism, then, offers liberals the best hope of addressing the political problem of moral disagreement whilst aspiring to something more substantive than the realists' proposed solution.

Rawls's transition of thought from his seminal work of *A Theory of Justice* to his second book *Political Liberalism* was largely shaped by his recognition of the fact of 'reasonable pluralism' and how this presents a problem for his earlier position. Reasonable pluralism is manifest when reasonable people, who regard themselves as free and equal citizens and are willing to cooperate under fair terms of cooperation when assured that others will do so, are committed to a diversity of incompatible comprehensive doctrines. A comprehensive doctrine is an exercise in both theoretical and practical reason and roughly translates to a worldview. Examples of such include moral doctrines such as Utilitarianism and Kantianism, as well as religious doctrines such as Catholicism and Judaism. Each doctrine typically entails an associated conception of the good: a view on what is of final value in human life and how values are to be weighed against each other when they conflict.

One important aspect of reasonable pluralism is moral disagreement, insofar as it involves diversity in people's conceptions of the good. For instance, some individuals will reason to the conclusion that consequentialist considerations are paramount and may typically resolve moral conflicts by appealing to something like the principle of utility. Others will be led to endorse a more deontological moral worldview, leading them to conceive a different account of what is of most value in life. Still others will refer to a traditional religious doctrine with

its own set of prescriptions and values to determine their conception of the good. Yet Rawls admits that as it is presented in *A Theory of Justice,* 'justice as fairness' represents a partially comprehensive doctrine of its own, and thus will not be endorsed by all reasonable citizens.[204] This is insofar as *Theory of Justice* invokes moral values and concepts beyond political justice, such as full autonomy, objectivity and moral justification. This, Rawls suggests, entails that his earlier project is not fully publicly justifiable, and presents difficulties in terms of both liberal legitimacy and, relatedly, stability.

According to Rawls's liberal principle of legitimacy:

> "Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason."[205]

This principle is one of Rawls's chief normative assumptions and is derived from the idea inherent in the political culture of liberal democratic constitutional regimes that only the collective body of the public, conceived of as free and equal citizens, can justify the enforcement of state statutes. Yet reasonable pluralism necessarily implies that no single comprehensive doctrine can be the object of consensus amongst reasonable people exercising freedom of thought. Thus, if we accept the liberal principle of legitimacy, then a society which is governed according to any particular comprehensive doctrine cannot legitimate the coercive power of the state.

Rawls thereby agrees with the realist that a universal consensus on moral principles, even amongst those he classes as 'reasonable', is utopian. However, he is not willing to give up on the liberal principle of legitimacy and denigrates *modus vivendi* political settlements as "political in the wrong way"[206]. Not only does such a solution infringe on the liberal principle of legitimacy, but Rawls deems it fundamentally unstable for the reason previously cited: it is ultimately hostage to the particular balance of power within a society.[207]

Moreover, reasonable pluralism entails that any society which is governed according to any particular comprehensive doctrine will fail to win the support of each and every reasonable citizen by addressing their freely exercised reason. And if some reasonable individuals within such a society do not have *internal reasons* as to why they should obey the laws as they are

[204] Rawls, J. *Political Liberalism*, New York: Columbia University Press (1993) p.xvi
[205] Ibid. p.137
[206] Rawls, J. *Justice as Fairness: A Restatement*, Harvard: Harvard University Press (2001) p.188
[207] Rawls, J. *Political Liberalism*, p.xli

specified by the state, it remains fundamentally unstable. Internal reasons are those reasons which one identifies with as normatively significant in light of one's own set of values. Without such internal reasons, an individuals' compliance with the state will always be contingent on externally imposed sanctions. Now of course a state may attain a high level of a kind of stability without meeting the demand that all reasonable citizens freely support its governance. Nonetheless, for Rawls the problem of stability is not merely the practical matter of how to ensure that citizens comply with the statutes of the state, willingly or not. Instead, he wants to attain stability *for the right reasons*, whereby reasonable citizens are motivated to comply for internal reasons. So a society governed according to any particular comprehensive doctrine cannot be stable in the sense which Rawls regards as essential. Thus, Rawls sums up his project in *Political Liberalism* as addressing the following problem: "How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical and moral doctrines?"[208]

Rawls's solution to this problem is to propose a 'freestanding' liberal political conception of justice which is not derived from any particular comprehensive doctrine. Rather it aims to be as neutral as possible between controversial philosophical, religious and moral positions, and so can act as the focus of an 'overlapping consensus' on the most appropriate conception by which to regulate society. This does not represent a mere modus vivendi, as those who live under the political conception endorse it for internal reasons rather than grudgingly accepting it out of pragmatic considerations. For although it is a moral conception, it is one which is compatible with any reasonable comprehensive doctrine, and thus reasonable pluralism does not preclude reasonable citizens from endorsing it. This is because, rather than its fundamental ideas being derived from any particular comprehensive doctrine or conception of the good, they generated from the widely accepted values implicit within the public political culture of liberal democratic states. Moreover, a political conception of justice need only be endorsed by citizens as appropriately regulating the political sphere. It does not strive to provide answers to questions which go beyond this distinct subject matter, such as whether moral facts exist or what a flourishing life consists in; citizens are left to freely answer such questions on their own. Adherents of such diverse comprehensive doctrines as Catholicism, Utilitarianism and Kantian Liberalism can find moral reasons to affirm the liberal political conception from their own internal perspectives. It thus acts as a 'module' which can slot into any reasonable system of values and widely be regarded as the proper basis of societal

[208]Rawls, J. *Political Liberalism*, p.xviii

governance. Citizens widely recognise it as legitimately defining the scope of public reason in terms of limiting what can be publicly justified to others in the public political forum. We will return to the viability of this solution later.

3. The Burdens of Judgement

To take a step back: why does Rawls conceive of reasonable pluralism as an inevitable consequence of maintaining freedom of thought? Because, he suggests, the theoretical and more significantly the practical reasoning of all reasonable individuals is subject to the 'burdens of judgement'. These burdens represent constraints on our reasoning and lead us to reach different conclusions on matters relating to our comprehensive doctrines.

Rawls proposes six burdens which I briefly name and summarise below, although he admits that this list is not exhaustive:

Assessment of Evidence: The relevant evidence is often hard to assess.

Difference in Weighting: Different people give different weight to different considerations.

Conceptual Vagueness: Concepts are vague and in need of interpretation.

Different Experiences: We are subject to different life experiences.

Normative Conflict: Normative considerations sometimes conflict.

Range of Values: There are multiple values, and societies must choose which to prioritise.[209]

These factors are presented in a somewhat disjointed, piecemeal manner, which might leave the reader confused as to how they relate to each other or apply to individual cases of reasoning. Nonetheless, it is possible to draw out a coherent account of the burdens of judgement which hopefully clarifies how Rawls most plausibly conceives of them as leading to a reasonable diversity of both general comprehensive doctrines and more particular conceptions of the good. To illustrate this proposed reading I will demonstrate how the burdens can lead to reasonable disagreement over a particular moral issue – namely, whether or not it is morally permissible to consume meat and other animal products.

When we are assessing the moral permissibility of consuming animal products the first thing which we might notice is that there are many distinct normative considerations which we

[209] Rawls, J. *Political Liberalism*, pp.54-58

might take into account. On the one hand, there is the appreciation of the distinct gastronomic experiences which can only be achieved from eating meat and dairy products, and the supposed nutritional advantages of an omnivorous diet. On the other hand, we might conceive of animal rearing agricultural practices as involving the suffering of sentient beings, causing environmental damage and cultivating vicious character traits, such as brutality or merely a general insensitivity to the welfare of fellow creatures. Assuming that we accept that these are all legitimate candidates for normative considerations relevant to the issue, we can see that the two sets of competing concerns come into conflict. Whilst there may indeed be means of minimising such conflicts, in some cases trade-offs must be made. For instance, we could potentially still gain the gastronomic pleasure we accrue from eating animal products to some extent whilst minimising the animal suffering and environmental impact associated with raising animals. Nonetheless, given the current state of agricultural technology, we could not do so without massively restricting our consumption of animal products and increasing their price. Hence, *Normative Conflict* applies in this and many other cases.[210]

However this burden isn't in itself a direct cause of disagreement. For although we might all recognise that normative considerations sometimes conflict, we might nonetheless agree on what trade-offs should be made between them when they do. Rather, *Normative Conflict* represents a background factor concerning the ways in which normative considerations can interact with one another which sets the scene for reasonable moral disagreement. This is also the case with regard to *Range of Values*. As some have noted, the fact that societies are limited in the range of values which they can simultaneously realise presents reasonable people with the *occasion* for disagreement rather than causing it directly.[211]

Yet when we take burdens of *Assessment of Evidence*, *Difference in Weighting* and *Conceptual Vagueness* into account, reasonable disagreement is to be expected over the issue under consideration. *Assessment of Evidence* entails that given the difficulty in assessing evidence reasonable individuals will often come to different conclusions regarding, for instance, how much negative environmental impact animal husbandry has, how much animals suffer from contemporary farming practices and the extent to which an omnivorous vs. a vegan diet is

[210] Rawls denies that he is committing himself to a strong thesis concerning the multiplicity of values in pointing to this burden. To do so would be to incorporate a philosophically contentious position into part of his overall project, in a way that would undermine his aim of maintaining neutrality between reasonable worldviews. Recall from chapter 1 that I too commit myself only to descriptive value pluralism rather than normative/metaethical value pluralism.

[211] Wenar, L. 'Political Liberalism: An Internal Critique.', *Ethics*, p.41-42

beneficial to one's health. So even if individuals were to agree on how much weight should be attributed to each normative consideration, they might nonetheless disagree on the facts which pertain to the considerations at stake and determine the stakes of the trade-off involved.

As *Difference in Weighting* makes clear, however, reasonable individuals will come to different conclusions regarding the relative importance of the relevant considerations, even when there is agreement on which are in fact relevant to the issue at hand. For instance, some might take animal suffering to be an extremely important normative consideration. They might take it to easily outweigh what is in their eyes the relatively negligible concern that minimising it diminishes our capacity to enjoy the full range of eating experiences available to us. Yet others might take the opposite view. There is also the further difficulty that some will not even recognise the same sorts of considerations as being pertinent. For instance, one might come to the conclusion that animal suffering is not a legitimate normative consideration. Perhaps they might base this on the view that only the suffering of beings with certain psychological characteristics which non-human animals lack is morally salient.

Conceptual Vagueness meanwhile highlights the further complicating factor that the concepts which we refer to when taking the considerations into account might themselves be subject to differing interpretations. For example, two individuals who are independently attempting to determine the extent to which raising animals necessitates animal suffering might come to very different conclusions because they are relying on different notions of what 'suffering' entails. One might take a thin conception of suffering which is constituted by directly painful experience, whilst another could employ a thicker conception which also includes being deprived of autonomy, companionship and access to a certain environment which promotes flourishing.

