Coherentism in Moral Epistemology: Moral Science or Elaborate Intuitionism?

By

Tomasz George Dembinski Stainer

March 2018

This thesis is submitted in partial fulfilment of the requirements of the award of MA by Research

Abstract:

Moral foundationalists maintain that among our moral beliefs there are some that are justified non-inferentially by something other than just more beliefs. Also, that all of the other beliefs in the system can trace their chain of justification back to members of this special 'epistemically privileged' group. The foundations for moral justification might be in the self-evidence of certain beliefs, the alleged infallibility of some basic moral intuitions, the word of a God, or natural facts about the world, to name a few options.

The core problem with moral foundationalist theories is that there simply are no obvious viable candidates for this epistemically privileged group, there are serious concerns with the options available.

Moral coherentism on the other hand, entirely rejects the claim that there exists this epistemically privileged group of non-inferentially justified beliefs that can function to justify the rest of the system. Instead, they maintain that all moral beliefs are justified only by one's other beliefs, and that the level of justification correlates with the level of coherence.

The problem is that the leading form of moral coherentism, Wide Reflective Equilibrium, holds that justification is achieved by coherence only between types of *moral* belief; specially the more general, 'background' moral beliefs cohering with more specific intuition-based moral beliefs. But with nothing other than different types of subjective moral intuitions regulating each other, this model is open to the charge of intuitionism in disguise, and fails to give any objective justificatory force.

My argument is that the problem with Wide Reflective Equilibrium is it only considers inter-moral connections to measure for coherence. But, if we add empirical, nonmoral beliefs to the process (and overcome the worries this extra step creates), that we yield a significantly more objective, and a positive, theory of moral justification.

Chapter 1 gives an introduction to the paper. Chapter 2 first explains and outlines WRE, and then argues that WRE is still 'rigged' to fit our personal subjective intuitions even after the key literary responses are considered. Chapter 3 presents my alternative version of WRE with the added role for nonmoral beliefs. Chapter 4 considers the relevant objections ad replies to my thesis. Finally, Chapter 5 concludes the work and presents some final thoughts.

Contents

Chapter 1

Introduction

Everybody holds moral beliefs, but what is of fundamental interest to the moral epistemologist is what justifies these beliefs. Much how epistemology generally asks if there is a knowable reality, moral epistemology asks whether or not there is any kind of moral reality, and if we can know anything about it; can we have any meaningful answers to moral questions, or is moral enquiry pointless; is it all just subjective and relative? Some think that there are no real answers to our moral questions at all and all we have are various feelings and emotions regarding particular situations; these feelings and emotions reflecting perhaps only evolutionary traits, or one's social circumstances and teachings. Others believe that the feelings and emotions we have about moral situations are actually indicative of correct moral answers, which might explain why people seem to have so many basic moral beliefs in common with one another. Others still, do believe there are moral answers one can access, but that the way to them is through a particular model or moral theory; such as 'what is right is whatever maximises pleasure and minimises pain', or 'what is right is whatever fits with x-set of moral principles the best'. Moral epistemology questions and critiques positions on the status of our moral beliefs in order to provide answers on whether or not moral knowledge is possible, and if so how so, and if not why not. I will start by briefly outlining some of the main positions within the topic and lead to where moral coherentism fits into the picture.

The moral skeptic maintains that acquiring moral knowledge is impossible for one of two broad reasons: either that while there are moral answers in theory, we (persons) have no reliable way of accessing that moral reality – our moral theories are flawed, our intuitions are worryingly subjective, that there is no successful model for getting correct/incorrect, or even better/worse justified moral answers. Or, that moral knowledge is impossible because there simply

are no answers to moral questions that are out there to potentially become known – that perhaps moral statements are really statements about how we feel and what emotions we have towards various situations. To the two questions of moral epistemology – 'is there a moral reality?' and 'can we have knowledge of it?' the skeptic can answer either way to the first, but always answers no to the second.

Positivist opposition to moral skepticism traditionally falls broadly into two key groups; foundationalism and coherentism. Both of the groups bear some similarity in the sense that they hold to a basing requirement: that for a person to be justified in holding a moral belief, (in addition to the belief being justified) the reason they have for believing it must be because it is justified – rather than, say, because of a guess, or because of some irrelevant superstition. Where these two groups that answer the skeptic differ is in their account of how this epistemic justification happens:

Moral foundationalism is discussed in many forms, but all variations share the view that among our beliefs there are some that get their justification from something other than inferences from our other beliefs – and, that all of the other beliefs in the system can trace their chain of justification back to members of this special group. This group of non-inferred beliefs is often termed 'epistemically privileged' or 'basic', as the chain of justification always stops with them and they require no further appeal to any other beliefs, whereas all non-epistemically privileged/non-basic beliefs must be inferred or deduced from other beliefs in order to be justified. Where moral foundationalist accounts differ is in what beliefs count as being basic/epistemically privileged: they might be the moral beliefs that we just *feel* absolutely certain about, they might be justified by a god via some religious scripture, they might be self-evident to all reasonable persons. Further, most moral foundationalist accounts maintain that the basic beliefs themselves are moral in nature; that we have a basic set of moral beliefs (be they from a moral 'sense', or god, or reason), and our other moral beliefs can track their justification back to these basic beliefs.

Moral coherentism on the other hand, entirely rejects the claim that there is an epistemically privileged group of non-inferentially justified beliefs that act to justify the rest of the system. Instead, they maintain that all moral beliefs are justified only by the other things that one believes. It may indeed be that a coherentist system has some foundational-like beliefs in the sense that one might hold some quite broad or general moral beliefs, such as 'one ought to minimise pain and maximise pleasure' or 'one ought to do as god says' or 'one ought to do what reasonable and unbiased people would agree to do' - these broad moral beliefs may inferentially justify most other moral beliefs, but rather than enjoying epistemic privilege found in foundationalism they must themselves be justified solely by the other things one believes.. There are many variations of coherentism that specify different criteria for measuring this coherence, but generally the "more and better" connections that exist between moral beliefs, the more coherence is achieved and the more justified the contained beliefs are (Sayre-McCord, 1996, p.152).

The main difference between these two approaches to moral justification is that foundationalists will always argue there are some basic beliefs which are epistemically privileged and therefore require no further inferential justification, and that all other moral beliefs link back to these basic beliefs – whereas coherentists reject exactly this claim, and hold that all beliefs are inferentially justified by the other things that we happen to believe.

Non-moral beliefs don't tend to feature in the foundationalist theories for two key reasons: Firstly, because of Hume's is/ought distinction: briefly, that one cannot make a logical inference to something that 'ought to be' from something that 'is' without a further assumed moral belief (which would then itself count as foundational) - if x action maximises pleasure, to know that I ought to do x action would require a further non-inferred *moral* belief (and thus foundational) that I ought to do things which maximise pleasure. However much x does in fact maximise pleasure, I seem to need a moral foundation that 'I ought to do such-and-such types of things' to know that therefore I ought to

do *x*. The force of Hume's distinction means that almost all moral foundationalist theories have a basic *moral* foundation. Secondly, non-moral beliefs are often excluded because it leaves the position open to being identified as a reductive naturalist theory; where all moral beliefs are either reduced to, or supervened on by, non-moral beliefs: for example, 'stealing is wrong' may just mean 'stealing increases suffering', provided that goodness can be reduced to minimising suffering while maximising pleasure, then doing good actions can be explained in terms of the nonmoral features of minimising suffering. Because of both the difficulties with overcoming the is/ought distinction and the avoidance of reductive naturalism, the role of non-moral beliefs in moral epistemology tend to be minimal to non-existent. Similarly in moral coherentist theories, probably the most well-known of which being John Rawls's wide reflective equilibrium model, the role of non-moral beliefs is either completely excluded or greatly minimised due to an avoidance of Hume's is/ought issue and of the charge of reductive naturalism.

The main issue with moral foundationalist theories is finding members that fit the epistemic privilege class of beliefs, that can be agreed upon by all or even most individuals. This may be acceptable for a moral relativist, but to those seeking more objective moral justification this point will be problematic. The beliefs that would be foundational cannot be inferred from other beliefs and must ground the justification of all the other inferred beliefs – but, especially if they are limited to moral content, finding those candidates is difficult. Aside from some religious foundation, moral beliefs that feel certain or appear self-evident to some, will not necessarily seem that way for others. Therefore, the content of potential foundational beliefs will differ between individuals, and so will the resulting moral justification. Further, moral foundational beliefs won't allow for revision or rejection – at least, not by other beliefs in the system, because they are supposed to get their justification separately from the inferred beliefs. But there are inevitably multiple issues and dilemmas with potential foundational beliefs; take Kant's categorical imperative as a foundation and the famous hypothetical problem of having to tell the axeman where one's family is hidden, or of rule utilitarianism permitting slavery, or of the personal

circumstances that can affect one's moral intuitions. At every turn it seems as though 'basic beliefs' still might need to be open to revision, but it is difficult to see how a foundationalist system can allow that move.

The main problem with the moral coherentist theories in the literature, as I discuss in detail in the next chapter, is that the moral beliefs only get their justification in virtue of the other moral beliefs within the system. But if the only fundamental source, or input, of potential beliefs coming into the system in the first place is via a moral sense/intuition, then the whole system of beliefs will be built to cohere with those: if you start with x types of moral intuitions and I start with y types of intuitions, then what is 'most coherent with the rest of the one's other beliefs' will likely be different for each of us. Thus, either the coherentist has to accept a strong form of moral relativism, or that something other than inter-moral belief coherence is doing the justificatory leg work.

Chapter 2 will begin by outlining Rawls's process of Wide Reflective Equilibrium from his *A Theory of Justice (1999)* by explaining how the WRE system works and the specific moral arguments that are made by Rawls using WRE. I then discuss an initial strong criticism by Richard Hare relating to WRE being 'rigged' to fit our initial moral intuitions (Hare, 1989). In the middle section of the chapter I consider a potential line of response to this criticism from Norman Daniels (1996) regarding the background theories in WRE acting as an 'independence constraint' on the moral intuitions used. I move on to discuss in more depth the role of the background theories and their very effectual, but possibly assumptive, contained moral positions. I then discuss how these particularly invested background theories in Rawlsian WRE might be swapped out for alternatives, and the effects this has on the belief that ends up being justified by the system. This is intended to both show a key weakness in Rawls's specific WRE but also to demonstrate how the structure can be abstracted from Rawls and applied to other theories.

Chapter 3 will begin by reiterating the suggestion that the 'independence constraint' in Rawlsian WRE doesn't go far enough to objectively scrutinise

one's quite possibly flawed, moral intuitions. I will consider the 'multiple systems problem' (MSP) criticism (of coherentism in general), in light of the moral independence constraint. I then will move on to arguing for my response to this problem within WRE, which in short will maintain that: by introducing a key formative role for non-moral beliefs in WRE one might be able to better and more independently constrain the moral intuitions, while at the same time avoiding granting either the non-moral beliefs, or the moral intuitions, epistemic priority - keeping the model coherentist in structure. I will argue that this altered version of Rawls's original structure a) provides grounds for a stronger response to the multiple systems problem b) allows for positive moral progress beyond simply refining subjective moral intuitions c) while giving a better prescriptive account, also incidentally gives a more accurate descriptive account of how individuals actually do carry out moral reasoning. Chapter 4 introduces some of the main potential criticisms of my version of WRE and also of coherentist theories in general, and then attempts to respond to these objections. The chapter is structured in an 'Objection…Reply' format for the most part. The final part of the fourth chapter involves some clarification and discussion regarding the objections and replies, and any remaining worries from this. Finally, in chapter 5 I make my conclusions and discuss my proposed altered version of WRE and the objections and replies surrounding it – clarifying where some of WRE's problems may be solved, and importantly highlighting where I may have remaining concerns.

Chapter 2

Wide Reflective Equilibrium

John Rawls explains his well-known method of wide reflective equilibrium (WRE) in *A Theory of Justice* as a part of the process of arguing for the acceptance of his two key moral principles, the 'principles of justice'; namely the principle of maximum liberty, and the difference principle. I want to abstract the WRE method that Rawls uses for selecting his 'principles of justice' from its original context in order to discuss the potential broader application of the WRE structure to other moral questions. In this chapter I first explain what WRE is and how it works, I then discuss how this method of justification could be applied to other moral theories and principles outside of Rawls' particular conception to show that it can be separated but also why this might in fact form an initial problem with the theory, and finally I will consider how Rawlsian WRE might respond. The importance of this stage, and the primary objective of the present chapter is the separation of the WRE structure from Rawls's specific arguments for his two particular principles by way of showing how the same structure can be used to justify different sets of principles to the ones Rawls puts forward. I then also discuss why, although this separation allows one to better understand the structure apart from its contents, it also reveals an obvious flaw.

In Part I of Norman Daniels' book *Justice and Justification* (1996) he unpacks and details the justificatory structure of Rawls' Wide Reflective Equilibrium (WRE), carefully identifies the important stages of justification in this process, and then proceeds to address some of the potential points of criticism. I use Daniels's descriptions of the various stages in WRE as a point of reference throughout this chapter because I have found his method of clarifying and

labelling the individual steps extremely helpful when considering potential problems with WRE.

The process of WRE involves three key categories of beliefs seeking 'equilibrium' with each other within the overall system of beliefs; a) CMJs, b) moral principles, and c) background theories. The initial idea is that you start with some initial individual moral judgements about particular moral instances, and then separate the obviously bad ones from the those that you are pretty confident about to yield 'a) CMJs – CMJs (Daniels, p.22) (Rawls, 1951, pp.182-183) (*fig.1*). The CMJs might be thought of here as our sure-felt moral intuitions. You then propose differing sets of b) moral principles until you find a set that seems to fit with and articulate the CMJs best. Then, adjustments and trade-offs are made between the a) CMJs and the b) moral principles, until a balance or 'equilibrium' is achieved: for example, some strongly felt CMJ may force us to adjust a moral principle to accommodate it where it previously did not, or a strongly supported core moral principle may force us to adjust, revise, or even reject one of our CMJs.

But, as Daniels points out, this would only provide a 'narrow' space for moral reflection because if for whatever reason one started off with especially bad or biased CMJs, then the rest of the system would be built around cohering with those bad judgements. Narrow WRE does allow for the rejection of CMJs, but only if the majority of the other CMJs support the principle being used to reject that particular one. But if the core or bulk-of CMJs are deeply flawed for whatever reason then my principles will also reflect those flaws.

The 'wide' scope of reflection of WRE is to be found by making additional adjustments and trade-offs to both the CMJs and to the moral principles in light of relevant 'background theories' to further test the initial two groups of belief for their acceptability in the system (Daniels, p.23). With these three basic categories in mind, and without prioritising either the CMJs, the moral principles or background theories, we make balancing adjustments and trade-offs between each of the described categories until we reach an equilibrium, or a maximally coherent system of beliefs:

"By going back and forth, sometimes altering the conditions of the contractual circumstances [principles]*, at others withdrawing our judgments* [CMJs] *and conforming them to principle... I refer to* [the method] *as reflective equilibrium."* (Rawls, 1999, p. 18)

To detail the key justificatory steps in the three-part WRE process, Daniels stipulates four 'levels' of justification in WRE (*fig.1*):

Level I refers to the equilibrium that exists between a) CMJs and b) principles

Level II refers to the principle-choosing arguments produced from c) background theories

Level III refers to relevant c) background theories

Level IV refers to "*body of social theory*" used to test I and III for practical feasibility (Daniels, 1996, p.23)

I intend to continue this use of Daniels' level-based system in assessing the various stages of the WRE process, as I have found it to be a helpful way of breaking down and analysing the main structure, while remaining fairly impartial (at least at this stage) on any final epistemological standpoint.

The first obvious potential problem with WRE at this stage is the complaint that it only rearranges the initial individual moral intuitions. This is essentially the charge that WRE is in fact foundationalist in the sense that it only amounts to a type of intuitionism. That the b) and c) categories are 'rigged' to fit our CMJs (Hare, 1989). This objection can also be explained as a charge of circularity: It is our background theories and moral principles that are supposed to constrain our CMJs (in case we have bad CMJs), but the only support or constraint that the background theories and moral principles that are doing the CMJ-constraining get, also seems to come from the body of CMJs. It therefore might seem here that the CMJs are acting to regulate themselves in a roundabout way. This can be seen to form a circle of justification, with nothing to ground that justification unless we consider the CMJs to have special epistemic priority as an input – which might solve the problem here, but would make the theory a foundationalist one with the CMJs being the foundation. Given that CMJs are really just our moral intuitions with

obviously questionable ones set aside, this circularity would seem to amount to an over-complicated form of a fairly basic and subjective intuitionism.

The problem that fuels this charge is that the only input or 'data' available to be worked with in the three-part system is coming from the a) CMJs category. Both the background theories *and* the principles are allowed to be adjusted or indeed to gain support from CMJs. But then one can argue that surely too much weight is being put upon those CMJs – they appear to take a fundamentally more important role than the other two categories. The rigging charge is therefore basically as follows: If b) and c) are essentially a product of/or subject to a), then b) and c) will always be adjusted to best-fit a), or rather 'rigged' to match those first intuitions.

The reason why the rigging of principles and background theories to fit CMJs would be such a huge concern for WRE is that the CMJs only appear to be derived from people's subjective opinions, given that they are produced by a simple system of filtering through intuition-based judgments to find ones we 'feel' are surely right or wrong (Brandt, 1990, p.272). If the whole of the WRE system is rigged to fit those initial moral intuitions, then it would seem to be just an over-thought, or complicated form of moral intuitionism. On this account, although WRE can be described in three parts, the only actual inputs into the process seem to be in the form of CMJs – which are essentially just sure-felt individual intuitions.

Further to this concern, given that it is widely acknowledged that individual's feelings and intuitions are often unknowingly warped by various biases (self-gain, class, gender, historical and cultural circumstances etc.), it does seem problematic, certainly in terms of objectivity, to suggest that moral principles can be justified by their match to the individual's intuitions. Brandt's example suggests how this could become a practical problem: "[the principles] *may have been acquired only just now, say as a result of having read Marx."* (1990, p.272). For example, if two people have just read wildly opposing and extremely powerful political texts, then we can imagine that their methods for filtering out these

'obviously good or bad' judgements would also be pretty different, which will (given rigging) affect the entire system.

I think that Daniels anticipates, and perhaps partially concedes, this point in *Justice and Justification* when he writes;

"The fact that these sources of error have been minimised does give CMJs some modest degree of epistemic priority…" (1996, p.)

An initial answer to the problem might be that the rigging complaint ignores the true relevance of the background theories, and therefore underestimates the potential for 'extensive revision' of these initial and perhaps more subjective intuitions (Daniels, p.27). The reflective process would be 'rigged' if all we do is find some principles that match our intuitions, and reflect no more. But this is the case only in narrow equilibrium. Wide reflective equilibrium involves the additional appeal to the relevant background theories to independently regulate the partial equilibrium that exists between a) and b), while a process of adjusting, trading-off and constraining, is carried out between all three of the categories – with no single part being held completely immune to adjustments. Using our more general and perhaps more agreeable background theories to regulate the CMJs and moral principles, while keeping all three categories open to revision, is supposed to quell the worry that WRE is rigged to only match one's initial CMJs. The crucial idea is that the b) moral principles are not simply the product of our initial a) CMJs, because we are constantly applying an independent check using the c) background theories.

But in response to that answer, one could still maintain that the introduction of background theories doesn't really solve the problem of CMJs appearing to form a hidden foundation for the WRE system, because using background theories at level III of WRE only gives the illusion of further revision simply by adding an extra element. These level III background theories are not drawn from any external or independent source in Rawls's WRE, but are themselves also the product of our set of CMJs that were used for level I equilibrium process in the first place. So the rigging worry remains, that if our individual

(and subjective) CMJs are used to regulate and form both the moral principles *and* background theories – then introducing those background theories does little more than the moral principles can do to *independently* check anything and to stop CMJs from being too heavily weighted within the system. Again, if background theories are still subject to our CMJs, then if we happen to have terrible CMJs, it's hard to see how the background theories can save the system from reflecting those.

To reiterate, the complaint is that the method of using background theories as well as CMJs for selecting between potential sets of moral principles, and ultimately the judgements produced by this, is just as 'rigged' to fit our individual moral intuitions as moral beliefs in a narrow equilibrium would be (Hare, 1989) (Brandt, 1990). At the core of these criticisms is the common concern that ultimately too much weight is being placed on the role of the 'CMJs' – giving them epistemic priority over the background theories and the principles.

The problem can be further extrapolated as follows: Our CMJs act to support or constrain both the b) moral principles in finding partial equilibrium at Level 1 – *and* the c) background theories at Level III, which themselves are used to select between the possible sets of b) moral principles at Level II as a 'tool of further revision'. The obvious concern being that both the principles and the background theories seem to be determined fundamentally by one's CMJs, and those appear to be simply individual subjective judgements made using only our immediate intuitions (Hare, p.82).

The result of the rigging problem is that the WRE process can only serve to systemise a particular set of subjective moral intuitions into a complicated scheme of principles and supporting theories – and that those theories and principles would ultimately be the product of those initial moral intuitions. The issue is that there is a distinct lack of any kind of *objective* justificatory force on this account if WRE just systemises those moral intuitions because of course, people often have very moral intuitions influenced perhaps by their class, culture, gender, or various other external/accidental factors.

The main reply in the literature to this strong problem of rigging is explained by Norman Daniels as follows: The background theories that are used to regulate the equilibrium process between a) and b), must be 'independently constrained' by a different set of CMJs that are significantly 'disjoint' from ones that were used to regulate the equilibrium between a) and b) (Daniels, 1996, p.27). This 'independence constraint' for WRE amounts to a requirement for background theories to be drawn from a distinct body of CMJs; distinct from the ones that are used in level I equilibrium (Daniels, p.27-28). This allows for an independent checking role for the background theories in the sense that the CMJs that are used for supporting principles at level 1 and the CMJs used to justify background theories at level 3, *are independent from each other*. Whether Rawls' WRE actually passes this test or not is discussed further on, but establishing what this independence constraint actually is, and why it should be required of any WRE, is crucial to understanding Daniels's main response to the worry that WRE is just a disguised form of moral intuitionism. The success of WRE justification may thus depend on a kind of revisionary-tool or independent check, what Norman Daniels calls the 'independence constraint', that: Any a) CMJs that are used to support or adjust c) background theories at level III, must be 'significantly disjoint' from the CMJs that are used to regulate the b) moral principles at level I (Daniels, 1996, p.49). The benefit to Daniels' independence constraint is that the c) background theories are not just pre-determined to match the level I equilibrium result, because they are not simply a product of the same set of a) CMJs used at level 1, but are drawn from an entirely distinct body of CMJs.

The abstract idea for this independent check is that our moral beliefs are better justified when, rather than just using our intuitions about the one particular subject – say, abortion, we make then a further appeal to some more fundamental and general moral theories that we have (background theories). While these more fundamental and general moral theories may also garner support and regulation from our moral-sense judgments (the CMJs), those intuitions have nothing to do with our initial intuitions that were specifically about the subject of abortion. Instead, the intuitions that support our

background theories are likely to be about general moral concepts such as 'fairness', 'justice', what 'the good' is, what a person is, etc. Essentially the justification for our moral beliefs is increased by WRE because we're appealing to diverse and separate sets of considered moral attitudes in the process. This is why WRE is so substantially different to basic, traditional intuitionism.

Daniels maintains therefore that any c) background theories used in WRE must to a significant degree satisfy the 'independence constraint' if they are to escape any of the charges of rigging. But, it is important to remember, as I discuss in more detail further on, that even given this 'constraint' there is still a valid concern that although the CMJs used at levels II and III are 'disjoint' from each other - they are still intuition based and not disjoint from the *individual*.

<div align="center">***</div>

The first part of this chapter lays out the basic method of WRE, but importantly has also detailed the underlying justificatory structure of Rawls' WRE, crucially; the application of an independence constraint on the background theories, and the four main justificatory levels. However, a secondary aim of this chapter is to be able to demonstrate that the WRE structure can be separated from its original Rawlsian context, and discussed in its own right, to eventually be applied to entirely different moral arguments. What is of fundamental interest here, is whether or not the general justificatory structure of WRE, especially the independence constraint discussed by Daniels, could be a valid means of justification for any kinds of moral principles at all – and not whether or not John Rawls's two particular principles of how society ought to be governed are acceptable or not. Hence, at heart this is a project in the field of metaethics and not of applied ethics.

In order for me to be able to discuss WREs general structure any further, I first need to sufficiently separate the justification mechanisms involved from the specific moral arguments it makes in Rawls's text, because it largely appears to be only through (or with heavy reference to) those specific Rawlsian

arguments that the WRE approach is ever discussed in the literature. I want to separate WRE from its original contents by showing how other non-Rawlsian theories can fit into the same structure. Specifically, how altering/switching-out the c) background theories at level III, changes what particular principles might get picked at Level II. The separation of Rawls's theories from their wider justificatory structure underscores the point that his particular argument is indeed one very specific WRE, but that 'WRE' as a justificatory force must therefore be more generally applicable. Demonstrating this point is will involve calling into question some of the crucial, but certainly more contentious, theories that Rawls employs in his specific account.