Finally, *Different Experiences* suggests that people's individual life experiences also contribute to reasonable disagreement. Rawls might be seen here to cite this as an additional independent factor which leads people to disagree, external to the burdens *Assessment of Evidence*, *Difference in Weighting* and *Conceptual Vagueness*. Yet it is not clear how one's unique life experience could contribute to disagreement other than via influencing one's weighting of values, interpretation of concepts and assessment of evidence. So to take a more charitable interpretation, it makes more sense to instead suppose that he is appealing to *Different Experiences* in order to help provide a partial explanation for *Assessment of Evidence*, *Difference in Weighting* and *Conceptual Vagueness*. The implicit claim seems to be that the reason that people

assess evidence differently, attribute different weights to different considerations and interpret concepts differently is partly due to their unique life experiences. To illustrate, one who has been brought up on a farm is likely to have different views concerning the permissibility of eating animal products than another who grew up in a strictly vegetarian commune. This is in virtue of the fact that they are subject to different life experiences which mould their values, concepts and the ways in which they interpret evidence relevant to the case.

These are all intuitively plausible suggestions as to why it may be impossible to reach a consensus on issues such as whether or not we are morally obliged to be vegan, even amongst reasonable persons. In section 5 I will attempt to reinforce Rawls's account by appealing to my proposed account of moral psychology. First, though, I will scrutinise Rawls's contention that one criterion of reasonable people must be that they accept the burdens of judgement and the reasonable pluralism that it implies in order for the project of political liberalism to succeed.

4. Rawls vs. Wenar on Reasonableness

Thus far in my exposition of Rawls I have continually referred to 'reasonableness' without clearly explaining what is meant by such a term, and given that the notion plays a key role in Rawls's schema, this must be addressed. In doing so, I will turn to Leif Wenar's interpretation in his paper 'Political Liberalism: An Internal Critique', and go on to evaluate his claim that the burdens of judgement play an unnecessary and counterproductive role in Rawls's overall argument. Responding to this charge will help me elucidate my contention that recognition of the burdens of judgement is a necessary prerequisite for being willing to forego proposing political terms based on reasons drawn purely from one's own comprehensive doctrine. A willingness to make this concession undergirds the possibility of a stable society united under the banner of a liberal political conception of justice. Hence, my buttressing of the case for the burdens by appealing to moral psychology is important in verifying Rawls's claim that we can realistically hope to eventually reach the overlapping consensus which he envisages.

Wenar notes that "Justice as fairness proceeds through *Political Liberalism* to the soft rhythm of the reasonable"[212] and goes on to cite over thirty different ways in which the term is deployed throughout the book. Indeed, in the previous section I have already referred to

[212] Wenar, L. 'Political Liberalism: An Internal Critique.', *Ethics*,p.34

reasonable pluralism, reasonable persons, reasonable comprehensive doctrines and reasonable disagreement. To make sense of this diverse application of the word, Wenar suggests that it is most charitable to interpret Rawls as intending his proposed definition of 'reasonable persons' as revealing the grounding concept of reasonableness, by reference to which we can derive an understanding of what is meant by the term in all its various guises.

Wenar explains that Rawls takes reasonable persons to have each one of the following five attributes:

"Reasonable persons:

1. a) possess the two moral powers - the capacity for a sense of justice and for a conception of the good; b) Posses the intellectual powers of judgement, thought and inference; c) have a determinate conception of the good interpreted in the light of some comprehensive view; d) are able to be a normal, fully cooperating member of society over a complete life;

2. are ready to propose and willingly abide by principles and standards that constitute fair terms of cooperation, on the condition that others will reciprocate;

3. recognise the burdens of judgement;

4. have a reasonable moral psychology; and

5. recognise the five essential elements of a conception of objectivity."[213]

The full details of these listed attributes need not concern us as yet – for now it is enough to note that they are characteristic of the reasonable in *Political Liberalism*, and as suggested, we should understand all its uses in relation to these characteristics. For instance a reasonable comprehensive doctrine is a doctrine which someone who possesses these attributes could affirm and which does not deny that reasonable persons have these attributes. Meanwhile, reasonable disagreement is disagreement which occurs, or could occur, between reasonable people.[214]

Whilst Wenar is sympathetic to Rawls's overall project of developing an inclusive liberal political conception of justice that can garner widespread acceptance, he holds that it is problematic to insist upon such an expansive notion of the reasonable. Wenar argues that in incorporating the latter three items on the list in his conception of reasonableness, Rawls

[213] Ibid, p.37 Note that criterion 4 might suggest circularity in Rawls's definition of reasonableness. Nonetheless, as Wenar highlights, Rawls's conception of a reasonable moral psychology is based on independent grounds. It relates to an individual's ability to be motivated by conception dependent (as well as object dependent and principle dependent) desires. Wenar goes onto suggest that this criterion is philosophically controversial, as it involves a non-Humean conception of desire. This aspect of Wenar's critique of Rawls's definition of reasonableness is beyond the scope of this chapter.

[214] Wenar, L. 'Political Liberalism', p.38

goes beyond what justice as fairness requires as a political conception and risks saddling his project with some philosophically contentious and needlessly alienating commitments. In fact, Wenar claims, only the first two aspects of reasonableness are necessary in terms of achieving Rawls's aims, and offers an alternative 'limited presentation' of justice of fairness which takes only these two as the defining features of reasonableness.

I will not engage with the whole of Wenar's case here, but concentrate on the more specific claim that building in recognition of the burdens of judgement as an aspect of reasonableness is unnecessary and counterproductive. As Wenar notes, Rawls affirms that recognition of the burdens of judgement serves the purpose of ensuring that reasonable persons will see the inappropriateness of proposing terms of co-operation which are grounded entirely in their own particular conception of the good, and be willing to limit themselves to appealing to public reason when engaging in public political discourse. Yet Wenar suggests that this cannot be the case, since he claims that the first and second criteria of reasonableness are enough to cement this condition. In particular, on his interpretation the capacity for a sense of justice along with the second criterion ensures that reasonable persons will only be willing to impose rules on others which could be endorsed by all, and believe it illegitimate to use political power to repress different comprehensive doctrines. Thus, "There is nothing else here for the burdens of judgement to do, beyond what the limited conception of the reasonable person has already done."[215]

Rawls further suggests that recognition of the burdens of judgement helps convince people to accept liberal constitutional principles such as freedom of conscience. Nonetheless, Wenar contends that the existence of comprehensive doctrines which apparently reject the burdens of judgement, and yet endorse such principles, proves that such recognition is not necessary for doing so. He points to modern Roman Catholicism as an example of such a doctrine. Contemporary Vatican doctrine is emphatic in its commitment to the right of religious freedom. Yet, according to Wenar, along with many other common religious doctrines it roundly rejects an appeal to the burdens of judgement as an explanation of moral and religious disagreement. For whilst the burdens suggest that such disagreements often result merely from the shared difficulties we all face when engaging in reasoning, a religious faith such as Catholicism "characteristically presents itself as universally accessible to clear minds and open hearts."[216] For the religious believer, disagreement is instead to be explained in

[215] Ibid, p.42
[216] Ibid, p.44

terms of factors such as "…worldly temptation, demonic intervention, divine predestination and so on – forces within the horizons of the religious doctrine's sure scheme of value and fact.'[217] They thus deny the influence of the burdens of judgement and so fail to qualify as reasonable on Rawls's terms.

Of course, Wenar concedes, Rawls could simply respond that the fact that Catholicism and other religious beliefs might fail the test of a reasonable comprehensive doctrine does not compel him to moderate the requirements of reasonableness. He could simply admit that such religious doctrines and their adherents will remain unreasonable just so long as they remain committed to denying the burdens of judgement. After all, Rawls is for the most part working in the realm of ideal theory, not trying to avoid offending against every currently existing comprehensive doctrine's sensibilities to get them on board with his political conception of justice.[218] However this is problematic insofar as it excludes a large swathe of individuals who are nonetheless willing to tolerate those of other comprehensive doctrines and accept the liberal political conception of justice. A religious believer might well consider the pluralism of comprehensive doctrines that we encounter in contemporary society as unfortunate and avoidable. They might attribute it to such factors as selfishness, wilful ignorance or irrationality. Yet this is not to say that they would eagerly impose their own conception of the good on others, given half the chance. In fact such religious believers are often morally committed to standing against such an imposition, rather than merely recognising the practical impossibility of realising it. Merely their commitment does not depend upon the recognition of the burdens of judgement, or so Wenar claims. As such there is no relevant justification internal to Rawls's project to deny them the status of reasonable persons. Further, even if political liberalism is primarily an exercise in ideal theory, there is a need to address the concern of the political realist that the project should ultimately be practically realisable. Insisting that those who reject the burdens of judgement are necessarily unreasonable can potentially present a problem for stability if it entails that a large proportion of people are alienated from the political conception as it is presented. This is a point which I will expand upon later. In any case, to summarise for now, Wenar not only claims that

[217] Wenar, L. 'Political Liberalism', p.44

[218] Larry Krasnoff argues that despite highlighting it himself, Wenar doesn't sufficiently address this point, and as a consequence his argument is deeply flawed. (See Krasnoff, L.'Consensus, Stability and Normativity in Rawls's Political Liberalism' *The Journal of Philosophy*, (1998) pp.279-280) This is a consequence of Krasnoff's insistence that the practicality of developing an overlapping consensus is irrelevant to Rawls's position, concerned as it is with political justification. In contrast, as will become clear throughout this chapter, I take such practicalities to be salient considerations, and thus do not think Wenar's point can be dismissed on the grounds that it is of mere practical significance.

recognition of the burdens of judgement is a redundant criterion for reasonableness, but that it is a requirement which is inimical to the general aim of keeping the political conception acceptable to as wide a range of worldviews as possible.

In reply one could, in the first instance, contest this general characterisation of how religious believers who endorse freedom of conscience typically conceive of moral and religious disagreement. Whilst Wenar cites the tenet of papal infallibility and the work of various Catholic theologians in making his case that Catholicism is hostile to the burdens of judgement, this does not necessarily preclude the possibility that most adherents of the faith take a far less strident view. It doesn't seem likely that the majority of vaguely liberal minded Catholics today would maintain, for instance, that a person born into a predominantly Hindu community refrains from accepting Catholicism only by virtue of their own irrationality, selfishness and/or weakness of will. Rather, they are apt to admit that such an individual's failure to accept the truth of Catholicism is more likely to relate to her own individual life experience, being as it is more conducive to the acceptance of a Hindu comprehensive doctrine. Moreover, whilst Wenar suggests that he is appealing to Catholicism only as a representative example of all religious doctrines, it might be argued that most other religious worldviews found within liberal democratic societies are more receptive to the burdens of judgement than he gives credit for. For instance, adherents of moderate Protestantism, Islam and Judaism seem to accept that whilst those who have a conception of the good other than that proscribed by their comprehensive doctrines are wrong, they are not necessarily guilty of selfishness or irrationality.