<p style="text-align:center">***</p>

Probably the most important and defining step in Rawls's specific WRE is his social contract theory shaped by the 'Original Position' (OP) thought-experiment, because this is how Rawls selects his two 'principles of justice' from any relevant competing principles.

To very briefly reiterate the OP process: One imagines a group of individuals coming together to decide how to to govern a society, but these individuals are subject to the 'veil of ignorance', that is, they are unaware of any information about any of their specific interests and position in the society until the social contract has been agreed (*fig. 2)*. Individuals do not know how they will end up (or the likelihood of) in the society, so therefore have to devise the rules on the basis they might be placed in any societal position. Rawls maintains that certain CMJs and background theories that we (all) have, compel us to construct our moral principles in order to explicitly avoid the influence of self-interest or other biases that could result from having more knowledge than the other individuals. Eventually these OP individuals will agree on two important principles relevant to political/public justice: 1) To maximise individual liberty, as far as it doesn't take from that of others, and 2) inequalities in wealth and resources are only permissible if they benefit the worst-off in society. (Rawls, 1999)

This well-known OP thought experiment features in Rawls' overall theory as the principle-selecting device used at Level II, and it is meant to be the inevitable embodiment of certain relevant Level III background theories – specifically his ideal of a well-ordered society, and his theory of the person. Rawls' ideal of a well-ordered society is most explicitly laid out in *Reply to Alexander and Musgrave* with the mention of aspects such as, "*Everyone accepts, and knows that others accept, the same principles (the same conception) of justice*" (Rawls, 1974, p.634), where it is patently obvious to see how the background theory directly influences the OP condition (unanimity in the above case).

Similarly, Rawls set out his theory of person in *A Kantian Conception of Equality* as essentially a list of statements about the fundamental nature of people, such as simply *"Persons are rational"*. Again, it should be obvious to see how the background theory here affects the OP conditions, in this case the condition that OP individuals are assumed as rational and self-interested people, so will select Rawls's two principles on that basis – that is, as an insurance option against being left in a 'worse off in society' category that gets no benefits in terms of wealth and resources (Rawls, 1975, p.94).

On reflection, one may therefore argue that the real function of Rawls's thought experiment is to carry a deeper argument or 'deep-theory' about fundamental, yet possibly contentious, beliefs within the relevant Level III background theories (esp. Dworkin, 1973) (Daniels, 1996, p.48). I focus closely on this discussion about deep-theory*,* with the aim of showing that the justifying or rejecting of moral claims using a WRE structure does not necessarily lead to the two Rawlsian principles. I discuss several different perspectives on this from Norman Daniels (1996), Ronald Dworkin (1973), Thomas Scanlon (1973), David Lyons (1989) and to make this point as explicit as possible. All of these accounts give a somewhat different take on which particular background theories are most relevant to choosing moral principles, and on whether or not the contained deep-theories do in fact draw upon separate bodies of intuitions in their equilibrium process (amounting to an independence constraint of the sort discussed above).

To suggest that the actual function of Rawls's OP is in fact the result of a deep-theory, I first must consider the argument that the OP does not function like an actual contract, therefore placing the role of the imagined agreement into question (Dworkin, 1973, p.18). The agreements of hypothetical contracts seem to differ greatly from the agreements of an actual contracts in terms of their justificatory force. It therefore perhaps cannot act as a binding or independent justificatory device in its own right, in the same way that a real contract can do. It is from this questionable role position the OP is put in, that it is suggested that the main purpose of the thought experiment is to carry a specific deep-theory, contained within the background theories, that is fundamentally responsible for which particular of moral principles get chosen among other competing ones at level II (Dworkin, 1973) (Lyons, 1989).

The imagined agreement made between OP individuals seems to be made under very contrived and principle-determining conditions; most notably the veil of ignorance. But imagined, and strange conditions such as the veil of ignorance, may be inappropriate to apply to real people in actual cases. I'll use Ronald Dworkin's mini illustrations, with some minor embellishments[1], to make the point about hypothetical agreements here a little clearer.

Take an initial example: you and I are playing a game of poker when we discover that the deck is one card short. At this point, you argue we throw-in the hand and take our money back – you point out that we *would* have agreed to this solution before the game actually began. But I refuse to quit because I'm winning and I want the money of course – plus I point out that I never actually agreed to your solution beforehand. But I seem to have missed the point - your key argument here is not that I am somehow *actually* bound to an agreement that never took place – but rather, the conditions imposed by my hypothetical agreement are "*obviously fair and sensible*", so that only someone

---

[1] There are some areas of the poker game example that can be altered to make the solution seem more 'obviously fair' – such as how it is discovered that a card is missing. Also, I've included a couple of my own versions of Dworkin's examples

with an immediate opposing interest/unfair advantage would ever reject my solution. (Dworkin, p.18).

The point that this example underscores is that the hypothetical agreement itself is not the core reason for accepting the end-solution: the core reason is that the solution is meant to seem obviously fair and sensible, and the hypothetical agreement is just there to illustrate and allow me to see that obvious-ness from an impartial perspective. Similarly, Rawls' hypothetical contractual agreement doesn't appear to add anything substantial to the argument for the principles once it's accepted that excluding specific advantages for individual people or groups, is an obviously fair and sensible thing to do (particularly when deciding the 'rules to govern social institutions'). In both the poker and OP scenarios, if one accepts that the solutions 'throwing in the hand' or say, 'applying the difference principle' are both obvious solutions to problems once particular self-interests are discounted, *and* that discounting particular self-interests when making those types of decisions is also an 'obviously fair' step – then the argument would be complete – and so the point that I would have also agreed to the same solutions but under very different and completely imagined conditions, is irrelevant. This example clearly begins to put the OP's fundamental use into question.

More specifically, the poker analogy suggests that Rawls' hypothetical agreement, like the poker agreement, can't function as an independent argument for accepting the resulting principles that are produced. There clearly needs to be something separate from the contract that informs us of what is and is not 'obviously' fair, to then be able to know whether actually imposing the OP conditions on real people is relevant or not. But however this step is taken, once it is, then arguably the contract device becomes redundant – because you need a particular set of beliefs to justify Rawls's imagined contract conditions – but that same set of beliefs can then go on to do all of the theoretical leg-work that the contract thought experiment was intended for in the first place. If the theories and intuitions that the OP conditions represent

are all completely acceptable and lead to Rawls' two principles, then it isn't quite clear what additional work the OP is doing.

It would be a mistake though to suggest that hypothetical agreements cannot ever be useful or relevant on the above basis. To give a practical example, if a doctor were to find a man unconscious and bleeding, he may well imagine what kind of agreement the man would make under different and non-pertaining conditions – perhaps if he were still bleeding heavily but was conscious and able to communicate. The doctor of course 'hypothesises' that the man would indeed have consented to a blood transfusion under the different conditions (that he was conscious), and so gives him the transfusion. The hypothetical agreement does indeed seem relevant in this example because it is a necessary and fundamental part of the doctor's solution and reasoning (Dworkin, p.19). For only circumstantial reasons, namely unconsciousness, the argument could not be completed without using the imagined agreement.

The question that arises from these two examples is whether Rawls' OP is in fact more like the poker example, or more like the doctor-patient example – and what it is that *is* fundamentally different about those two examples.

The answer is perhaps found in a distinction between types of interest. Ronald Dworkin makes a two-way distinction between the type of individual's interest being discussed; antecedent interests, and actual. It would be in one's antecedent interest, for example, to bet on the football team that had the best-odds of winning the premiership – however, it is in their *actual* interest to bet on the team that actually does win the premiership. Or, it may be in my actual interest for example, to holiday in the town with the best average weather record for August, but it is in my actual interest to holiday in the town that really does end up with the best weather that August. It can be helpful here to think of 'actual interest' as one's retrospective interests here – (but there can sometimes be an important difference, in that the terms 'antecedent' and 'actual' do not always refer to either earlier or later points in time respectively, it just seems that this is most often the case).

The reason therefore, why a hypothetical agreement seems acceptable in the doctor case is that the antecedent interest and the actual interest match each other. The patient would have consented to the blood transfusion if he had actually been asked anyway, therefore his antecedent interest was the same as his actual interest. The only conditions imposed on the patient in the hypothetical, are not ones which fundamentally oppose his real interests. But the OP argument is substantially different, and it is comparably much closer to the poker example than the doctor-patient one. It is in OP individuals antecedent interest to consent to Rawls' principles given the extreme 'thick-veil' condition hypothetically imposed upon them. However, it is obviously not in the actual interest of rational self-interested, mutually disinterested, individuals to consent to Rawlsian principles – particularly not the 'difference principle'. Rawls arguments hold that because the OP individuals would have consented to Rawlsian principles if they were asked under certain conditions, that it is reasonable to apply those rules to them now, under very different conditions when they do not consent. But this akin to arguing that I should sell you some object of mine that I have only just recently discovered is rare and valuable, for the price that I would have sold it to you the previous week when I did not know it's worth (Dworkin, p.19).

In the poker case a hypothetical agreement is used as a device to exclude considerations of self-interest from the debate in order to make the procedure 'fair', because rejecting the 'throwing hand in' solution is only a rational choice if one is motivated by an unequal advantage – and crucially this process therefore presupposes that prioritising any particular group interests when deciding on a principle for all groups is an 'obvious' example of unfairness. It's this type of presupposition that constitutes the relevant deep-theories discussed here; they are the relevant backgrounds theories that are supported by strong or 'obvious' intuitions.

Rawls' OP is perhaps then a tool to exclude any considerations of individual's particular interests and talents, on the assumption that embedding specific interests intended for individual advantages within public rules is somehow

inherently and 'obviously' unfair. The OP agreement may, like the poker game agreement, be functioning as a kind of vehicular device for pointing to the relevant c) background theories, or 'deep-theories' held by the Original Position individuals regarding the nature of fairness. The contractual conditions and the resulting principles are a just reflection of these deeper held beliefs.

But then there still remains the crucial question of why one should bother with the detour of using the OP, to make an allegedly 'obvious' point from a broader and more general deep-theory about fairness. I have demonstrated using some of Dworkin's examples above that the same argument can be made without using the OP thought experiment – because, if you accept what Rawls clearly claims is obvious about the nature of fairness, then you can argue straight to the principles from those points, without using the OP.

One possible reply to the objection that the OP is an unnecessary device is that it is in the OP individuals actual interest, not just in their antecedent interest, to consent to Rawls' principles once obviously unfair considerations are excluded. The OP perhaps functions as a device for removing any obviously unfair considerations by using the uncertainty condition imposed on OP individuals. Put another way, Rawls's illustration may show that when obviously unfair considerations are excluded from the process, ('obviously unfair' in a way that any rational persons would agree is so), adopting Rawls' two principles of justice are, after this 'wider' consideration, in the best interest of rational self-interested individuals; even for the entirely rational, self-interested, and 'mutually disinterested' people described in Rawls's background theory of the person. It might not in fact be in the actual interest of rational self-interested people for obviously unfair considerations to be allowed when constructing the rules by which to govern their society by, this could set a terrible precedent that may at some point disfavour them. It might seem to be in one's *immediate* interest to allow unfair considerations if it happens to benefit them in that instance, but if rational self-interested individuals would agree that considerations the OP removes are 'obviously unfair', and they don't generally

believe obviously unfair societies are a good, then removing obvious unfairness may very well be in their actual interest – even if the immediate imbalances favour them. In terms of the type of interest on this view, it seems at least possible that the antecedent interest does match the actual interest – only it's a more 'widely' considered and 'reflected' actual interest, rather than just an immediate one.

But in reply, it would seem that Rawls's argument for his two principles could still be made without using the OP, given what is supposed to be 'obvious' to the OP individuals on his account. If it can be demonstrated simply on grounds of 'obviousness' that special advantages provided to particular persons or groups is unfair - then the main argument for adopting two principles would be complete, without reference to a hypothetical arrangement. To clarify this point: If rational/ self-interested individuals would obviously think that giving particular persons or groups special advantages is unfair, *and* that applying this general rule would indeed lead to the level of insurance expressed by Rawls' two principles - then one could surely just prevent "*simply by stipulation"* the providing of these 'special advantages' to any particular individuals – and then the argument for the two Rawlsian principles would again be complete. Moreover, the criticism that 'it is unfair to actually impose the terms of a hypothetical agreement when the relevant conditions and interests are completely different', could be avoided. Despite the response that removing unfairness may be an actual interest of the OP individuals, it still seems valid to question the necessity of the veil of ignorance process. The OP is therefore still more like the poker example than the doctor one, because the doctor can't justify his course of action purely by stipulation; so his imaginary agreement serves a practical purpose. But, with the same basic beliefs about fairness, the poker players can both see that throwing in the hand would be obviously fair immediately, in the same way that OP individuals could see (with matching conceptions of fairness) that prioritising particular groups and excluding others in the distribution of resources is just obviously unfair. This brings the discussion back to Dworkin's notion of deep-theory, and that carrying this

fundamental deep-theory about fairness through the OP scenario is the true aim of the WRE process.

There are a variety of different arguments presented in the literature, about which particular background theories, or rather, which specific premises of those particular background theories, might constitute the relevant deep-theory. I will briefly unpack three quite different positions regarding what might constitute the deep-theory in Rawls's WRE, and I will also explore what might happen to the conditions of the OP, and ultimately perhaps to the principles, if these deep-theory conditions are altered or dropped altogether.

Ronald Dworkin maintains that the fundamental, and all-determining deep-theory in Rawls' WRE is the OP individuals:

"*right to equal concern and respect in the design and administration of the political institutions that govern them*" (Dworkin, 1989, p.50-51).

Also importantly, Rawls himself does actually explicitly assume this individual right in one of his later papers *A Kantian Conception of Equality* (1975, p.94):

"*they [OP individuals] each have, and view themselves as having, a right to equal respect and consideration in determining the principles by which the basic arrangements of their society are to be regulated.*"

Dworkin's argument for why this specific right is what constitutes the all-important deep-theory in Rawls' model, is broadly as follows: Political theories involving hypothetical agreements could be split into three distinct categories of deep-theories they are trying to carry; goal-based theories (e.g. the goal to maximum average utility), duty-based theories (e.g. a duty to not lie), and right-based theories (e.g. a right to maximum possible freedoms). Dworkin argues that Rawls' deep-theory can't be goal-based because it could never justify the individuals' use of their veto power on the basis that a decision was not in their

self-interest, (but was nevertheless in the interest of some wider goal) – but the OP is concerned with how individuals make decisions while acting in self-interest, *and then* get to Rawls' principles of justice. The relevant deep-theory also can't be a duty-based theory, for the same reasons; a duty-based deep-theory could not account for the OP individuals apparent power to veto any decision that would conform them to the deep-duty, but would at the same time compromise their self-interest.

Dworkin's remaining option, is that Rawls' deep-theory must therefore be a right-based theory, since this is the only option remaining that could explain both the unanimity condition *and* the idea that OP individuals will choose the two principles when acting in self-interest. With this established, Dworkin goes on to eliminate some of the specific options for a right-based deep-theory: For example, he argues that while Rawls' relevant deep-theory must be a sufficiently abstract right, in order for it to be in all OP individuals antecedent interest.

But at the same time, Dworkin maintains that the deep-theory right cannot simply be the "very general" right to liberty because, while this general right to liberty might be in OP individuals *actual* interest (because, *as it turns out*, a general right to maximum liberty *will* be likely to aid most people's interests) – Dworkin argues that this fact could not be realised by the OP individuals behind the veil of ignorance, because they know nothing about their individual conception of the good or their particular interests - and crucially, which possible options they are most likely to end up with post-OP. So, for all OP individuals know, they could end up with other interests that could only be satisfied by somehow restricting the liberties of others – moreover, they don't know the probabilities of how likely or unlikely it is that they will have these particular kinds of 'other interests'. This argument can also be supported by Richard Hare's point, that Rawls' denies the OP individuals any knowledge of any particular facts that they could use to work out probabilities concerning their particular interests or conceptions of the good (Hare, 1989, p.102-103). Therefore, Dworkin suggests that the relevant right-based deep-theory in Rawls must be the

"*right to equal concern and respect in the design and administration of the political institutions that govern them*" as this is the only possibility that makes sense of the strong unanimity condition, and the notion that OP individuals will self-interestedly choose Rawls' principles.

In contrast to Dworkin's point of view, Thomas Scanlon maintains that the relevant and OP-determining deep-theory is actually the social ideal of the 'person as a rational chooser of his own ends and means' that Rawls includes among his conditions for a 'well-ordered society' (Scanlon, 1973, p.1033). It is from this theory about the social ideal, Scanlon holds, that Rawls gets to the specific circumstances and conditions of the OP. Scanlon further points to two alternative social ideals that used in could have been used in an otherwise similar WRE system; the communitarian ideal, and the perfectionist ideal. The communitarian ideal is one where conflicts of interest among society members are eliminated by people's relations of "love and sympathy" toward their fellow persons – compare this ideal to Rawls' ideal, of calculating self-interested persons following their own ends and means (yet ultimately still arriving at his two principles) (Scanlon, p.1025). If we reflect on the following statement about ideal societies,
"*Let us begin, then, by trying to describe the kind of person we might want to be…in this way we arrive at the notion of a well-ordered society*" (Rawls, 1975, p.94)
It would seem that the ideal person of Rawls (rational, self-interested etc.) and that of a communitarian (sympathetic towards community) are drastically different – this would result in different imagined-contract conditions because the deep theory is different to Rawls's.

Alternatively, the perfectionist ideal also pushes a very different deep-theory to Rawls'; namely that an ideal society is one where individuals would conform to a single considered-perfect ideal such as some religious ideal of the perfect person, or an ideal of a perfect person in terms of "*human excellence*" in arts, science, culture etc. So the ideal society certainly wouldn't be made up of the 'rational choosers of their own ends and means' as it would be for Rawls, but would instead be made up of excellent individuals judged to be so by their conformity to some 'perfect person' template. The key point to take from this

though, is that Scanlon provides two credible, and two very different social ideals than the one Rawls happens to use. Perhaps there is some future argument to be made from the Rawlsian side as to why the social ideal of the person as an autonomous chooser is indeed the most reasonable, or coherent, or most widely-held, ideal to have in this context. But this is not argued for, or even explicitly mentioned in *A Theory of Justice* at all – this seems to just be an assumption of what is ideal (for Rawls).

David Lyons, not initially dissimilar to Dworkin, suggests that the deep theory in Rawls's WRE amounts to a collection of assumptions about impartiality and fairness, and is what determines the specific conditions of Rawls' OP (Lyons, 1989). David Lyons's analysis of WRE is initially similar to Dworkin's in the sense that he begins by questioning the idea of imposing the outcomes of hypothetical agreements. The contract conditions allow Rawls's OP individuals to select moral principles only on the basis of "*their general knowledge about human beings",* so that they are all starting from the the "*exact same premises"* (Lyons, p.151). Lyons' main argument ultimately, is that Rawls' two principles can only be justified in terms of their level of coherence with a collection of assumptions regarding fairness. And fairness as it's understood by Rawls, amounts to the considering of moral judgements from an impartial, third person, 'general human' perspective.

Lyons's does ultimately differ from Dworkin's because the deep-theory assumptions regarding fairness is seen to raise problems, whereas Dworkin ultimately sides with Rawls. He essentially points to two distinct concerns here: Firstly, the beliefs about the nature of fairness seem to indeed just be a collection of assumptions on Rawls' part, rather than actual argued-for theories;
"*The contract argument presupposes the value of fairness. (which, in turn, commits us, on Rawls' view, to other values, such as impartiality)"* (Lyons, 1989, p.159).]

Secondly, even with any complaints regarding assumptions about fairness set aside, Lyons maintains that a purely coherence-based justification of moral

beliefs is highly problematic anyway on the basis that, unlike coherence justification of empirical beliefs where observations provide some external and objective input, moral beliefs have no such reliable source to cohere with (other than moral intuitions and assumptions) – so even a maximum coherence with one's basic fundamental moral beliefs (such as the nature of fairness or the person) doesn't suggest anything epistemologically interesting. Again it brings one back to the elaborate intuitionism discussed above.

What is the upshot of all of the particular deep-theory arguments, in relation to separating the WRE from Rawls' arguments? What the Scanlon, Dworkin and Lyons arguments show here is that the deep-theory of Rawls's WRE is not obvious to identify, seems to be founded in moral assumptions, and therefore perhaps changeable with other rival deep-theories. The arguments suggest that whether or not imposing the conditions of Rawls' OP contract is reasonable, specifically the unanimity condition and the notion OP individuals will choose Rawlsian principles in self-interest as insurance, depends upon a fundamental but *assumed* 'rights' within the 'background theories' that are applied. Considering the above evidence in support of this point, it's reasonable to argue that changing or altering the nature of the particular deep-theory, will necessarily alter the reasonability of imposing the conditions of Rawls' OP.

Moreover, the above arguments bring into question the entire relevance of the Rawlsian thought experiment, which was a defining feature of his WRE. It is not the case that the OP is just the unavoidable result of using the WRE structure, it is only obvious once one adopts Rawls's specific background theories. Dworkin's work for example, points out precisely that this is a background assumption about the nature fairness, hence we end-up with Rawls's difference principle. But his principles crucially depend on this background theory, and if one does not adopt it, or (more likely) priorities other background theories such as the notion of 'market desert' within their concept of fairness, then the OP isn't the obvious 'principle choosing apparatus' that it is claimed to be: A different idea of fairness could very well

change the 'veil of ignorance' condition and one might want to take account of the 'consideration in the design of the principles' that the individuals *deserve,* rather than taking a purely egalitarian view of fairness.

Scanlon's arguments highlight a credible alternative to Dworkin's initial identification of the relevant deep-theory – specifically, a significant alternative to Dworkin's insistence that Rawls' WRE must be a rights-based theory. Instead, Scanlon maintains that its Rawls' social ideal of persons 'as a rational chooser of his own ends and means', that constitutes the relevant and all-determining deep-theory As was clear from the example, if the deep-theory ideal changes to say a perfectionist one, then the ideal society may be one in which persons conform to *x*-qualities (religious or virtue-based for example), and this will in turn change the conditions of the contract; namely the veil of ignorance condition that Rawls uses to show that, blatant unfairness aside, we should accept his moral principles because rational self-interested people would do so to insure themselves against the worse-off position – on the perfectionist ideal it wouldn't matter what 'rational self-interested own-ends-choosers' would do because the assumed ideal person is entirely different.

Finally, Lyons stance on the deep-theory in Rawls' WRE maintains that the deep-theory in Rawls is in fact a collection of assumptions about fairness as impartiality (as opposed to say, fairness as desert) – again here it becomes clear that changing the deep-theory regarding the nature of fairness changes how the moral principles in WRE are selected: because, 'fairness as just desert' as opposed to 'fairness as equal rights' would almost certainly prohibit the extent of the difference principle; some might just deserve more resources than others so ought to get them, whether or not is directly benefits the very worse off.

By demonstrating and explaining how altering various aspects of the background theories, changes the OP conditions, and changes the product of that particular WRE, I hope to have shown how the underlying justificatory structure can be considered apart from the particular Rawlsian contents that it

was conceived in; to use Norman Daniels's terms, I can now explain the testing of Rawls' arguments as testing the 'soundness' of argument, but testing the overall structure as testing the 'validity'. For the following two chapters, I intend to move away from discussing Rawls' particular arguments, and to test the validity.

Showing how other alternative deep-theories fit into the framework has allowed me to abstract to a level of discussing the fundamental validity of the WRE, coherence-based, structure. But more importantly, the above discussion of alternative deep-theories demonstrates how plausible the changes and alterations to the particular background theories that Rawls himself used actually are. It is clear that altering, or adding to, the background 'deep theories' of, say fairness to perhaps incorporate a concept of dessert as well as equality for example, will drastically alter the conditions of the OP, or even the entire relevance of the OP thought experiment to WRE. The specific but subjective intuition-based background theories that Rawls employs in his WRE is absolutely essential to the outcome of the process and ultimately to the judgments and principles that Rawls ends up justifying.

Moreover, switching the background theories for alternatives allows for different moral principles to become justified, and that is a potentially large worry in terms of objectivity: If everything that WRE justifies is ultimately determined by these 'deep' moral background theories, and these are themselves only the product of an individual's intuitions, then different individuals with different moral intuitions will be able to use a WRE structure to justify entirely different moral judgments. But this strong relativism is the very thing that a *positive* theory of justification must avoid. Also, the above discussion shows that Rawls's specific background theories that he uses to justify the principles are far from obvious. This idea of switching Rawls's background theories for possible alternatives is quite plausible – therefore, the notion that using Rawls's WRE will inevitably justify his two principles for everyone, or even for most people, is problematic. For example, most of us know plenty of people that would include at least some notion of just desert in their concept what fairness is, but this drastically alters the 'OP' idea, and

ultimately the outcome of the whole process. Fairness as equality over fairness as just desert is not just *obvious,* and if the background theories of WRE can be easily switched then chances of achieving any objective justification will be minimised. The structure will be open to a strong multiple systems problem. For WRE to get past, or at least minimise, the effect of this objection, there needs to be something else affecting the moral background theories and rest of the equilibrium process besides general and perhaps strongly *felt* moral intuitions.