However this would not serve as an adequate defence against Wenar's main point, which is that it is perfectly consistent in theory to reject the burdens of judgement whilst maintaining a commitment to liberal values for independent reasons. So, what I mean to show is that recognition of the burdens of judgement along with the first and second aspects of reasonableness is necessary to guarantee an attitude supportive of a liberal political conception of justice.[219] This is on the grounds that the first two aspects of reasonableness

[219] This is not to deny that there might be comprehensive doctrines which deny the burdens of judgement and yet are compatible with the liberal political conception of justice for independent moral reasons, internal to its conception of the good. However I doubt Wenar's claim that we can say this of many comprehensive doctrines which exist within contemporary liberal society. For instance, Wenar states that Catholicism's commitment to religious freedom is enough to secure a Catholic acceptance of the liberal political conception. But this merely ensures that Catholics refrain from promoting state enforcement of adherence to Catholicism. It does not bar them from proposing terms of cooperation which are drawn exclusively from their conception of the good and thus cannot be publicly justified. For

do not, in fact, rule out riding roughshod over those whose comprehensive doctrines we take to be necessarily irrational or immoral. The first and second aspects of reasonableness only commit individuals to refrain from proposing terms of cooperation which they take others to be reasonable in rejecting. If one takes someone to be necessarily unreasonable by virtue of them rejecting the terms that one favours, then it is entirely consistent for a reasonable person to refuse to see the need to publicly justify such terms to those who disagree with them. I thus contend that Rawls is correct to maintain that the burdens of judgement "are of first significance for a democratic idea of toleration."[220]

To elaborate, recall that the first criterion of reasonableness specifies that reasonable individuals have the capacity for a sense of justice. The second criterion meanwhile entails that reasonable citizens are ready to propose and willingly abide by principles and standards that constitute fair terms of cooperation on the condition that others will reciprocate. Wenar interprets these criteria alone as committing individuals to something like the liberal principle of legitimacy; he assumes that they ensure that those who are reasonable in his limited sense will not be willing to repress different comprehensive doctrines, and will propose only those terms of cooperation which others can be expected to endorse. They are thus committed to ensuring the public justifiability of political arrangements.[221] However, I take this to be interpreting the implications of the first and second criteria of reasonableness to be broader than is actually the case. For an individual having a capacity for a sense of justice and a desire to cooperate with others under fair terms does not necessarily entail that they will be reluctant to propose terms of cooperation that others might in fact reject. On the contrary, if one is convinced that those who might reject one's favoured terms of cooperation are doing so because they are necessarily irrational, wilfully ignorant or motivated by the distorting influence of self or group interest, then one generally does not see the need for public endorsement. The implicit practices and values of liberal democratic political cultures do imply that it is a breach of justice and fairness to impose terms on citizens without due reason for discounting their dissent. However, insofar as people perceive their opponents as morally and/or epistemically unjustified in their disagreement, they generally do take themselves to be justified in ignoring them. For individuals take such a failure to endorse the terms drawn

instance, the doctrine of free faith says nothing against promoting censorship of blasphemy, or the restriction of abortion, for reasons which are recognised only amongst Catholics.

[220] Rawls, J. *Political Liberalism*, p.58

[221] Wenar, L. 'Political Liberalism' pp.40-41

from their own comprehensive doctrine to be an indication that their opponents are unreasonable and therefore beyond the need for public justification.

To elaborate, when we morally disagree with one another it is often tempting to ascribe to our opponents the sort of debunking explanations discussed in Chapter 2 and 3 (such as wilful ignorance, irrationality and/or selfishness). For instance those who endorse left wing policies, such as market regulation and the redistribution of wealth, will often accuse those who oppose them of doing so because they are either greedy or in the grip of false consciousness. They may insist that their opponent's typical appeal to economic theories which highlight the efficiency of free market practices is spurious, and borne out of a desire to maintain a status quo which is favourable to their class interests. Meanwhile, those on the political right will sometimes claim that left wingers are motivated by envy rather than social justice, and that the economic arguments against *laissez faire* capitalism are so obviously misguided that they can only be made by those who are either irrational or have been blinded by socialist ideology.

The burdens of judgement suggest a more charitable potential interpretation of one's opponent's reasoning. Insofar as individuals are receptive to the possibility of this alternative explanation of their disagreements, they are more likely to accept that to ride roughshod over their opponent's conception of the good is to fail to respect the need for public justifiability when it comes to political decision making. Note that this is not on the grounds that they come to see that they are unjustified in remaining firmly committed to their own conception of the good. As Rawls makes clear, reasonable persons can be convinced that they are unquestionably correct with regards to their comprehensive doctrine and conception of the good, even in the face of their recognition of the burdens of judgement.[222] Although it is inevitable that individuals will typically take their opponents to be mistaken in coming to different conclusions, what is important is that they do not regard them as necessarily unreasonable. Individuals within democratic societies implicitly adopt the first and second criteria of reasonableness due to the widely shared background values and practices in the public political culture, and this leads to a perceived need for public justification. But people

[222] Rawls, J. *Political Liberalism*, p.63 This may seem like an odd claim – surely, one might naturally counter, recognising that our own reasoning capacities are hindered by the burdens of judgement gives us good grounds to be wary of being too confident in our conclusions? Nonetheless, Rawls claims that it is not necessarily unreasonable to remain stalwart in one's beliefs even in the face of recognition of the burdens of judgement, and that to insist otherwise would be to rule out many otherwise reasonable comprehensive doctrines.

are typically only apt to feel the need to provide a public justification to those whom they take to be reasonable, even if misguided.

What the burdens of judgement do is convince people that those who they may otherwise take to be unreasonable in virtue of their disagreement are in fact not necessarily so. On the other hand, those who reject the burdens of judgement are likely to take themselves to be justified in proposing terms of cooperation deriving solely from their particular comprehensive doctrine that their opponents can't accept. Since they take opponents to be exemplifying an uncooperative attitude in their rejection of such terms, they do not take the requirement of public justifiability to prohibit themselves from doing so. So, the path from the burdens of judgement to a commitment to the political conception runs through recognition that those who hold contrary comprehensive doctrines and associated conceptions of the good potentially do so for reasons which, from their own internal perspective, are genuinely moral and rationally justifiable.

In sum, Wenar is wrong to assume that the first two criteria of reasonableness are enough to secure a commitment to the political conception of justice in all cases. Only the recognition of the burdens of judgement can ensure that those who share those values which Rawls takes to be implicit in liberal democratic public political cultures will be willing to exercise the necessary restraint to facilitate an overlapping consensus on the political conception.

Nonetheless, there is a related point which threatens to undermine Rawls's case for the possibility of an overlapping consensus. That is, despite recognition of the burdens of judgement being crucial for Rawls's project to be feasible, in the real world many people simply do not always recognise the extent of their influence. Many who share liberal values and would otherwise meet the criteria of reasonableness simply reject the proposition that those whom they disagree with on certain issues are not necessarily doing so out of irrationality or self-interest. They might not actually reject the burdens wholesale – they could concede that in some instances they apply, but nevertheless hold that they do not allow for as much reasonable disagreement in certain areas as Rawls takes them to. For instance, whilst many individuals today would accept that general religious disagreement might stem from the burdens of judgement, people are typically less inclined to give a similar explanation for disagreement over particular moral issues, such as abortion or euthanasia. Given that people are so strongly invested in the moral judgements which they favour, it should not come as a shock that they are reluctant to attribute reasonableness to those who come to different conclusions. They will thus be motivated to reject the influence of the burdens in such cases.

In Rawls's presentation the burdens of judgement are not especially well established, leaving us with little to say to such people. For the most part he simply asserts them as facts concerning the nature of reason that plausibly account for the pluralism which he takes to be inevitable under liberal institutions, as evidenced by its growth since the European reformation of the church. But as Wenar points out, Catholics could reasonably favour a very different interpretation of this historical story which does without the burdens as a factor in shaping pluralism, and instead emphasise human immorality or wilful ignorance in explaining it.[223] Indeed, it is not only religious believers who have a conception of reasoning which is incompatible with Rawls's account of reasonable pluralism. Many, if not most, agnostic and atheistic believers similarly account for the extent of the moral disagreement by citing the moral blindness or viciousness of those with whom they disagree rather than appealing to something like the burdens of judgement. If people continue to explain pluralism in this sort of way, then they will take public justifiability to have a relatively narrow scope, and thus remain unreasonable.

An implication of this is that if we want Rawls's project to be practically realisable, we need to present the burdens of judgement in a more convincing manner. Now one might argue that since, as has been previously suggested, Rawls is working within the remit of ideal theory, it still remains the case that nothing hinges on people actually recognising the burdens of judgement in his presentation of political liberalism. One might contend that what is or is not practically conducive to the establishment of an overlapping consensus on the political conception of justice is beside the point: all that matters is that it is realisable under certain idealised assumptions. Nonetheless Rawls himself is perfectly clear that the relative feasibility of his project actually being realised is of no small importance to him. He suggests that although he need not guarantee that we will, in fact, eventually achieve the sort of society legitimately and stably governed by the political conception of justice he imagines, it is not enough that he describes a mere conceptual possibility. Rather, it is incumbent on him to provide the grounds for a "reasonable faith in the possibility of a just constitutional regime."[224] Insofar as this is the case, and widespread recognition of the burdens of judgement gives us further grounds for such a hope, it is vital for the Rawlsian project to provide as compelling a case for them as possible.

[223] Wenar, L. 'Political Liberalism' pp.43-45
[224] Rawls, J. *Political Liberalism*, p.172

5. The Import of the TEA model

I now want to suggest that the case for the burdens of judgement as some of the major sources of moral disagreement can be improved by appealing to the TEA account of moral psychology which I argued for in Chapters 4 and 5. If such a psychologised account of the burdens of judgement can be made then this buttresses Rawls's claims regarding the reasonableness of pluralism, and empirically vindicates what I have argued is a crucial element of his central thesis. Moreover, if this presentation of the burdens of judgement is more amenable to the wider population than Rawls's rather abstract and contestable account, it has the potential to help persuade a wider range of individuals of reasonable pluralism. This, in turn, improves the chances of realising the conditions under which an overlapping consensus could develop and vindicates Rawls's reasonable faith in his project being practicable.

To summarise my argument of the later chapters: the available empirical evidence heavily implicates our affective tendencies in shaping our interpretation of moral concepts, what we morally value and to what extent. They are thus liable to be strong determinants of the details of our particular conceptions of the good.

For instance, those with a strong disgust response are likely to interpret the concept of moral purity as involving refraining from certain sexual practices, and place greater normative weight on values relating to sexual purity than those with a weaker disgust response. Meanwhile, those with a stronger inequity aversion response are more likely to interpret justice as necessarily involving the redistribution of wealth, and weight equality higher than those without such an affective disposition. To the extent that this is the case, moral disagreement can be partially explained in terms of differences between the emotional dispositions of individuals. Furthermore such differences stem from both variations in the non-shared environmental influences that individuals are exposed to and genetic differences between individuals.