Chapter 3

The Role of Nonmoral Beliefs in WRE

The aim of chapter 2 was to explore possible responses for Rawls's WRE to Hare's strong 'rigging' problem, by giving an initial reply using Daniel's commentary on the issue: that by employing the c) background theories and then not giving any epistemic priority to a), b) or c), and developing an 'independence constraint' for the background theories, that we can get past the response that the system is simple subjective intuitionism in disguise.

Ronald Dworkin's work on this, although he defends Rawls' theory ultimately was used to argue that the actual function of WRE is instead to carry a particular deep theory of Rawls's into his coherentist arguments for moral principles (Dworkin, 1996). I then used some other examples of Rawlsian deep theory, such as Thomas Scanlon's work on WRE, to show (unlike what Dworkin argued for ultimately) that Rawls' deep theory is not as straightforwardly acceptable as he seems to assume in *A Theory of Justice* (1999). Other rival deep theories, such as fairness as desert, might just as easily be considered within the WRE framework. I demonstrated this point by showing how a variety of arguments can fit into the deep theory/WRE structure – such as positing a rivalling deep theory about say, a 'right to just desert', or 'right to one's own property', as opposed to Rawls' 'right to equal respect and consideration'.

But the WRE version as put forward by Daniels with his so-termed 'independence constraint' – and the introduction of the role of deep theory, seems is just a better and more focused version of the original basic WRE that suffers significant criticisms about being rigged to fit with one's CMJs. Hare

argued that WRE merely systemises our subjective intuitions, or at least just the 'considered' intuitions of Rawls' WRE. Ronald Dworkin's stance varies from this only slightly in that it's not all of the certain-intuitions that are fundamental to the whole system, but only a particular subset of beliefs (essentially just one belief for Dworkin) that forms what he terms 'deep theory'; the abstract right to equal consideration discussed in chapter 2.

My two main points made by the end the previous chapter though are that the CMJs justifying
deep-theory are equatable to the intuitions that merely get systemised in Hare's description of WRE. Once it is demonstrated that a variety of other rival deep theories can just as easily replace the one that Dworkin highlighted, such as an abstract right to market desert, then the rigging concern becomes relevant once again. Dworkin's deep theory analysis of WRE would not seem especially defining in an epistemological sense then, because individuals have substantially different deep theories, or at least prioritise them differently to Dworkin and Rawls and therefore can use this version of WRE to justify to equal degrees, entirely contradicting moral belief systems.

Moreover, the independence constraint that Daniels suggests does not get past this issue: The independence constraint is supposed to be that the CMJs used at level 1 in partial equilibrium with the b) moral principles – and the CMJs used to support and constrain background theories at level 3, be separate or 'disjoint' from each other. But, as seen with the deep theory analysis, even if the relevant CMJs justifying the deep theory (about, say 'fairness') are disjoint from the CMJs that used in partial equilibrium at level 1 with the proposed moral principle (such as inequalities in resources are only justified when they benefit the very worse off) – it doesn't stop them from being *our* CMJs, and individuals do have opposing strong intuitions about various deep theory options. In turn, the deep theory that gets justified will simply be the one that *we* happen to have strong intuitions or 'CMJs' about. In that case, Hare's complaint that WRE is rigged to match intuitions we have, is still relevant even after Daniels's independence constraint.

The criticism that is therefore implicit in both Richard Hare's work on this, and in the line I took given Dworkin's account, is what is usually termed in epistemology a 'multiple systems' problem (MSP): That simply justifying moral claims based on their coherence with only our initial subjective intuitions, or even 'extensively revising' intuitions through the background beliefs - that themselves are still regulated by other moral intuitions (of ours) still doesn't get us that much further in terms of *objective* justification than the classical intuitionist theory does; all this being the thrust of Richard Hare's complaint in the first place. The only difference found by considering Dworkin's supporting arguments regarding the idea of deep theory was that it is more realistic to suggest that it's perhaps one or two, and not *all*, of our basic sure-felt intuitions, that so fundamentally affects the direction of the rest of the belief system; in the sense that many of the sure-felt intuitions are shared by most, but it is the few that differ that really matter when it comes to moral disagreements and dilemmas.

This lack of objectivity that results from basing the WRE system in our intuitions is problematic for justification because if the multiple systems problem gets traction, then WRE produces nothing interesting for moral epistemology, unless one bites a strong-subjectivist bullet so to speak. In that case reflective equilibrium would just describe (probably completely incorrectly) how people *do* carry out moral thought processes, and why it is that people hold such moral differences; why different individuals adopt different moral belief systems, and therefore have moral disagreements (Haidt, 2008, p. 1024-1026). But that would have to be a question for moral psychology or anthropology because just a description of moral thinking makes no positive justificatory claims about the moral judgments that result, it just describes how we got to it - again, probably incorrectly in that case (Daniels, 1996).

In short, the foundation of the criticism that is discussed throughout the second chapter comes from a wider concern with all coherence-based theories in epistemology; that the underlying justificatory structure can be used to justify multiple, and therefore "rival" belief systems – and then can't 'progress'

any further without assuming some subjective notion as fundamentally true (Klein & Warfield, 1994). But if WRE can't then further select between these competing rival systems, without *assuming* a deep-theory-type of intuition base, then it can be said to be able justify multiple systems at the same time – which epistemologically speaking is problematic.

Even after considering Dworkin's deep-theory response, and therefore the partial weakening-down of Hare's stronger critique that the strong role of intuitions in WRE can be used to justify practically anything, it is still apparent by demonstrating the variety the potential deep theories in chapter 2 that Rawls' specific arguments for his two principles are not the inevitable result of using Rawls's WRE structure, but they are only a possible result. Rawls' WRE is just a WRE among a selection of competing rivals. But the key worry is that if we can have many WRE-justified systems (based on different 'deep theories'), this does indeed leave the whole model open to a multiple-systems type of response.

This is a worrying issue of under-determination with WRE. That measuring the coherence of our moral judgments with only our sure-intuitions, and with deeper background theories as reflective equilibrium does, will end up justifying contrasting moral belief sets. Even though we are taking the products of WRE to be a constructivist moral theory, instead of the realist view, it still seems to lack the needed epistemological resources to determine (or at least narrow down) between potential rival systems. At best, using the Rawlsian WRE model we can maybe rule out 'obviously bad' beliefs – or at least ones that seem obviously bad to most – but solving more pressing or contentious moral matters such as the political arguments in Rawls remains a challenge.

<p style="text-align:center">***</p>

My suggested response to the MSP still present in WRE is essentially that while it is the case that the Rawls/Dworkin/Daniels version of WRE justifies multiple rival systems of belief, this only happens to such a worrying degree because the Rawlsian model of justification intentionally excludes the role of

relevant nonmoral or non-normative beliefs, and also the role of non-doxastic assent to beliefs; in fact it takes explicit steps to exclude them. Norman Daniels especially, in avoiding either a reductive naturalist account or avoiding the is/ought rule, makes his so-termed independence constraint exclusively an inter-moral belief constraint, where only the significantly disjointed CMJs regulate the background theories.

I suggest therefore that while the traditional Rawlsian WRE model discussed in Chapter 2 essentially fails due to a strong version of the MSP, once we take into account the relationship between empirical beliefs or experiences and our moral beliefs and the role this can have in a WRE system, we are then in a much better positon to further narrow down rival moral belief systems and conflicting moral beliefs, therefore increasing the level of justification and tackling (or reducing - at worst, still only partially) the MSP. That said, it is also important to avoid claiming that the nonmoral supervenes on the moral because then we would have a naturalist/foundationalist account, where the nonmoral 'facts' and experiences essentially act as foundational for the moral beliefs. But a valid coherence theory has to avoid giving any one group of beliefs an epistemic priority in this sense.

The level of justification that might further be achieved by considering relevant empirical beliefs and experiences, compared to only inter-mora support, will essentially determine how strong my answer is going to be. If, even when employing relevant empirical beliefs alongside (not instead of) the internal moral reflection method in WRE, there still remains an issue of underdeterminiation, then the force of the MSP would still somewhat remain after considering my altered model, then the MSP would still hold to some degree because we would still end up with multiple, equally coherent, and yet still contradictory systems of beliefs - even when all of the relevant empirical facts are considered.

However, I would still then intend to argue that even on this weaker interpretation of my solution, that the consideration of relevant empirical beliefs will still inevitably narrow-down potential rival belief systems, more so

than not considering them would do. So at worst, perhaps there is still a Multiple Systems Problem, but the 'multiple', given my answer, is likely to be a lower one.

The stronger interpretation of my answer is therefore that by considering all of the relevant empirical information alongside our moral reflection, one could *in theory* make all moral dilemmas and disagreements more tractable and answerable in some objective sense. Given that this ideal of the theory is obviously unlikely in practice, the argument for why disagreements and dilemmas will likely remain is that conducting all of the relevant moral reflection (removing all inconsistencies) – and considering *all* of the relevant nonmoral beliefs and experience in the process, is just impossible to actually do. Having a perfectly consistent 'web' of moral beliefs, coherent with a perfectly consistent and complete system of nonmoral beliefs and experience – will at best always remain an ideal (BonJour, 1985). But still, on the stronger interpretation of my argument, the coherence-based model of moral justification is more epistemologically significant because it provides a positive account of justification, as well as a more descriptively accurate one, in comparison to Rawlsian WRE which doesn't include the role of relevant nonmoral beliefs at all.

Further, by implying an account of coherence-justification rather than coherence-truth, the fact that employing empirical facts might not in practice narrow multiple rival systems down to one single system, is not completely detrimental to the theory – so long my inclusion of the empirical facts can narrow the rival systems down more so than the original WRE model can do. The complaint that the model still cannot be reduced down into one single guaranteed system, doesn't refute my argument because the type of system being put forward is an ongoing and continuous process.

In that sense, WRE with added nonmoral beliefs is analogous to the justification of scientific beliefs: In both there is a continuum of data inputs that are, or might be, relevant to the justification of the potential beliefs at hand. Further, these inputs don't simply come to a natural end and are never

really considered 'complete' or 'exhausted'. In the natural sciences we bring our data inputs, our investigations, to an end once one particular belief or set of beliefs (consistent with all the other things we believe) is probabilistically more likely, to a high enough degree than some opposing belief. But never in practice do these beliefs rise to some perfect or unrevisable level. In the moral case, the inputs of our inter-moral reflection *and* of our employing relevant nonmoral beliefs, don't come to an end by themselves, because in practice we can always seek more and better relevant nonmoral beliefs, and we can always carry on eliminating any inconsistencies, however minor, between our connected moral beliefs.

In Rawls WRE, the only key inputs into the system, the 'data', are the CMJs. This means that the Rawls/Daniels/Dworkin account of WRE attempts to build everything ultimately upon these sure-intuitions; the a) CMJs and a') disjointed CMJs that constrain c) background theories. But this is what leaves the strong MSP issue, even given the arguments regarding deep-theory, because as was demonstrated by swapping-out the deep theories in the previous chapter, we are still left with many possible options – all rival systems appearing to be coherent in their own right. The only exception for nonmoral beliefs on the Rawlsian account is in testing for the 'feasibility' of the principles (Daniels, 1996) (Rawls, 1999).

But once we add the role nonmoral facts, in constraining our moral intuitions and background theories further, it seems that we may yield an even wider WRE, that allows for increased (independent) revisions of our individual moral judgment. What is more, this new approach to WRE (with added nonmorals) is an ongoing process that is never considered to be complete, so it can be compared to natural sciences in this way at least – i.e. that the gaining of data and potential increasing of coherence and justification is a continuous process, and fluid one, where the 'data' inputs never really come to an end in the same way that observational data supporting some scientific theory doesn't simply end.

What adding nonmoral beliefs really means for the individual resulting moral

judgments, is that they can be narrowed down to an epistemologically preferable option – or be part of the 'preferable set' (Engel, 2012). This resulting WRE-with-nonmorals system would measure the belief system's coherence in three important stages – the overall coherence between: 1) considered moral judgements 2) results of reflective processes/revisions of our deeper/ 'core' moral beliefs, 3) results of reflective processes/revisions of our nonmoral beliefs. The third of these is the important added step that I, among others, am suggesting presently (Engel 2012) (Timmons, 1990). That is, of applying relevant nonmoral beliefs to the Rawlsian WRE reflective process in order to get an alternative version of WRE that can be considered more objective than classic WRE; Rawls' original version, at least enough to weaken the MSP.