What is the upshot of all this for Rawls's account of the burdens of judgement? For one it vindicates his claim that people's conception of the good will inevitably differ as a consequence of the uniqueness of their life experiences. Recall that I earlier suggested that, on my interpretation of Rawls, this burden is to be conceived as fundamental insofar as he intends it to explain the other burdens. One of the major reasons we will always differ over the proper weightings of values, our interpretations of concepts and assessments of the evidence is because the different influences that we've been exposed to lead us to do so. I have argued that non-shared environmental influences not only shape our concept

acquisition and the way in which we assess evidence directly, but also the affective dispositions which indirectly make some concepts and assessments of evidence more psychologically appealing to us than others. Rawls's contention that our unique life experiences contribute to moral disagreement thereby gains some credence. Environmental influences upon us do indeed play a strong role in determining our conception of the good, especially our interpretation of value concepts and the degree of normative weight that we assign to them.

However the evidence which indicates the importance of genetic influences on our affective dispositions also suggests an additional way in which the burdens of judgement can be reinforced. It implicates another major factor which contributes to moral disagreement and which is both inevitable and does not stem from either irrationality or the distorting influence of self-interest on the part of agents. This is that it is simply the case that individuals who possess different genetic dispositions will tend to come to endorse different conceptions of the good. This is because the affective mechanisms which are partly shaped by such dispositions influence which values most strongly resonate with us.

Insofar as this is the case, the TEA model of moral psychology which I propose lends extra weight to Rawls's account. Those who are even minimally committed to certain liberal political values have good internal reasons to exercise restraint in terms of drawing on their own conception of the good when proposing political terms of cooperation. My account provides another reason to suppose that the conceptions of the good of those one morally disagrees with can sometimes be interpreted as both reasonable and internally rational, rather than necessarily stemming from either wilful ignorance or selfishness. It further vindicates the notion that enforcing a society-wide adherence to a particular conception of the good necessarily involves reneging on the liberal principle of legitimacy. Given the evidence, it is fair to suppose that the only kind of society wherein one could conceivably reach full consensus on a conception of the good whilst maintaining freedom of conscience would involve ensuring that the life experiences and genetic makeup of individuals would be utterly identical. Even if such a society were ever to be realisable in practice, it would necessitate such overwhelming control of the lives of individuals that only those with a complete disregard for liberal values would be willing to implement it.

This reinforced, psychological account of the burdens of judgement thereby helps show that Rawls is right to highlight the possibility of reasonable disagreement with regard to conceptions of the good. Given my argument that the burdens of judgement are a crucial

element of Rawls's scheme, this represents a bolstering of the overall argument for political liberalism. Moreover, my account has the advantage that it is not reliant on any abstract claims regarding the nature of reasoning itself, but is rather based upon empirical evidence regarding how our moral psychology operates. Rawls claims that one epistemic requirement of reasonable citizens is that they recognise the five essential elements of a conception of objectivity, and one of these involves drawing inferences and making judgements "on the basis of mutually recognised criteria and evidence"[225]. This being the case, an account of the source of reasonable disagreement backed up by empirical findings is less likely to prove controversial amongst those who might otherwise reject the current presentation of the burdens of judgement.

Hence the TEA model could play an important role in fostering the conditions under which an overlapping consensus is possible. It has the potential to persuade the wider population that moral disagreement can be, and often is, the result of affective variation rather than those that they disagree with being necessarily unreasonable. Once people are convinced of this, they will realise that their own comprehensive doctrine is not necessarily publicly justifiable, and therefore cannot legitimately ground a political conception of justice.

6. Plural Conceptions of Justice

Despite my claim that appealing to my TEA model can help improve the prospects of developing an overlapping consensus, I now want to turn to a line of argument which might suggest the opposite conclusion. I have suggested that emphasising the role of affective variance as one of the major sources of moral disagreement can help reinforce the wider population's commitment to a political conception of justice. However, one might claim that the very same evidence which justifies this emphasis might also imply that such a commitment will never be universally shared.

Recall that Rawls believes that an overlapping consensus on a political conception of justice is achievable in spite of the intractable diversity of reasonable comprehensive doctrines which the burdens of judgement imply, as a consequence of there being certain latent political values shared within liberal democratic societies. However, many commentators have noted that it is not clear that he is justified in claiming that supposedly widely shared values will be enough

[225] Rawls, J. *Political Liberalism,* p.111

to motivate people to put their non-political values to one side when engaging with fellow citizens in the public political forum.[226]

For one thing, many otherwise reasonable people have strong moral commitments which conflict with what Rawls's proposed conception of justice recommends. To take an oft discussed example, many Catholics subscribe to a conception of the good which maintains that abortion is morally equivalent to murder. Such a Catholic may accept that someone might, via the burdens of judgement, come to reasonably disagree with such a conclusion. Nonetheless, they might hold that the importance of preventing what they conceive of as murder is such that they would not be willing to forgo proposing terms which outlaw abortion. They could recognise that such a restriction might very well stand against the liberal principle of legitimacy and require riding roughshod over other liberal political values. Nonetheless, despite their acknowledgement that such concerns are important moral considerations, they could still hold them to be ultimately negotiable in the face of the non-political values drawn from their particular conception of the good.

Rawls insists that most people will take the political values of the conception of justice to be "very great values and hence not easily overridden."[227] He furthermore contends that many people in fact subscribe to only *partially* comprehensive doctrines, which encompass fewer non-political values and are more loosely articulated than fully comprehensive doctrines.[228] As a consequence, they will be more open to integrating those political values embedded within the shared, freestanding political conception into their deliberations over matters of basic justice. Thus potential conflicts of values will be minimised and people will be willing to accept that only political values should govern the political realm, despite some maintaining conceptions of the good which may seem to recommend different political arrangements. However, as Fabian Freyenhagen suggests, these are just not convincing enough reasons for us to refrain from worrying that some people simply will not be willing to forgo the social enforcement of their deepest held moral commitments.[229] Insofar as the account of the influence of the burdens of judgement is reinforced by the TEA model, this seems even more

[226] For a selection of commentators who present some variations on this claim, see Caney, S. 'Anti-perfectionism and Rawlsian Liberalism' *Political Studies* Vol. 43, 2, (1995) pp. 248–264 Clarke, S 'Contractarianism, Liberal Neutrality and Epistemology' *Political Studies* (1999) pp.627-642 Gaus, G 'Reasonable Pluralism and the Domain of the Political: How the Weaknesses of John Rawls's Political Liberalism can be Overcome by a Justificatory Liberalism' *Inquiry* (2010) pp.259-284

[227] Rawls, J. *Political Liberalism*, p.139

[228] Ibid, p.175

[229] See Freyenhagen, F. 'Taking Reasonable Pluralism Seriously' in *Philosophy, Politics and Economics*, (2011) pp.327-334.

pressing a worry. For my account suggests that the manner in which one weighs distinct values against each other is partly dependent on one's affective dispositions. And since such dispositions vary even amongst those who have been exposed to similar environmental influences, some of those who are brought up in a background liberal political culture might still attach great moral weight to considerations which conflict with liberal political values.

Moreover, even if one concedes that our shared political values dramatically reduce the scope for reasonable disagreement concerning conceptions of justice, one might suggest that there is still room for the burdens of judgement to entail at least some such political disagreement. Most people within our own political communities do typically lend some weight to the values of freedom and equality, and regard them as the most salient within the political sphere. Nonetheless, they certainly do not agree on how these are to be interpreted and/or balanced against each other in instances where they conflict. In fact, it seems that some of the most pressing disagreements which make setting the terms of cooperation difficult are precisely those concerning how we interpret and balance various political values.

Rawls does hold that, initially, consensus can only be reached on the constitutional essentials of the liberal state and that a full blown overlapping consensus on a conception of justice is something which develops over time.[230] Whilst we might not currently agree on the liberal political conception of justice, we at least all conceive all citizens as free and equal, which entails an implicit commitment to democracy and the liberal principle of legitimacy. This constitutional consensus represents a step towards the eventual consensus on the liberal political conception of justice. He also has faith that the way in which he models and unpacks the various moral conceptions and considered judgements to which the vast majority of people living in liberal democratic societies are implicitly tied leads us to much more convergence in our conceptions of justice than is presently the case. For instance, the thought experiment of the original position still plays a role in political liberalism by way of prompting us to model our widely shared yet vague political values of fairness, freedom and equality so that we may move to more determinate principles of justice.[231] Finally, the political conception of justice is not meant to prescribe every statute of the state; only the way in which the 'basic structure' of society is governed. The political decisions that fall outside this remit are to be determined according to a democratic procedure guided by the requirements

[230] Rawls, J. *Political Liberalism* p.158-164
[231] Ibid, pp.22-28

of public reason. So, in Rawls's schema, some reasonable political disagreement is accounted for and a means by which it can be legitimately resolved is offered.

Nonetheless in his later work 'The Idea of Public Reason Revisited', Rawls comes to recognise that, although reasonable people can come to a consensus regarding the constitutional essentials, they may not actually come to a complete convergence on the particular conception of justice which he proposes. Rather, there exists a family of liberal political conceptions of justice which different people will hold to be more or less reasonable.[232] The suggestion of a plurality of reasonable conceptions of justice does not trouble Rawls, however, on the basis that individuals need not recognise the conception of justice which governs the political realm as necessarily being the most reasonable. As long as they endorse it as reasonable to some extent, he holds that society can be stable for the right reasons (i.e. because citizens accept it on the basis of a moral rather than practical justification). This subtle move represents a weakening of the requirement of the liberal principle of legitimacy but is necessary given the clear potential for reasonable disagreement concerning political conceptions of justice.

Yet even given this more modest ambition, my reinforcement of the burdens of judgement might imply that it is too much to hope for. For if the extent to which we take values to be normatively salient is largely dependent on our particular set of affective dispositions, then whether we can be expected to endorse the sacrifice of non-political values in the name of political values as reasonable depends on our emotional repertoire. Of course, insofar as we find his arguments convincing, Rawls's modelling of our supposedly implicit moral conceptions can help push us in a generally liberal direction. And living in a liberal democratic society might not only typically inculcate us with certain political values but also cultivate the affective bases which underpin them, insofar as our affective dispositions are malleable in the face of cultural influences. However, since some of the major determinants of such dispositions are genetic and non-shared environmental influences, which are variable across individuals and beyond the reach of our general social environment, we can expect this willingness to endorse the priority of political values to similarly vary regardless of how

[232] Rawls, J. 'The Idea of Public Reason Revisited' in *Political Liberalism* pp.450-452 Note that although Rawls concedes that there will be some variation within the range of liberal conceptions of justice, to count as reasonable each conception must meet certain criteria of liberal legitimacy. Specifically, all conceptions must require publicly funded elections, universal health care and limitations on inequality. As such, despite his concession to reasonable political disagreement, Rawls yet maintains that many individuals within contemporary liberal democracies do not currently hold a reasonable liberal conception of justice.

society is structured. Therefore there is reason to be pessimistic at the prospect of a universally endorsed overlapping consensus on even a range of liberal conceptions of justice.