In this chapter, following the previous assessment of Rawls' WRE, I'm arguing that a moral-only WRE is not objective enough to be epistemologically significant, because it excludes nonmoral beliefs. Including nonmoral makes the process of justification much more able to narrow down the potential multiple systems of belief sets. But in addition is also more descriptively accurate because, as examples below should demonstrate, we in fact *do* use a number of nonmoral beliefs to reach our moral conclusions, and they certainly appear relevant to the level of justification that we end up with.

Essentially, this chapter demonstrates the role of nonmoral, empirical beliefs within a WRE structure. I point out that the Rawls/Daniels version of WRE of chapter 2 leaves this important role out, and as a result suffers the strong-relativist issues that were initially raised by Richard Hare. I go onto argue that once the role of nonmoral beliefs is added to the WRE model, sufficiently take into account the role of immediate experience (non-doxastic assent), consider the theory as a coherence-justification rather than coherence-truth, that a coherence based WRE model that includes the role of nonmoral beliefs yet doesn't *reduce* the moral to them, can sufficiently justify moral believes and lead to moral progress. I discuss a selection of illustrative cases aimed at bringing

nonmoral facts into the justification of our moral beliefs - particularly taken from some more recent literature on this topic. I then conclude by discussing why the nonmoral may have been left out of Rawls's WRE, and then how a refined and a more inclusive WRE model can overcome these worries, with reference to the relevant case examples. The cases ultimately represent the essential justificatory involvement of nonmoral beliefs within a wider belief web *(fig. 3)*. The first of these examples is termed Modest Non-Doxastic Coherentism (MNC) initially coined by Engel Jr. from a 2012 paper, I then consider a well-known piece by Elizabeth Anscombe, *Mr. Truman's Degree,* and thirdly a discussion on abortion from Van Roojen's *Metaethics* book.

Modest Non-doxastic Coherentism

Mylan Engel Jr. puts forwards a coherentist approach as a system of moral justification that pays particular attention to the relevant nonmoral beliefs - "*facts*"-, but also experiences, as part of a wider ethical argument for vegetarianism. His paper on what he terms 'Modest Non-doxastic Coherentism' breaks down the method of justification he ultimately uses.

'**Modest**': This approach is termed 'modest' because it doesn't take a strict anti-foundationalist stance by completely denying the existence or the possibility of any basic beliefs, as many other coherentist approaches do. It is only anti-foundationalist in the sense that it doesn't require that *all* beliefs are justified in a linear basic-to-nonbasic fashion. There still are still basic beliefs, but crucially they don't rely upon any further *beliefs* for their support (which makes them possible justificatory-regress-stoppers), and they are also justified immediately. However, Engel's basic beliefs are not foundational because they don't carry a special epistemological privilege and so are still potentially revisable.

For example, my belief that there is currently something red before me is justified by my immediate sense experience of redness – a justification that is not positively supported by any other *beliefs*, but by an immediate experience. My immediate experience of the redness needs no further justification, it can act as a regress stopper – but yet it might still be revisable – it is not unquestionable or infallible: Suppose I had another belief in a memory that I had ingested hallucinogenic drugs that same day – I may then have reason to suspect and reject my immediate experience of redness. All that is required for a belief to be 'basic' is that it rests on no *other* beliefs for its justification and is justified immediately. It needn't be totally unrevisable.

Non-doxastic: Coherentist theories in epistemology are usually *doxastic* theories, meaning that the justification of beliefs is pinned upon the other things that one *believes*. The initial issue here though is that doxastic theories

pay no attention to experiences because they focus solely on our pre-existing and informed beliefs. However, our immediate perceptual experiences for instance, might not be belief forming themselves, but are still relevant to the justification of other moral beliefs.

Engel terms his version 'non-doxastic' simply because he does include the value of immediate perceptual experience from our senses, current mental states, "memorial seemings", and our immediate moral experiences in the justification of our moral beliefs. He just denies the strictly doxastic claim that the only thing doing justification work is the rest of one's existing *beliefs*.

Coherence: Finally, it is worth noting before discussing a case example argued by Engel, that his approach is a coherence theory of *justification*, not a coherence theory of truth. It argues that one moral option is better than another because it is *better justified* – not because, upon reflection, it is more likely to be true.

"MNC: If p coheres with S's beliefs and other internal states better than any proposition that competes with p, then S is justified in believing that p."

Engel presents a case for vegetarianism using a coherentist method of justification, crucially taking into account the coherence measures between both moral and nonmoral "*empirical*" beliefs. He maintains that virtually all individuals "*of normal moral sensibilities*" hold the following beliefs[2] (p.56):

Moral Beliefs

1. Other things being equal, a world with less pain and suffering <u>is better</u>

---

[2] I have moved one of the 'moral beliefs' from the Engel paper into the 'nonmoral' category: "Many nonhuman animals (certainly all vertebrates) are capable of feeling pain" – as this seem to be an empirical rather than moral belief.

than a world with more pain and suffering.

2. A world with less *unnecessary* suffering <u>is better</u> than a world with more *unnecessary* suffering.

3. Unnecessary cruelty is wrong and <u>ought not</u> to be supported or encouraged.

4. We <u>ought to</u> do what we reasonably can to avoid making the world a worse place.

5. A <u>morally good</u> person will take steps to make the world a better place and even stronger steps to avoid making the world a worse place.

6. Even a "minimally <u>decent person</u>" will take steps to help reduce the amount of unnecessary pain and suffering in the world, when she/he can do so with little effort on her/his part.

7. It is wrong to cause an animal to suffer for no good reason.

8. It is wrong and despicable to treat animals inhumanely for no good reason.

In addition to this collection of relevant moral beliefs, Engel highlights three nonmoral beliefs that are essential to the (ultimately moral) justification of the claim that eating meat is wrong:

<u>Nonmoral Beliefs</u>

1. In modern societies the consumption of meat is in no way necessary for
human survival or human flourishing.

2. Virtually all commercial animal agriculture, especially factory farming, causes animals intense pain and suffering and, thus, greatly increases the amount of pain and suffering in the world.

3. Many nonhuman animals (certainly all vertebrates) are capable of feeling pain.

Coherence argument

Nonmoral belief 3 with moral belief 7 and 8 are supposed to show that "*virtually all*" people believe animals can indeed experience extreme pain and suffering, and that this ought to be avoided where possible. Moral belief 1 & nonmoral 1 show that we believe less suffering is better for the world, and that animal farming obviously increases that suffering. The result of holding moral belief 1 and 2 with nonmoral 2 mean that the world would be *better* without animal farming. The moral beliefs 1, 2 and 4 mean that even "minimally decent" individuals ought to do what is possible, especially when it's of little effort to them, to avoid adding to the total amount of unnecessary pain and suffering in the world. Moral belief 3 then means animal farming should not be encouraged (i.e. by buying and consuming meat).

Therefore, to be consistent with both our deeply held moral beliefs, and our nonmoral beliefs or 'facts', we must conclude that eating meat is morally wrong.

Mr. Truman's Degree

Elizabeth Anscombe wrote an essay protesting the honorary degree due to be given to former US president Harry Truman by Oxford University for allegedly ending the second world war by using the two atomic weapons on Japan. Anscombe objects to the consequentialist line of thought that morally allows or obliges one to perform an evil act in order to gain a better or a 'good' consequence (such as achieving surrender).

The case described is of President Harry Truman signing the order to drop two atomic bombs onto the Japanese cities of Hiroshima and Nagasaki during the second world war. The aim; having the Japanese surrender unconditionally, effectively ending the war. But to achieve this consequence, Anscombe points out that many innocent civilian men, women and children were deliberately killed in an indiscriminate and painful way. Also, that the Japanese wanted to negotiate a peace agreement, and that an *unconditional* surrender was by no means the only way to end the war, or to cause Hitler's government to cease.

It seems the action 'ending the war' can equally be termed 'murdering the innocent' given the *known* potential victims, mean the same thing in every practical sense here. Secondly, the argument in favour of the bombing seems to justify, even *obligate* the killing of innocent people to achieve some other goal; using innocents as a means to some end or 'for the greater good'.

This is obviously of moral concern, as Anscombe highlights, "…*it cannot be right to kill the innocent as a means to any end whatsoever…"* (1958, ¶26). But crucially for my present questions, Anscombe clearly uses a combination of both moral and nonmoral beliefs throughout her argumentation, to justify the conclusion that dropping the atomic bombs on Japan – and honouring Harry Truman for this with a degree, is morally wrong and unjust.

Moral Beliefs
1. Choosing to kill the innocent as a means to one's ends is always murder, and murder is one of the worst of human actions

2. "The proposal of an unlimited objective in war is stupid and barbarous"

3. Those who are not fighting or supplying those fighting, are 'innocent' in war

4. To aim at *killing*, even when one is defending oneself, is murderous and wrong

5. By flattering, praising, and defending a bad action, one shares some of the guilt of that action

Nonmoral Beliefs

1. The objective 'making no peace with Hitler's government', did not necessarily entail the *unlimited* objective of 'unconditional Japanese surrender'

2. There was nonetheless a fixation by the Americans, on the *unconditional* surrender of the Japanese

3. There was a disregard of the Japanese wanting to *negotiate* a peace agreement

4. Two atomic weapons where dropped; one on Hiroshima and one on Nagasaki. The decision to use them was President Harry Truman's

5. In the case of Hiroshima and Nagasaki it was certainly decided to kill the innocent as a means to an end

6. A very large number of innocent people died, all at once, without warning, without the possibility of escape

Coherence argument

Nonmoral belief 5 together with moral belief 1 is intended to show that the atomic weapon attack on Japan amounted to murder, as nonmoral 5 states the 'fact' that the decision knowingly amounted to killing the innocent as a means to an end – while moral belief 1 expresses that type of act as murder. To back

this up, nonmoral 6 confirms that a large number of innocent people did in fact die as a result. Connected to that, the 'choosing' part of 'killing the innocent as a means to an end' is meant to be verified by nonmoral beliefs 1, 2 and 3 and moral belief 1 because the nonmoral 1 and 2 suggest that the Americans demanded an *unconditional* surrender while moral belief 1 entails that those kinds of demands are 'barbarous and stupid' during war, and the fact 3, that there appeared to be another option (besides the bombing) of negotiating a peace agreement with the Japanese, show that the final decision was in need a matter of *choosing* to kill, rather than being forced to in self-defence. Further to that, moral belief 4 would entail that even if the peace agreement option failed and action did have to be taken in self-defence, that the sole aim of killing, certainly of innocents, would still be morally wrong. Finally, moral belief 5, and nonmoral belief 4 suggest that it would be wrong to ward Harry Truman the honorary degree from Oxford, as it is a fact that the final decision to cause all of the above to take place was indeed Harry Truman's to make – and he did, while moral belief 5 entails that rewarding that action is to share some responsibly for that action. Therefore, as a matter of consistency and inferential connection, as well as applying a large range of historical facts, it is more coherent and so more justified to accept that it would be wrong to award Mr. Truman's degree.

Meena and Malcom

This third case is a discussion on the Meena and Malcom example in Van Roojen's text (2015, p.38-40). I will outline the Meena and Malcom illustration, and then discuss how the nonmoral beliefs play a role in the justification of both parties' moral beliefs.

Friends Meena and Malcom have a fundamental disagreement regarding the morality of abortion. Malcom is strongly pro-life; he believes that aborting a foetus is never permissible; it is always morally wrong. Meena on the other hand is pro-choice, and believes that it is sometimes morally right to undergo an abortion. Malcom therefore believes the process should be illegal, while Meena thinks it should be legal and funded.

Crucially though, Meena and Malcom both have the same nonmoral beliefs regarding abortion. From the conception stage, to the forming of the embryo, the timing of the development of vital organs, when nerves develop and when the foetus becomes a baby and would be physically able to survive outside of the mother's womb. In addition, both Meena and Malcom know the same statistics surrounding birth, and therefore know all the dangers to the mother, the risks involved in undergoing the abortion, pregnancy as a result of rape statistics etc. Both friends have the same facts, and yet they have a moral disagreement about abortion.

The important difference to notice from this example compared to the previous two, is that both Meena and Malcom use a combination of moral and nonmoral beliefs to support opposing moral claims. The example also states that they use the same set of nonmoral beliefs in their reasoning to support their claims.