On the one hand this might be thought not to be a great problem for Rawls, for he portrays individuals who are willing to let their own non-political values take precedence when proposing terms of cooperation with others as, by definition, unreasonable. And political liberalism does not hold that an overlapping consensus on the political conception of justice must include those who are unreasonable in order to be legitimately realised – that aim would surely be utopian. However despite this being the case, the account of moral psychology which I have highlighted does suggest that a certain proportion of such 'unreasonable' persons are likely to persist, and, at the least, their coercion cannot be justified on the grounds that they are necessarily blameworthy of blatant epistemic failings or self-interest. Rather, their reluctance to refrain from imposing their particular conception of the good on others could result from their possessing an emotional makeup which causes certain non-political values to resonate with them particularly strongly, whilst the widely shared political values lack such affective resonance. This leads them to evaluate the former as more normatively significant than the latter. Again, one could yet insist that this doesn't present a problem for the legitimacy of a political conception of justice. Even if their failure to be reasonable can't be regarded as obviously epistemically or morally blameworthy from a non-comprehensive standpoint, their coercion may be justified insofar as it would otherwise be impossible to legitimate any form of state. Nonetheless, even if one takes this line, the prospect of a society in which a proportion of people will never have internal moral reasons to accept the political conception of justice presents a lingering problem for the kind of legitimate stability Rawls wants.

7. Rawlsian Realism

It is for this reason that I hold that something needs to be said to those individuals that Rawls takes to be unreasonable in order for a liberal state to be stable. Doing without a justificatory story to offer the unreasonable person arguably infringes upon even a weakened liberal principle of legitimacy, given that we cannot explain their unreasonableness in terms of self-interest or epistemic failings, and more importantly, neglecting to do so might also prompt them to undermine the stability of a liberal state. I suggest that Rawlsians must at this point borrow from political realism in order to plug this justificatory hole and that my reinforced account of the burdens of judgement helps in doing so. Recall that realists take the stubborn

fact of moral and political disagreement to ultimately undercut the possibility of ever realising the liberal hope for a society where all citizens willingly consent to a particular political authority. Instead, they argue, we must appeal to the importance of maintaining peace and security in order for the state to gain the assent of its citizens. For the realist a political authority attains legitimacy as long as it achieves these goals whilst also managing to manage an effective compromise, a *modus vivendi,* between the various conflicting political ends of its citizens. In the background context of western democracies, this compromise will typically take a liberal form, and most people will accept it for internal reasons, even if they do not take it to be ideal.

But a minority of individuals will yet possess a set of affective dispositions which lead them to denigrate the importance of liberal values; what leads them to assent to the political authority? Well, so long as they accept my reinforced account of the burdens of judgement, then they are in a good position to recognise that attempting to impose their own conception of the good on others would inevitably lead to conflict and instability. For even if someone regards it as permissible to violate the liberal principle of legitimacy, they might nonetheless come to see that permitting the enforcement of particular conceptions of the good would come at the price of a perpetual war between those who hold rival conceptions. Despite the individual's judgement that, all other things being equal, they would be willing to dominate dissenters, their recognition of perpetual disagreement could convince them that given the impracticalities of doing so they should be willing to sign up to a mutual agreement not to do so.

Rawls himself would not, of course, be satisfied with individuals employing this sort of rationale in order to concede to the political conception of justice, given his explicit statement that "No one accepts the political conception driven by political compromise"[233] and aforementioned disdain for *modus vivendi* style arrangements as ultimately unstable. Yet whilst it is true that it would not fully deliver the idealised sort of stability which Rawls is aiming for, as previously mentioned, he has already watered down his ambitions on this score by admitting that not all reasonable citizens will converge on the same liberal political conception of justice. What matters is not that individuals all enthusiastically endorse the same political conception of justice for purely internal reasons, which, as even Rawls eventually admitted, is impossible. Instead, we should settle for achieving the more realistic goal of ensuring that as many as possible have a strong enough mixture of both internal *and* pragmatic reasons to

[233] Rawls, J. *Political Liberalism,* p.171

not actively oppose the political conception. For this is the most pressing manner in which stability could be threatened.[234]

However, to the extent that people widely accept my reinforced account of the burdens of judgement, this latter worry will not be as serious as might once have been thought. For whilst the balance of power between those holding competing conceptions of the good might indeed fluctuate, people have good reason to believe that others will *always* dissent from their own conception. This is especially the context of the large scale societies of modern liberal democracies: the more individuals there are within a society, the more likely that a wide range of conceptions of the good will inevitably gather a sizable number of adherents. In fact, there is also the probability of there being some outliers who reject the prevailing conception of the good even within relatively homogeneous, small scale societies. As noted above, the only way to guarantee that people would all converge on their own preferred conception of the good would be through a combination of genetic manipulation and strict regulation of the environmental influences individuals are exposed to. Such measures are not merely taken as morally abhorrent by the vast majority of people in our current cultural context, but are also, importantly, practically impossible to implement. Thus even those who would be willing to sacrifice almost anything in the name of universal convergence on their conception of the good would have to recognise that their ambitions are simply not feasible, given what we can discern from the moral psychological evidence. As such, they have good pragmatic reason to adopt the Rawlsian attitude, and rely only on shared political values when proposing terms of cooperation, in order to avoid a perpetual struggle with those who will always fundamentally disagree with them.

All of this is not to say that we can be secure in the knowledge that unreasonable people will, in fact, refrain from trying to realise their own conception of the good through political means. Some unreasonable individuals might outright reject my psychologised account of the burdens of judgement, or else deny that the pragmatic benefits of maintaining peace and security in the face of perpetual conflict are enough reason to give up on their political ambitions. And there will always be fanatics who will stop at nothing to attempt to enforce their own values, even in the knowledge that such attempts are ultimately futile. However there is little we can do here in terms of offering such individuals a cast-iron case for not

[234] As Brian Barry points out, with reference to the IRA in Northern Ireland, even a relatively small minority of individuals who fundamentally disagree with the foundations of a particular society can undermine liberal democratic stability. Barry, B. 'John Rawls and the Search for Stability', *Ethics*, Vol.105, No.4 (1995) p. 904

doing so. At this point, all we can do is "contain them so that they do not undermine the unity and justice of society"[235], as Rawls suggests we can legitimately do with all those whom he identifies as unreasonable. Given that my stance provides us with at least something to say to such individuals, we can at least be more confident that we have retained an added layer of legitimacy, and be more hopeful that the number who reject the justification we offer will be small enough so as not to threaten the stability of a liberal state.

In sum, my reinforced account of the burdens of judgement suggests that a liberal society will always produce a minority who will weight certain non-political values highly, to the point that they will prioritise them over the liberal values of the political conception of justice when such values conflict. Such individuals may be deemed unreasonable by the likes of Rawls, but their very existence could threaten the stability of political liberalism. However, to deal with this problem, we can incorporate the realist strategy of acquiring stability through appealing to such citizens for their assent via the pragmatic necessity of compromise. Moreover, my reinforced account of the burdens of judgement strengthens the case for such compromise in fact being necessary, which means that more can be persuaded away from undermining liberalism for pragmatic reasons.

So even if it turns out that Rawls was too optimistic to hope for a society where almost all individuals were reasonable on his terms, we might nonetheless have a basis to hope to develop a relatively stable society governed by a liberal political conception of justice. Most members of this society would endorse the political authority for internal reasons, whereas others will assent to it based on an appreciation of the fact that we simply can't attain widespread agreement on anything more substantive. Whilst many might mourn the fact that we cannot find a better solution to the political problem of irrevocable moral disagreement, if they accept my proposed explanation of it as inevitable and potentially reasonable, they will simultaneously be forced to concede that a political liberalism with a realist slant is the most we can realistically hope for. In Rawls's own words, "We strive for the best we can attain within the scope the world allows."[236]

[235] Rawls, J. *Political Liberalism*, xvi-xvii
[236] Ibid, p.88

Conclusion

In this chapter, I have attempted to apply the implications of my proposed TEA model of moral psychology to liberal political theory. In doing so, I have moved from a purely descriptive endeavour towards normative enquiry. At first blush, this might appear controversial. The recent development of empirically informed moral psychology has ignited fierce debate within ethical theory. Whilst some have contended that this new understanding of moral psychology should help shape our values, others have remained stalwart in their resistance to the influence of the research program on our normative principles. The latter camp often cite Hume's law and the naturalistic fallacy in defence of their position: we cannot directly derive an 'ought' from an 'is', and this equally holds with regard to those facts concerning the psychological mechanisms which facilitate moral judgement itself.

However descriptive truths always influence our normative practices – our knowledge of them is crucial when we are concerned with how to instrumentally realise those moral values which we already hold dear. This in no way offends against Hume's law, which specifically states that only *direct* moves from an 'is' premise to an 'ought' premise are fallacious. Moving from an 'is' to an 'ought' indirectly via an intermediate ethical principle which the 'is' pertains to is not debarred.

When it comes to deciding how we should live together, moreover, facts regarding the motivations and behaviour of humans, and the extent to which they can be directed, become particularly salient. Determining whether a society run according to any proposed set of terms of cooperation will in actuality instantiate the values we desire to bring about or not depends upon what we think about how humans are, or have the potential to be. It is for this reason that some psychologists and biologists who have studied human nature from an ostensibly descriptive standpoint have unwittingly provoked intense ideological controversy.[237] Given that our moral values are such an important component in shaping our motivation and behaviour, then, facts concerning how people come to have these values, and the extent to which we can come to share them or not, become extremely relevant.

Here, I have demonstrated one way in which my TEA model can have important implications for a particular type of political theory. I have argued against Wenar that recognition of the burdens of judgement is in fact a necessary criterion for signing up to an overlapping

[237] See, for instance, Segerstrale, U. *Defenders of the Truth: The Sociobiology Debate*, Oxford: Oxford University Press (2000) for an account of the so called 'sociobiology wars', provoked by the presumed conservative implications of an evolutionary biological account of human social behaviour.

consensus on a liberal political conception of justice. Furthermore, the TEA model of moral psychology for which I have previously argued can help buttress Rawls's case for the burdens of judgement as an explanation of moral disagreement as reasonable. Since Rawls's current case for the burdens is inadequate, persuading individuals of the TEA model is a fruitful path towards creating the sort of conditions under which political liberalism could be realisable. Whilst this same model might also suggest that a certain proportion of people will always be unreasonable from a Rawlsian perspective, such people can nonetheless be convinced to refrain from attempting to politically enforce their own conception of the good on pragmatic grounds. The TEA model of moral psychology may therefore spell out the need to take a leaf from the political realist's book and subscribe to a principle of legitimacy which is not quite as demanding as liberalism traditionally requires. Yet doing so allows us to have reasonable faith in the possibility of attaining a stable society, regulated by a liberal conception of justice, which we can conceive of as legitimate.

Concluding Remarks

The overall argument of this thesis is as follows. Firstly, some moral disagreement both within and between cultural groups is fundamental, and that this stems from individuals sharing similar values that have distinct moral force, and yet differing in the relative normative weight that they assign to each. This, along with independent empirical evidence, indicates that my proposed Two-stage Enculturated Affect (TEA) model is the most plausible model of moral psychology. In turn, this model predicts that fundamental moral disagreement is ineliminable, due to individuals being subject to different environmental and innate psychological influences which determine how much weight they assign to values. Finally, these conclusions regarding moral disagreement can inform liberal political theory – in particular, it underscores the necessity for political liberalism to shift in a more realist direction.

In Chapter 1, I argued that the phenomenology of moral conflict lends credence to descriptive value pluralism. I held that when people are faced with situations of moral conflict where two or more competing values are at stake, they are often able to judge one resolution to be all-things-considered best. Nonetheless, I further suggested that their judgements will leave them with an unnerving sense of moral loss, which Steve De Wijze captures in his articulation of the experience of tragic remorse, and which they take to be an appropriate response. This, I argued, implies that individuals implicitly take a range of values to have distinct normative force, which can only be partly commensurated.

In light of this, a potential interpretation of moral disagreement is of it stemming from individuals weighting values differently, and thus coming to different all-things-considered resolutions when they conflict. Chapter 2 made the case for such fundamental moral disagreement existing between members of distinct cultural groups. In doing so, I discussed a range of anthropological examples of moral disagreement and highlighted the extent to which distinct groups often share similar values, but nonetheless differ in how much consideration they pay to each. I described various 'defusing explanations' which have been proposed that conceive such disagreements as apparent rather than fundamental. These defusing explanations, I argued, may be plausible in some cases, but cannot account for the whole range of intercultural moral disagreements which we encounter. In particular, I

considered Doris and Plakias's recent presentation of cases which seem resistant to defusing explanations, and build upon their argument.

Chapter 3 continued with the theme of substantiating fundamental moral disagreement, but focused on the claim that it also occurs between members of the same cultural group, including relatively narrowly delineated groups. I discussed preliminary reasons why we should conceive of the sort of moral disagreements encountered within our broad societal cultural group as fundamental, citing studies which suggest that such disagreements typically stem from differences in value weighting rather than non-moral disagreement. Further citing instances of moral disagreement within more narrowly defined groups, in particular, between peer groups and between professional ethical theorists, I argued that we also have good reason to sometimes take moral disagreement emerging in these groups as fundamental.

Chapter 4 started by offering an overview of Shaun Nichols' and Jonathan Haidt's accounts of moral judgement and the cultural construction of moral norms. I argued that both were incomplete and each needed to be modified. I articulated Nichols' 'sentimental rules' account of the precise link between affect and moral judgement, along with his 'affective resonance' hypothesis, which concerns the role of affective dispositions in shaping the cultural evolution of moral norms. Finding Nichols' work promising but incomplete, I moved on to discuss Haidt's Moral Foundations Theory, which I found to offer explanatory advantages over Nichols', yet again suffer from certain deficiencies. In light of this, I proposed my own TEA model, which combines the best insights from each theory and adds additional elements to offer a more complete explanation of moral psychology. It was claimed that, in the first stage, although our emotional dispositions are malleable in the face of cultural influences, our evolved, innate emotional learning biases help shape (but not determine) the moral values which are culturally constructed in communities. In the second stage, the moral values constructed within a particular community go on to influence the cultural evolution of moral norms.

Chapter 5, meanwhile, demonstrated how this model has more explanatory power with regards to the phenomenology of moral conflict and the incidence of fundamental moral disagreement when contrasted with competing accounts. I introduced two competing accounts of moral psychology: Jesse Prinz's Emotional Constructivism, which is less nativist than my own TEA model, and John Mikhail's Universal Moral Grammar, which is more nativist. After critiquing each as less well supported by empirical evidence than the TEA model, I compared them each in terms of their capacity to explain moral conflict, and the

pattern of intercultural and intracultural moral disagreement. I argued that while Emotional Constructivism could potentially explain the phenomenology of moral conflict, it fared less well than the TEA model in terms of explaining why distinct cultural groups often share values, and why individuals within the same group often morally disagree. Meanwhile, Universal Moral Grammar can explain the similarity in moral values across cultural groups, yet struggles to explain the phenomenology of moral conflict as well as inter and intracultural moral disagreement.

Finally, Chapter 6 unpacked the implications that the TEA model's explanation of fundamental moral disagreement has for liberal political theory. In light of the problem that ineliminable moral disagreement might raise for liberalism, I first discussed the possibility of turning to political realism as an alternative form of political theory which takes account of disagreement. I next articulated one prominent version of liberalism that takes moral disagreement amongst reasonable individuals as an important part of its justification; John Rawls' political liberalism. I argued against Leif Wenar that recognition on the part of the populace that moral disagreement can be reasonable, via the burdens of judgement, is indeed an important precondition if political liberalism is to work. I claimed that, nonetheless, Rawls does not do enough to convince us that moral disagreement is indeed reasonable. On the other hand, my TEA model can reinforce the burdens of judgement, and thus strengthen Rawls's overall argument. Finally, I claimed that this improved account of the burdens of judgement might imply that a minority of individuals would always have values which would lead them to be unreasonable, and that political liberalism thereby needs to move in the direction of political realism if it is to be both practicable and legitimate.

Throughout the course of this thesis, I have made several distinct contributions to our understanding of moral disagreement, its origins and implications. In chapter 1, I have described the often overlooked phenomenology of moral conflict, and utilised this to claim that individuals engage in moral deliberation in an implicitly value-pluralistic manner. As I noted, moral philosophers such as Berlin and Williams have argued from moral conflict to normative value pluralism before. However, my argument here is original insofar as it draws on the way moral conflict is experienced by agents in order to sketch a purely *descriptive* account of pluralism. Whilst the case for fundamental moral disagreement between cultural groups has been made before, chapter 2 brought new considerations to bear and defended Doris and Plakias's argument from various objections. Moreover, it applied the account of moral deliberation from the first chapter in order to offer some original insight as to what exactly is going on when moral disagreement turns out to be fundamental. Chapter 3,

meanwhile, continued to draw upon my account of moral deliberation to discuss a topic which has been seldom touched upon before – intracultural moral disagreement – and offered novel reasons for taking it to also be fundamental. In Chapter 4 I offered innovative critiques of Nichols' and Haidt's accounts of moral psychology and fused the best elements of both together, along with some additions, in order to produce my own TEA model of moral psychology. Chapter 5 offered further critiques of other competing models of moral psychology and furthermore demonstrated the superiority of the TEA model to such models in an innovative way: by testing their capacity to explain the phenomenology of moral conflict and fundamental moral disagreement as I conceive of it. Finally, in Chapter 6 I applied my TEA model to liberal political theory, and unpacked the implications that it has for the plausibility of Rawlsian political liberalism.

There are certain areas of philosophical interest where the research of this thesis could be applied further. In particular, my work may have important implications for ethical and metaethical theory. I will now note just some examples of areas in ethics and metaethics where my research could be fruitfully applied.

In chapter 1, I hinted that my argument for descriptive value pluralism might render value monist ethical theories such as utilitarianism and Kantian deontology implausible. However, I did not follow this line of thought, and instead concentrated merely on the descriptive psychological implications. Yet if my account of a value pluralistic moral phenomenology, underpinned by the TEA model of moral psychology, is indeed descriptively accurate, then one might indeed develop a critique of monistic normative theories on the basis of their incongruence with our moral psychology. As I mentioned, most would agree that our pretheoretical intuitions have at least some weight when we are attempting to formulate moral theory. An acceptance of this, when combined with descriptive value pluralism, might help make a case in favour of normative value pluralism.

On the other hand, one might argue that if my proposed TEA model of moral psychology holds, then we should instead reject moral intuitions as legitimate sources of moral knowledge. The neuroscientist Joshua Greene has argued that insofar as our moral intuitions are to a large extent a product of affect rather than more cognitive processes, they are not to be trusted.[238] Others have claimed that those moral intuitions which we take to be shaped by

[238] Greene, J. D. 'The Secret Joke of Kant's soul' in *Moral Psychology, Vol. 3: The Neuroscience of Morality*: *Emotion, Disease, and Development*, W. Sinnott-Armstrong, Ed., MIT Press, Cambridge, MA (2007) pp.35-79

evolutionary forces can be 'debunked', since evolution is likely to select for beliefs which prove adaptive rather than those which are true and the two do not necessarily go hand in hand.[239] Stronger versions of this argument hold that since all moral thinking has been contaminated by evolutionary influence, no moral judgement can be verified and we should ultimately be moral sceptics.[240] Since the TEA model holds that our moral intuitions are indeed the result of affect, which in turn is a product of evolution, my work here could help reinforce the empirical assumptions behind such arguments.

Moreover, as I noted briefly in the introduction and chapter 2, moral disagreement is often stressed by moral anti-realists and relativists whilst being dismissed as merely apparent by moral realists. Thus, my own analysis could also have important implications for metaethics. As Doris and Plakias note in their article, although some realists claim to be able to make room for fundamental moral disagreement, it is not clear that they can do so whilst remaining consistent.[241] Insofar as my model strengthens the case for moral disagreement and ultimately attributes it to differences in affective dispositions, this might increase the pressure on moral realists. At the very least, simultaneously affirming both moral realism and the correctness of my TEA model would seem to suggest that our ability to intuit moral truth is down to factors beyond our control. Only those who were blessed with a certain set of genetic predispositions and exposed to a certain range of environmental influences would weight values in such a manner which is congruent with the supposed external moral truth. Of course some realists are happy to admit that only a minority of moral experts are directly privy to the moral truth. However, given that there would be no principled and independent manner of determining who these experts are, this wouldn't bode well for our chances of attaining what realists think of as genuine moral knowledge.

In addition to these further applications of my research, there are also ways in which my argument could be developed further. Firstly, in chapter 1 my claims regarding the phenomenology of moral conflict may be restricted to individuals from within western cultural groups. For the presumption that individuals do in fact experience tragic remorse in the manner that De Wijze argues, and regard the response as appropriate, is based on a combination of anecdotal evidence and examples from western cultural narratives which feature moral conflicts. This may be enough to establish the response as typical within

[239] Singer, P. 'Ethics and Intuitions,' *The Journal of Ethics*, 9 pp.331–352.
[240] Ruse, M. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*, New York: Prometheus Books (1998), Joyce, R. *The Evolution of Morality* Cambridge, MA: MIT Press (2006)
[241] Doris, J. and Plakias, A. 'How to Argue about Disagreement' pp.304-313

western cultural groups, but things may be different in other societies. Given the extent of intercultural moral variation, which I have explored elsewhere in my thesis, we should be wary about generalising too much from the way that individuals from cultural groups defined broadly respond to moral conflicts.[242] Nonetheless, my case would be more complete if I could point to some stronger evidence. In light of this, it would be useful to conduct a cross-cultural survey which tested for whether individuals experienced tragic remorse and regarded it as appropriate within *all* cultural groups.