This example may appear to be a case where adding the nonmoral beliefs to the consideration has not helped make the problem more tractable; on the contrary both have used the facts to different ends.

However, I think there are some issues with this particular example. On the altered version of WRE I am arguing for, the belief systems of Meena and Malcolm in Van Roojen's (2015) illustration for example, would never be thought of as 'complete' systems that have *all* the relevant empirical information available, but are then nevertheless different. Rather, in dilemma moral situations, we can progress the debate by essentially 'researching' the issue further, in a similar way methodologically speaking, to researching issues of scientific disagreement. That is, by collecting a *better* data sample to work with as we would in any other research scenario. We collect more of the a) felt-moral intuitions that are relevant to the situation (and relevant ultimately to the a-b-c reasoning process) - and crucially – we collect more relevant (d?) empirical facts – and because these two processes never really come to an obvious end, (we can always do more moral reflection, and we can always seek better relevant empirical facts), moral claims always have some further opportunity to be justified *over* some other rival claim.

Further, Meena and Malcolm do not have *all* of the relevant facts of abortion, nobody does. The relevant empirical facts are part of a young and ongoing process. This is a misunderstanding in assessing the example; the notion that both friends have all the relevant facts but disagree on the morality. It is rare in any area of research, particularly the nervous system and the brain in biology, to know anything close to 'all the facts available' – and certainly impossible to predict all of the ones possible to know.

Also, even if people in general knew more about the science surrounding abortion, the chances that both Meena and Malcom themselves could know *all* of those facts is difficult to imagine – and that is just the biology side of this, but there are also historical facts, social facts, endless statistics. To say they have the same nonmoral information is extremely unrealistic.

But even if for arguments sake we grant that both friends might have 'all' the facts available, if those facts are very limited and moral intuitions take more of

a role in the WRE process, then disagreement is still obviously possible. This may be why the example of abortion is a particularly difficult moral situation – because all the 'relevant facts' it turns out, have a lot to do with things like consciousness, the brain, neuroscience, psychology, etc. – which are all fairly new sciences and very much works in progress.

Meena and Malcolm may disagree because the empirical facts around abortion are murky, and so if there aren't enough decent relevant facts available, then too much weight is necessarily put upon the inter-moral reasoning central to Rawls style of WRE, and then divergence more easily occurs. This extra weight is necessarily put on to inter-moral reasoning in the sense that, maintaining a positon or even taking some actual action regarding abortion is often imperative, so 'suspending belief' isn't an option when pushed to claim one way or the other.

Even if Meena and Malcom have the same facts, all of the facts, *and* the same basic moral beliefs – there is still the nondoxastic aspect discussed in Mylan Engel's example, argued for above, to be taken into consideration. Even if all of the parties' moral and nonmoral beliefs match each other, they may very well have different experiences, different feelings, different things effecting them at the time they decide to diverge which all play a role in the reflective equilibrium process.

My argument against the validity of this example is not however, intended to imply a reductive naturalist account of morality. I am not claiming that the moral status of abortion is really nothing more than a difference in one's facts rather than one's moral reasoning. What I am claiming about this particular example is that: the idea that Meena and Malcom have the same level of coherence in their belief systems, given how likely it is those systems are to be different in at least some ways, and given that they have drastically different opinions on abortion, is almost impossible to actually imagine. It needs further illustration on the part of the person trying to show this.

Given that the two friends are likely to have different feelings, experiences, and moral beliefs – they are also very likely to have different degrees of coherence

in their systems as a whole – even if they do have all of the same facts. The burden of proof must be on Van Roojen's example to show how equally coherent, and otherwise identical systems, could produce an opposing moral claim and remain equally coherent. I would hold that when all the relevant facts are considered, the background theories and sure-intuitions are applied in WRE, that completely equal coherence between Meena and Malcom's systems of belief would be impossible. One must hang together better than the other.

The Is/ought Gap

One of the key reasons why the role of nonmoral beliefs is often minimised or eliminated altogether among theories in moral epistemology, is to avoid the so-termed 'is/ought gap', a distinction famously made by David Hume (1739), but utilised by many philosophers since as an argument against reductive naturalism in ethics, and against moral positivism in general.

Hume maintained that all knowable propositions come from either logical arguments that are in some way traced back to sense experiences (analytical knowledge), or from empirical beliefs derived from actual sense experiences; this system is known as Hume's fork. But using Hume's fork, statements about 'what ought to be' don't seem to fit the criteria for knowledge; we don't directly observe how things 'ought to be' by sense experiences of the world, and the notion of a logical connection between any empirical beliefs and moral conclusions is yet to be demonstrated. How one moves from the descriptive to the prescriptive must be explained.

Essentially on this view, talking about how things *are* doesn't tell us how things *should be,* to move to any moral conclusions we seem to require some further, broader moral beliefs. But then one can argue that any specific moral judgements (such as 'eating meat is wrong') seems to depend fundamentally on these broader moral beliefs (such as 'causing unnecessary pain and suffering is wrong'), and not on the empirical descriptive beliefs (such as 'animal agriculture causes unnecessary pain'). In short, Hume is/ought gap casts doubt on the possibility of any logical connections between nonmoral and moral beliefs, and if you accept his empiricism and the 'fork', then it amounts to moral skepticism.

Hume remarks that philosophers often do move from talking about what '*is*' to what '*ought to be*' without properly explaining the nature of that relation. Hume doubts the idea that there can be a logical connection between 'what is', and 'what ought to be'; the two categories of statements seem to Hume "*entirely different*" from each other – how one could form this relation would need to be

explained (Hume, Pt.1, §1, 1739). But Hume, without such an explanation, therefore denies the possibility of this kind of logical connection.

However, I would maintain that the role of nonmoral beliefs does not need to be excluded from WRE in order to avoid the consequences of Hume's is/ought gap; in short, I would argue that Hume's gap is only really a critical problem for the reductive naturalist – but not for the coherentist. The reason being that the reductive naturalist must show that all moral statements can be logically derived from *exclusively* nonmoral premises in order to show that claims about what ought to be, are really just claims about what is. What Hume's argument shows is that however many relevant empirical claims we have, we still always need to refer to some, usually quite general, moral beliefs in order to justify our specific moral judgments; we must have moral premises to justify the moral conclusion otherwise it's impossible to move from the 'is's' to the 'ought's'. I agree with the distinction only up to that point. But it goes too far to suggest that Hume shows that it is *only* the moral premises doing the justificatory legwork. I disagree that the gap argument demonstrates that it is impossible for nonmoral beliefs to have any logical connection with moral beliefs, the argument in fact only shows that it is impossible for moral beliefs to be justified by exclusively by the nonmoral. But very crux of my WRE model is that moral justification needs *both* coherent moral, and nonmoral, premises.

The challenge of the is/ought argument is therefore to show how moral and nonmoral claims are in fact logically connected. The logical connection between the moral and nonmoral beliefs in WRE is one of logical compatibility and my various illustrations above demonstrate this relationship. If is/ought distinction was correct, then no particular set of empirical beliefs would be either more or less compatible with moral judgments, because the two classes of belief are allegedly "entirely different", and a logical relationship impossible. But this claim is demonstrably false. Take the vegetarianism example:

The nonmoral beliefs:

4. Eating meat is in no way necessary for human survival or human flourishing

5. Animal agriculture, especially factory farming, causes animals intense pain and suffering and, thus, greatly increases the total amount of pain and suffering in the world

6. Many nonhuman animals (certainly all vertebrates) are capable of feeling pain

…with the general moral beliefs:

1. A world with less unnecessary pain and suffering <u>is better</u> than a world with more unnecessary pain and suffering.

2. Unnecessary pain and suffering is bad and <u>ought not</u> to be supported or encouraged.

…are meant to justify the conclusion: Eating meat is morally wrong.

If the is/ought distinction was correct, then the nonmoral beliefs here couldn't be either more or less logically compatible with the moral beliefs and the judgment that 'eating meat is wrong'. According to the consequence of Hume's gap, changing the nonmoral premises should be irrelevant to the level of justification here. But this is patently not the case. Suppose instead I had the nonmoral beliefs:

1. Eating meat *is* necessary for human survival and human flourishing

2. Animal agriculture, at least in the developed world, is usually humane and minimises pain and suffering

3. Nonhuman animals are not capable of human-like emotions, and cannot reflect on 'pain' in a human sense – and cannot experience a complex feeling 'suffering' as this involves intelligent reflection

…with the same general moral beliefs:

4. A world with less unnecessary pain and suffering is better than a world with more unnecessary pain and suffering

5. Unnecessary pain and suffering is bad and ought not to be supported or encouraged

The conclusion that 'eating meat is wrong' is not obvious in this case as it was in the first case. In fact, the opposite claim that 'eating meat is morally permissible' would be more compatible with this particular set of beliefs – but the only thing that changed was the nonmoral, not the moral. To further exemplify this point, take the discussion of the use of nuclear weapons: I absolutely need the facts about the large-scale and uncontrollably destructive consequences of using nuclear weapons, *as well as* moral beliefs such as 'killing the innocent as a means to an end is always murder and murder is always wrong', in order to reach the moral conclusion that using the two nuclear weapons on Japan in the second world war was wrong.

This demonstrates that there *is* a relationship between nonmoral and moral beliefs when it comes to moral justification. Just as my discussion of moral 'deep theory' and Rawls's WRE in chapter 2 shows that swapping-out the moral background theories changes the moral justification, the above example demonstrates that swapping-out the set of nonmoral beliefs also changes the moral justification. Both particular moral background theories *and* a particular set of nonmoral beliefs are needed for the increased level of justification achieved.

Hume is right that it is impossible to move from talking (*exclusively*) about what is to what ought to be – because then the nonmoral would be basically foundational and one would be lead to a reductive naturalist position (or at least be very open to it). But this does not entail that it is impossible to move from talking about 'what is, and what ought to be' to 'what else ought to be'.

My version of WRE not only permits, but is fundamentally committed to nonmoral and background moral beliefs working together, with equal epistemic priority, to justify specific moral judgments. Moreover, I don't believe that Hume's concern actually rules this out at all.

Geoffrey Sayre-McCord also reflects this same idea in *Coherentist Epistemology and Moral Theory* (1999): Sayre-McCord argues that whatever the nature of the 'is/ought gap', it does not completely insulate one's moral beliefs from nonmoral beliefs. Further metaphysical, epistemological, social, and psychological considerations may all be relevant to the justification of our moral beliefs (p.150-151).

"…*moral coherentism doesn't need to maintain that nonmoral beliefs alone either entail or in some other way inferentially support moral conclusions; they may well hold that our moral views themselves establish the epistemic relevance of nonmoral considerations*" (p.151).

As I stated above, this means that the coherentist model can accept all the standard arguments for the 'is/ought gap' without being committed to arguing that all the evidence we have for our moral views must come from only other moral considerations. The nonmoral can still have a justificatory role, just not an exclusive one. Given just how implausible it is to see any of our moral views as having some special epistemic privilege, a great attraction of this version of WRE is its ability to make sense of our moral views being justified even in the face of the 'is/ought gap' (Sayre-McCord, p.151).

In sum, my version of WRE gets past the is/ought issue by two important methods: Firstly, it demonstrates that there must be a compatibility

relationship between moral and nonmoral beliefs, otherwise any set of nonmoral beliefs would be just as compatible with our general 'background' moral beliefs as any other set – which is shown to be an impossible position. Secondly, by not requiring moral justification to be exclusively nonmoral in content, the process does not need to move directly from talking about 'what is to what ought to be' as Hume worries, but moves from 'what is and what we believe very generally ought to be, to what ought to be'.

Increasing coherence

Finally, it is crucial to mention, in light of the three case examples discussed, the way in which coherence between beliefs, moral and nonmoral, can be increased and how we can assess competing belief sets for their coherence, to see which is justified over the other.

One obvious way is to simply reduce the number of contradicting beliefs. So for example, among the beliefs:

1. I need to eat meat to get enough protein and therefore to survive
2. In modern societies the consumption of meat is in no way necessary for human survival or human flourishing
3. I am a human

There is an obvious contradiction. Given the well-supported belief 3, that I am a human, I cannot simultaneously hold 1 and 2, on pain of a blatant contradiction. One of the first two must be rejected or altered. Removing contradictions within the beliefs amounts to what is often referred to as *minimal* coherence – it means that no belief in a set tells against another. (McCord, p.187) (BonJour, L.,1985).