Secondly, the evidence for intracultural disagreement that I cited in chapter 3 could be bolstered. There, I pointed to moral judgement surveys as a source of evidence for fundamental moral disagreement existing within cultures, and cited a few examples which demonstrated fundamental disagreement amongst individuals from similar cultural backgrounds. However, I also raised the possibility of conducting a meta-analysis of such surveys in order to find stronger evidence for this. Another means of gathering more evidence would be to conduct a new moral judgement survey which specifically aimed to sample respondents who had been exposed to extremely similar cultural influences and test the extent of disagreement amongst them.

Thirdly, my proposed TEA model would benefit from being spelt out in more detail. I contend that my account is an advancement over that of both Haidt's and Nichols', insofar as it provides a more detailed picture of in what sense and to what extent innate affective dispositions are culturally malleable, and how these interact with norms and values. Nonetheless there are still areas of the account which remain underdeveloped. For instance, more could be said regarding how exactly the cultivation and moralisation of innate affective tendencies works, and how the innate dispositions of individuals interact with cultural influences in a manner which leads to intracultural moral disagreement. Moreover, my articulation of the TEA model does not address why, exactly, different cultural groups selectively cultivate and moralise affective tendencies to different patterns in the first place. One possibility is that this is due to interactions with other innate dispositions which shape cultural evolution, such as prestige-bias. This could lead cultural groups to enculturate its individuals to develop and moralise affective tendencies more in line with those its high status

[242]This is not to say that the conclusions of chapter 1 are likely to be entirely culturally contingent. For instance, there are at least some examples of characters seemingly experiencing something like tragic remorse in response to moral conflicts within some non-western cultural narratives. See, for instance, Csikszentmihalyi, M. *Material Culture: Ethics and the Body in Early China* Leiden: Brill (2004) pp.354-355 for an example of recognition of moral conflict in early China.

members are innately disposed towards. There is not yet enough evidence available to determine if this is in fact the case, but this and many other questions could be addressed in a more developed articulation of the TEA model.

Bibliography

Abarbanell, L. and Hauser, M.D. 'Mayan Morality: An Exploration of Permissible Harms' *Cognition,* 115 (2010) pp.207-224

Anscombe, E. 'Modern Moral Philosophy' Philosophy Vol.33, no.124 (1958) pp.1-19

Appiah, K.A. 'More Experiments in Ethics', *Neuroethics,* (2010) pp.233-242

Aristotle, *Politics,* Book 1, Chapters 4-7, trans. Reeve, C. D. C. Indianapolis: Hackett (1998)

Barry, B. 'John Rawls and the Search for Stability', *Ethics,* Vol.105, No.4 (1995) pp.874-915

Bentham, J. *Introduction to the Principles of Morals and Legislation* Oxford: Clarendon Press (1907)

Berlin, I. *Four Essays on Liberty, Oxford: Oxford University Press (1969)*

——— *Against the Current: Essays in the History of Ideas,* London: Hogarth Press (1979)

——— *The Proper Study of Mankind: An Anthology of Essays,* London: Pimlico Press (1998)

——— *The Power of Ideas,* Princeton: Princeton University Press (2000)

——— *Liberty*, Oxford: Oxford University Press (2002)

Berlin, I. and Williams, B. 'Pluralism and Liberalism – A reply.' *Political Studies* (1994)

Blair, R. 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath.' *Cognition*, 57. (1995)

——— 'Moral Reasoning and the Child with Psychopathic Tendencies' *Personality and Individual Differences*, 26. (1997)

——— 'Psychophysiological responsiveness to the Distress of Others in Children with Autism' *Personality and Individual Differences* 26 (1999)

——— 'Reponsiveness to Distress Cues in the Child with Psychopathic Tendencies.' *Personality and Individual Differences* 27 (1999)

Blair, R. et al. 'The Psychopathic Individual: A Lack of Responsiveness to Distress Cues?' *Psychophysiology* 34 (1997)

Blair, R. et al. 'Theory of Mind in the Psychopath.' *Journal of Forensic Psychiatry,* 7 (1996) pp.15-25

Blair, R. Mitchell, D. and Blair, K. *The Psychopath: Emotion and the Brain*, MA: Blackwell Publishing (2005)

Bohannan, P. *Justice and Judgement Amongst the Tiv* Oxford: Oxford University Press (1968)

Boyer, P. *Religion Explained,* New York: Basic Books (2001)

Brandt, A. and Rozin, P. (eds.), *Morality and health* New York: Routledge, (1997) pp.119–169

Brandt, R. B. 'The Significance of Differences of Ethical Opinion for Ethical Rationalism' in *Philosophy and Phenomenological Research* 4 (1944) pp.469-495.

——— *Hopi Ethics: A Theoretical Analysis*. Chicago: University of Chicago Press (1954)

——— *Ethical Theory* Englewood Cliffs, N.J.: Prentice-Hall. (1959)

Brink, D.O. 'Moral Disagreement' in his *Moral Realism and the Foundations of Ethics,* Cambridge, Cambridge University Press (1989)

Caney, S. 'Anti-perfectionism and Rawlsian Liberalism' *Political Studies* Vol. 43, 2, (1995) pp. 248–264

Carruthers, P. Laurence, S. and Stich, S. (eds.) *The Innate Mind Volume 2*: *Culture and Cognition,* Oxford: Oxford University Press (2007)

——— *The Innate Mind Volume 3: Foundations and Future*, Oxford: Oxford University Press (2008)

Chagnon, N.A. *Yanamono: The Last Days of Eden,* San Diego, CA: Harcourt Brace Javanovich (1992)

Chang, R. (ed.) *Incommensurability, Incomparability, and Practical Reason*, Cambridge, MA: Harvard University Press (1997)

Clarke, S 'Contractarianism, Liberal Neutrality and Epistemology' *Political Studies* (1999) pp.627-642

Conee, E. 'Against Moral Dilemmas', *Philosophical Review 91* (1982) pp.87-97

Crowder, G. 'Pluralism and Liberalism', *Political Studies*, Vol. 42, No. 2, (1994) pp.293-305

Csikszentmihalyi, M. *Material Culture: Ethics and the Body in Early China* Leiden: Brill (2004)

Cullen, S. 'Survey-Driven Romanticism', *Review of Philosophy and Psychology,* (2010) pp.275-296

Cushman, Young, L. and Hauser, M. 'The Role of Conscious Reasoning and Intuition in Moral Judgment – Testing Three Principles of Harm'*, Psychological Science*, Vol. 17, No. 12 (2006) pp.1082-1089

Darwin C, Ekman P and Rodger P. "The Expression of the Emotions in Man and Animals" Oxford: Oxford University Press. (1998)

Davis, D and Solomon, M *Te Ara - The Encyclopedia of New Zealand*, found at http://www.teara.govt.nz/en/ updated 13-Jul-12

De Wijze, S. 'Tragic-Remorse – The Anguish of Dirty Hands' *Ethical Theory and Moral Practice* (2004) pp.453–471

Ekman, P. 'Basic emotions' in T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*. New York: John Wiley & Sons. (1999) pp. 45-60

Ekman, P., Sorenson, E. R. and Friesen. W. V. 'Pan-cultural elements in facial displays of emotions.' *Science*, Vol.164, No.3875 (1969) pp.86-88

Foot, P. *Moral Dilemmas and Other Topics in Moral Theory,* Oxford: Clarendon Press (2002)

Fraser, B. and Hauser, M. 'The Argument from Disagreement and the Role of Cross-Cultural Empirical Data', *Mind and Language*, 25, 5 (2010) pp.541-560

Freud, Sigmund. *Three Essays on the Theory of Sexuality*, trans. James Strachey. New York: Basic Books (1962)

Freyenhagen, F. 'Taking Reasonable Pluralism Seriously' in *Philosophy, Politics and Economics*, (2011) pp.327-334

Funk, C. et al. 'Genetic and Environmental Transmission of Political Orientations' *Political Psychology*, vol.34, no.6 (2013) pp.805-819

Garcia, K. and Koelling, R. 'The Relation of Cue to Consequence in Avoidance Learning' *Psychonomic Science, 4.* (1966) pp.123-124

Gaus, G 'Reasonable Pluralism and the Domain of the Political: How the Weaknesses of John Rawls's Political Liberalism can be Overcome by a Justificatory Liberalism' *Inquiry* (2010) pp.259-284

Gaus, G. F. *Contemporary Theories of Liberalism*, London: SAGE (2003)

Giroux, J. 'The Origin of Moral Norms: A Moderate Nativism Account' *Dialogue* 50 (2011) pp.281-306

Gould, S.J. and Lewontin, R.C. 'The Spandrels of San Marco and the Panglossian Paradigm' *Proceedings of the Royal Society, (*1979) pp.581-598

Gray, J. *Isaiah Berlin*, Princeton: Princeton University Press (1997)

Greene, J.D. et al. 'An fMRI Investigation of Emotional Engagement in Moral Judgment'. *Science* 293, (2001) pp.2105–2108

Gregor, M. (ed. and trans.) *The Cambridge Edition of the Works of Immanuel Kant,* trans. Cambridge: Cambridge University Press (1996)

Haidt, J. 'The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' *Psychological Review. 108*, (2001) pp.814-834

Haidt, J. and Joseph, C. 'Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues", *Daedalus,* Vol.133 No.4 (2004) pp.55-66

Haidt, J. and Keltner, D. 'Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition' *Cognition and Emotion*, 13 3 (1999)

Haidt, J., Koller, S., and Dias, M. 'Affect, culture, and morality, or is it wrong to eat your dog?' *Journal of Personality and Social Psychology, 65,* (1993), 613-628.

Halsam, S. A. Oakes, P. J. Turner, J. C. and McGarty, C. "Social categorization and group homogeneity: changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness" *British Journal of Social Psychology* 34 (2) (1995) pp.139-160

Hare, R.M. 'Moral Conflicts' *The Tanner Lectures on Human Value* (1978)

Harman, G. 'Moral Psychology and Linguistics' in ed. Brinkmann, K. *Proceedings of the 20th World Conference of Philosophy: Vol. 1: Ethics* Bowling Green, Ohio: Philosophy Documentation Center (1999) pp.107-115

Hauser, M. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* New York: Ecco Press (2006)

Heine, B. 'The Mountain People: Some Notes on the Ik of North-Eastern Uganda' *Africa: Journal of the International African Institute*, Vol. 55, No. 1, (1985), pp 3—16.

Heuer, F. and Reisberg, D. 'Emotion, Arousal and Memory for Detail' in ed. Christianson, S. *The Handbook of Emotion and Memory,* Hillsdale, NJ: Lawrence Erlbaum (1992)

Hiebert, P.G., *Anthropological Insights for Missionaries*, Grand Rapids: Baker Book House, (1985)

Hill, T.E, 'Kantian Pluralism' *Ethics,* (1992) pp.743-762

Horberg, E.J, Oveis, C. and Keltner, D. 'Emotions as Moral Amplifiers: An Appraisal Tendency Approach to the Influences of Distinct Emotions upon Moral Judgment' *Emotion Review* Vol.3 No.3 (2011) pp.239-240

Hume, D. *A Treatise of Human Nature: A Critical Edition* David Fate Norton and Mary J. Norton (eds.), Oxford, Clarendon Press (2007)

Inbar, Y. et al. 'Disgust sensitivity predicts intuitive disapproval of gays.' *Emotion* 9 (3) (2009) pp.435-439

Joyce, R. *The Evolution of Morality* Cambridge, MA: MIT Press (2006)

Kant, I. *The Metaphysical Elements of Justice: Part 1 of The Metaphysics of Morals*, trans. John Ladd Indianapolis: Bobbs-Merill, (1965)

Kant, I. trans. Gregor, M *Groundwork for the Metaphysics of Morals* Cambridge: Cambridge University Press (1998)

Kelly, D et al. 'Harm, Affect and the Moral Conventional Distinction' in *Mind and Language*, Vol. 22 No. 2 (2007), pp. 117–131

Kleinsmith, L. and Kaplan, S. 'Paired-associate Learning as a Function of Arousal and Interpolated Interval' *Journal of Experimental Psychology* (1963) pp.190-193

Koenigs, M. et al. 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements' *Nature* 446, (2007) pp.908–911

Krasnoff, L. 'Consensus, Stability and Normativity in Rawls's Political Liberalism' *The Journal of Philosophy*, (1998) pp.269-292

Lemmon, E.J. 'Moral Dilemmas', *The Philosophical Review* 70 (1962)

Little, M. (ed.), *Moral Particularism*, Oxford: Oxford University Press (2000)

Lord, C. Ross, L. and Lepper, M. 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence', *Journal of Personality and Social Psychology*, (1979). 37 (11) pp.2098-2109

Mackie, J. L., *Ethics: Inventing Right and Wrong*, London: Penguin Books (1977)

Marcus, R.B. 'Moral Dilemmas and Consistency' in *The Journal of Philosophy*, Vol. 77, No. 3. (1980), pp. 121-136.

Mason, H.E (ed.) *Moral Dilemmas and Moral Theory,* Oxford: Oxford University Press (1996)

McCormick, J.P. *Carl Schmitt's Critique of Liberalism,* Cambridge: Cambridge University Press (1999)

McGrath, S. "Moral Disagreement and Moral Expertise" in Shafer-Landau, ed. *Oxford Studies in Metaethics* 3: (2008) pp. 87-107

Mead, M. *Coming of age in Samoa: A psychological study of primitive youth for western civilization*. New York: Morrow (1961)

Mikhail, J. 'Universal Moral Grammar: Theory, Evidence, and the Future', Trends in Cognitive Sciences, 11, (2007) pp.143-152

——— 'Emotion, Neuroscience, and Law: A Comment on Darwin and Greene' Vol. 3 No. 3 *Emotion Review* (2011) pp.293-295

Miller, R. and Cushman, F. 'Aversive for Me, Wrong for You: First-person Behavioral Aversions Underlie the Moral Condemnation of Harm' *Social and Personality Psychology Compass* (2013) pp.707-718

Miller, R.W. *Moral Differences – Truth, Justice and Conscience in a World of Conflict* Princeton: Princeton University Press (1992)

Modell, B., and Darr, A. 'Genetic counselling and customary consanguineous marriage.' *Nature Reviews Genetics*, 3, (2002)

Moody-Adams, M. *Fieldwork in Familiar Places – Morality, Culture & Philosophy*, Cambridge, MA: Harvard University Press (1997)

Nichols, S. "Norms with Feeling: Towards a Psychological Account of Moral Judgment." Cognition, 84 (2002) pp.221-236.

——— *Sentimental Rules – On the Natural Foundations of Moral Judgement*, Oxford: Oxford University Press, 2004

——— 'On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution' *Philosophy of Science,* 69, (2009) pp.18-30

Nisbett, R.E. *The Geography of Thought – How Asians and Westerners Think Differently*, London: Nicholas Brealey Publishing (2003)

Nisbett, R.E., and Cohen, D. *Culture of Honor: The Psychology of Violence in the South,* Boulder, CO: Westview Press. (1996)

Öhman, A., and Mineka, S.' Fear, Phobias and Preparedness: Toward an Evolved Module of Fear and Fear Learning.' *Psychological Review,* 108, (2001) pp.483-522

Okasha, S. 'Darwinian Metaphysics: Species and the Question of Essentialism' *Synthese* (2002) pp.191-213

Otsuka, M. 'Double Effect, Triple Effect and the Trolley Problem: Squaring the Circle in Looping Cases', *Utilitas*, 20 (1) (2008) pp.92-110

Perkins, A. M. et al. 'A Dose of Ruthlessness: Interpersonal Moral judgment is Hardened by the Anti-anxiety Drug Lorazepam' *Journal of Experimental Psychology: General*, 142, (2013) pp.612–620

Plato, trans. Lee, D. *The Republic*, London: Penguin Classics (2003)

Prinz, J. *Gut Reactions: A Perceptual Theory of Emotion*, Oxford: Oxford University Press (2004) pp.131-157

——— *The Emotional Construction of Morals* Oxford: Oxford University Press (2007)

——— 'Can Moral Obligations be Empirically Discovered?' *Midwest Studies in Philosophy* 31 (2007) pp.271-291

——— *Beyond Human Nature: How Culture and Experience Shape our Lives* London: Allen Lane (2012)

Rawls, J. *Political Liberalism*, New York: Columbia University Press (1993)

——— *A Theory of Justice: Revised Edition,* Oxford: Oxford University Press (1999)

——— *Justice as Fairness: A Restatement*, Harvard: Harvard University Press (2001) p.188

Ross, W.D. *The Right and the Good* Oxford: Clarendon Press (1946)

Rozin, P., Lowery, L., Imada, S. and Haidt, J. The Moral-Emotion Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Ethics (Community, Autonomy, Divinity) *Journal of Personality and Social Psychology, 76*, (1999) pp.574-586

Ruse, M. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*, New York: Prometheus Books (1998)

Scheidel, W. 'Brother-sister and parent child marriage outside royal families in ancient Egypt and Iran: a challenge to the sociobiological view of incest avoidance?' *Ethology and Sociobiology* 17: (1996) pp.319-340.

Schnall, S. Haidt, J. Clore, G.L. and Jordan, H. 'Disgust as Embodied Moral Judgment.' *Personality and Social Psychology Bulletin*, 34 8, (2008) pp.1096-1109

Schonebaum, S.E (ed.): *Does Capital Punishment Deter Crime?,* San Diego: Greenhaven Press, (1998)

Segerstrale, U. *Defenders of the Truth: The Sociobiology Debate*, Oxford: Oxford University Press (2000)

Silverberg, J. and Gray, P. *Aggression and Peacefulness in Humans and Other Primates* New York: Oxford University Press (1992)

Simner, M. 'Newborn's Response to the Cry of Another Infant', *Developmental Psychology*, 5 (1971) pp.136-150.

Singer, P. 'Ethics and Intuitions', *The Journal of Ethics* (2005) pp.331-352

Skitka, L. J., et al. (2002). 'Dispositions, Scripts, or Motivated Correction?: Understanding Ideological Differences in Explanations for Social Problems.' *Journal of Personality and Social Psychology*, 83, p.470

Sleat M. *Liberal Realism: A Realist Theory of Liberal Politics*, Manchester: Manchester University Press (2013)

Smart, J.C.C "Extreme and Restricted Utilitarianism", The Philosophical Quarterly, (1956) pp. 344-354

Smart, J.C.C, Williams, B. *Utilitarianism: For and Against* Cambridge: Cambridge University Press (1973)

Sousa, P. 'On Testing the 'Moral Law'', *Mind and Language* 24 (2009) pp.209-234

Sperber, D. *Explaining Culture: A Naturalistic Approach.* Oxford: Blackwell (1996)

Sterenly, K. 'Moral Nativism: A Sceptical Response' *Mind and Language Vol.*25 (2010) pp.279–297

Stocker, M. *Plural and Conflicting Values*, Oxford: Clarendon University Press (1990)

Thomson, J.J. 'A Defense of Abortion' *Philosophy and Public Affairs*, Vol.1 No.1 (1971) pp.47-66

——— 'Killing, Letting Die, and the Trolley Problem', *The Monist*, 59, (2), (1976) pp.204-217

Thornhill, N. W. 'An Evolutionary Analysis of Rules Regulating Human Inbreeding and Marriage' *Behavioural and Brain Sciences*, 14, (1991) pp.247-293.

Thucydides, Crawley, Richard (trans.) *The History of the Peloponnesian War*, (431 BCE) The Internet Classics Archive, http://classics.mit.edu/Thucydides/pelopwar.mb.txt

Tolhurst, W. 'The Argument from Moral Disagreement', *Ethics,* Vol.94. No.3 (1987) pp.610-621

Turiel, E. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press. (1983)

Turkheimer, E. 'The Three Laws of Behavioural Genetics and What They Mean' *Psychological Science* vol. 9 no. 5 (2000) pp.160-164

Turnbull, Colin M. *The Mountain People*. New York: Simon & Schuster, 1972

Unger, P. 'Causing and Preventing Serious Harm.' *Philosophical Studies* 65 (1992) pp.227–255

Sinnott-Armstrong, W. (ed.) *Moral Psychology Volume 1: The Evolution of Morality – Adaptations and Innateness* Cambridge, MA: MIT Press (2008)

——— *Moral Psychology Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press. (2008)

——— *Moral Psychology, Volume. 3: The Neuroscience of Morality*: *Emotion, Disease, and Development*, Cambridge, MA: MIT Press (2008)

——— 'Is Moral Phenomenology Unified?' *Phenomenology and the Cognitive Sciences,* 7 (2008) pp.85-97

Wason, P.C. "On the failure to eliminate hypotheses in a conceptual task", *Quarterly Journal of Experimental Psychology* (Psychology Press) (1960) 12 (3) pp.129–140

Weber, M. *Political Writings,* eds. P. Lassman and R. Speirs, Cambridge: Cambridge University Press (1999)

Wenar, L. 'Political Liberalism: An Internal Critique.' *Ethics*, vol.106 (1995) pp.32-62

Westermarck, E. *A History of Human Marriage*. New York: Macmillan (1891)

Wheatley, T. and Haidt, J. 'Hypnotic Disgust Makes Moral Judgements More Severe' *Psychological Science* Vol.16 No.10 (2005) pp.780-784

Williams, B. *Problems of the Self: Philosophical Papers 1956-1972* London: Cambridge University Press (1973)

——— *Moral Luck*, Cambridge: CUP (1981)

——— *In The Beginning Was The Deed: Realism and Moralism in Political Argumen*t, Princeton: Princeton University Press (2005)

Williams, M.C. *The Realist Tradition,* Cambridge: Cambridge University Press (2005)

Youssef, F. et al. 'Stress Alters Personal Moral Decision Making', *Psychoneuroendocrinology*, 37, (2012) pp.491–498.