However minimal coherence alone is not sufficient for the justification of moral beliefs, because a set of beliefs, while not contradicting one another, may also be completely unconnected, as this very simplified example shows:

1. I am a human
2. Camels can't breathe underwater
3. Lying is morally wrong
4. Charity is good

5. Eating meat is morally wrong

This set of beliefs clearly meets the criterion of minimal coherence, because none of the beliefs tell against the others. But the beliefs are completely unrelated and unconnected. No one belief gives any positive support for any of the others, they just don't contradict each other. It cannot be a fully coherent justification for the belief that 'eating meat is morally wrong' simply on the basis that nothing tells against it (p.187). To increase the level of coherence there have to be some positively connected beliefs (BonJour, L.,1985) , such as:

1. Eating meat is not necessary for human survival or flourishing
2. Pain and suffering is bad
3. Meat farming greatly increases pain and suffering

4. <u>Eating meat is morally wrong</u>

Thirdly, and most importantly for my purposes of showing how adding nonmoral beliefs to WRE increases the coherence, in addition to belief sets being connected and minimally coherent to increase coherence they must be increasingly comprehensive (BonJour, L.,1985) . This part is crucial to highlighting the importance of nonmoral beliefs, but also will provide the foundations for a response to the multiple systems problem discussed further in the next chapter. To make a set of beliefs more comprehensive you have to *add more* beliefs to the set that are non-contradictory, and ideally also provide positive support. But the key to increasing comprehensiveness is by adding more beliefs that fit. Consider two examples of justification for the same moral belief:

a)
    1. My friend Jane had an abortion
    2. I like my sister, she is generally a good person
    3. Women have a moral right to decide what to do with their bodies

<u>4. Abortion is morally permissible</u>

b)

1. My friend Jane had an abortion

2. I like my friend Jane, she is generally a good person

3. Women have a moral right to decide what to do with their bodies

4. Having the right to an abortion is important for gender equality

5. Banning abortions increases the prevalence of illegal abortionists

6. Illegal abortions are extremely dangerous

7. Many victims of sexual assault require abortion rights

8. Human foetuses have not entered human personhood

9. Even the 'right to life' for fully developed humans does not include the right to use somebody else's body to stay alive (I can't demand my mother gives her kidneys)

10. Up to at least 28 weeks, human foetuses cannot experience pain

<u>11. Abortion is morally permissible</u>

Belief set b) is clearly a much more coherent set than set a). Both are minimally coherent, but b) is far more comprehensive; it contains many more beliefs that fit with the conclusion that abortion is sometimes permissible

Discussion

The above cases show in a practical sense, how the involvement of our nonmoral 'empirical' beliefs are essential to the justification of our moral beliefs within a coherentist structure. It should be obvious from the above examples of vegetarianism, nuclear war, and abortion – that our nonmoral considerations play a vitally determining role in what moral beliefs we end up being able to justify over the other alternatives.

However, what is also clear from the above examples is that the arguments discussed are one's that are still very contentious and where moral opinions clearly remain divided strongly on both sides of the arguments. There remain plenty of arguments both for and against vegetarianism, possessing or using nuclear weapons, and on abortion rights.

But the key argument I am suggesting here, as highlighted particularly by the abortion case example, is that the justification process does not simply come to some end – *especially* when one considers the role of the relevant nonmoral 'facts'. The acquisition and application of our nonmoral beliefs to moral justification is a continuum, an ever on-going process that is itself constantly up for revision.

Pragmatically speaking, perfect consistency between the relevant beliefs remains as much of an ideal as the notion of 'having all of the relevant nonmoral facts' to begin with – It is not as if this ideal is ever *actually* achieved with the moral disagreements still remaining. What is important is that, where our moral beliefs are in present disagreement, as with the abortion and vegetarianism arguments, we usually have some idea how to go about getting more and better nonmoral facts and to apply then them to our moral considerations.

For example, Meena and Malcom may both know all of the *current* relevant biological, social, and historical 'facts' surrounding abortion (although realistically this is obviously almost impossible) and still be in disagreement over the morality of abortion, but surely both Meena and Malcom should accept that having a *better* set of facts regarding, for example, foetal development, foetal awareness, pain, all the millions of historical and legal cases around abortion and women's rights, pregnancy, human health, relevant statistical information etc. would leave either of them in a *better* position to justify their views and could possibly lead them to revise their views in some future circumstance. Moreover, this 'better set of facts' can always be a larger, more informed/evidenced, and more comprehensive set of facts than it previously was – so a constant revision and reflection of these is worth doing, it *better* justifies the moral beliefs at hand.

What makes the justification still very much coherentist in nature though, is that the relevant facts don't carry a foundationalist-like epistemic privilege over our moral considerations. If this were the case, then we would have a reductive account of morality in which we essentially reduce all our moral beliefs to nonmoral beliefs. In that case the moral beliefs would simply rest upon what set of facts one held and would only vary according to those facts – the route that Norman Daniels is so keen to avoid (and thus strongly plays down the role of the nonmoral in Rawlsian WRE) (Daniels, 1996). But the nonmoral beliefs are not supposed to have this privileged role – on the contrary, our nonmoral beliefs - particularly our scientific beliefs – are just as 'up for revision' as the moral considerations are; in light of new evidence. Of course there are deeper held, more certain, empirical beliefs just as there are deeper held and more certain moral beliefs, but nothing remains totally unrevisable and foundational, and crucially our moral as well as our nonmoral beliefs are justified exclusively by what else it is that we believe (or are experiencing – see Engel's MNC) at the time.

Chapter 4

Objections and Replies

I have so far suggested an alternative version of WRE that adds the role of nonmoral beliefs and experiences. The core idea here being that by adding nonmoral beliefs to the process and crucially not giving any class of belief, moral or nonmoral, epistemic priority, that we can yield a more objective and more accurate moral justification than we could with the original Rawlsian model. I have also pointed to the role of nondoxastic assent to show how we can get justification from moral intuition, from immediate experience, and crucially from permissively justified beliefs. There are still a variety of concerns and potential criticisms involved in introducing the role of nonmoral beliefs and of nondoxastic assent to beliefs. This section explores a series of objections and replies concerning four key areas of potential criticism: The issue of infinite regress and how a coherentist model can stop this while avoiding both circularity and foundationalism; arguments against the nondoxastic aspect and permissive justification; problems with measuring/increasing the level of coherence (particularly by increasing the size of the belief system); and finally the notion that the altered WRE model could still be rigged to fit our intuitions, and still remain open to a multiple systems problem. I use a cascading objection-reply format here for each area of concern.

1. Infinite Regress

Objection:
Moral coherentism, and coherentist theories of justification in general, have no way of stopping an infinite regress of justification. Foundationalists can more easily stop the regress with their foundational beliefs – because these are beliefs

that don't depend on any other beliefs for their justification; the regress stops at those basic, foundational beliefs. But coherentists, in denying any epistemologically privileged set of beliefs, have no way of stopping the regress without resorting to circular reasoning which has no justificatory force at all. (Therefore moral coherentism must be false).

Reply:

The regress argument in epistemology essentially traces and questions the line of justification of any given belief: for a belief to be justified, it must either be inferred from another belief, or the buck stops with a belief that doesn't require anymore justification. Intuitively it would seem that the beliefs we hold are mostly inferred from other beliefs we have. But the obvious question remains of 'what justifies *that* belief?' And then 'what belief justifies the one that justified the first one?' etc. It would appear this leaves two options: either the claim that the chain eventually stops with beliefs that require no further justification (in the moral case this might be, for example, strong moral intuitions such as Ross's prima facie duties that we 'just know' by our moral sense, are justified). Or, we are left an option of circular reasoning where the chain eventually links back to the belief one began with. Either way, the regress must be stopped somehow, or else the chain of justification just goes on and on ad infinitum. Initially therefore, the debate around regress seems to favour foundationalism because they have foundational beliefs within their arguments that can act to stop the regress. The coherentist must therefore either accept and defend, or reject the charge of circular reasoning – if coherentism rejects the possibility of acceptable circular reasoning then some beliefs must be able act to stop the regress.

The coherentist has only two options at this stage: either to show that there are some non-inferentially justified beliefs that can act to stop the infinite regress, but that these are not infallible or self-evident in the way that a foundational belief would be. Or else coherentism must defend some form of circular reasoning as being epistemologically preferable. I will explore further the first

option here, that there are some beliefs that might stop the regress, but they are not foundational.

The first option is one that is also argued by Sayre-McCord: that there are indeed some non-inferentially justified beliefs that can act to stop the justification regress, but, these do not carry any special 'epistemic privilege' in the sense that they are absolutely certain or infallible, or unquestionable (1996) – which is what is claimed about actual foundationalist beliefs. The non-inferred beliefs on Sayre-McCord's account can be altered or even rejected entirely in the face of other strongly-held (yet inferred from other beliefs) beliefs that we hold. This amounts to a kind of compatibilism between foundationalism and coherentism because the beliefs in question can be viewed as foundational only in the sense that many other beliefs rely upon them and that their support is something beyond simply 'other beliefs' – but, they are not unquestionable or unrevisable because they still need to cohere with what we already believe in order to be accepted, and if they don't, they're rejected – even though they are not inferred from what we already believe in the first place, which is the crucial distinction here. These non-inferred beliefs therefore don't hold epistemic privilege above our inferred beliefs, because the inferred beliefs we have, can in some cases cause us to reject them. What this particular position does is attempt to bridge a gap, or rather equate, non-skeptical foundationalism with non-circular coherentism.

It is important to explicitly note the type of coherentist approach that this is not: This position all but explicitly rejects the possibility of any acceptable circular reasoning, which is decidedly different from many coherentist approaches that only emphasise the 'hanging together' of a coherent system of beliefs as being central to the justification. This overall 'hanging together' coherence is of course necessary on the Sayre-McCord model, but is not

sufficient to get past the regress. Sayre-McCord's version instead appeals to foundational-like beliefs, that can still be rejected by our other beliefs, but don't come *from them* – so they can act to support the inferred beliefs while avoiding circularity, but reject epistemic privilege. This position is able to answer a variety of initial concerns with a positive coherentist approach to moral questions, particularly on the issue of regress.

The first concern with this answer might be that there simply are no foundational-like beliefs that exist, in terms of non-inferred yet justified beliefs. That is the classic skeptical view that there are simply no plausible candidates for the type of beliefs apparently needed to hold-up the rest of the belief system (Sayre McCord, p.154). Where the skeptic would agree here is that for our beliefs to be justified, they must eventually link back to some either foundational or foundational-like beliefs that are themselves justified – but not by any other inferred beliefs; they must have a justification independent of other beliefs. Where the skeptic disagrees here is simply that no such beliefs exist or are available to us; that all of our beliefs are justified by other beliefs, which leads to circular reasoning, which removes any possible real justificatory force.

However, a possible response to this concern, that there simply are no foundational-like/not-inferred-from-beliefs beliefs, is that the bar set by the skeptic for such beliefs to become justified is far too high. For if the foundational-like beliefs must be certain and infallible, then everything beyond pure mathematical conclusions seem to be ruled out as unjustified. But the criterion is misunderstood: the kinds of belief that can stop the regress don't need to come with some guarantee of absolute certainty – they simply need to have that quality of being *not*-inferred from other beliefs. They might however, be inferred from something else, say, an immediate sensory experience or a strongly felt intuition about some immediate circumstance. Furthermore, in terms of just stopping the regress, foundational-like beliefs don't really need to be anywhere close to certain because they can be traded-off for better beliefs in different circumstances or contexts (especially given that external circumstances do often change). What is important, at least just for denying an

infinite regress within any coherentist system, is not infallibility, but that whatever justifies the foundational-like beliefs is not an inference made from *another belief.*

But what then, if not an inference from other beliefs, could give support? Take immediate sensory experience as an example of this alternative support: I might hold a belief that 'there is a red postbox in front of me' (A). Now there is a variety of supporting evidence that comes with this belief A, but a main one of course is that *'I can see* a red postbox' with my eyes, right there in front of me (B). My belief that 'there is a red post box in front of me' doesn't initially require any further positive support from any other beliefs – it is justified simply because I can see it. I don't seem to need any other beliefs to evidence my 'seeing a red postbox'. Yet this doesn't automatically entail that my belief (A) is by any means certain or infallible. It could still be false and ultimately rejected: I might for example, hold beliefs that force me to reject my immediate experience (B) – like I may believe that I took hallucinogenic drugs just beforehand, or I might be in a country where post boxes are strictly always green and never red – plus the light is at an unusual angle, or I might have a belief that I am and always have been red-green colour blind. The point is that just the seeing/experiencing a red postbox presently (B), seems to sufficiently justify my belief (A) – provided that I have no reason (in the form of counter-beliefs) to doubt that current experience. But the (B) experience itself is not a belief, it just causes one, which is precisely what is needed to stop the regress.

Furthermore, I don't even need to have positive beliefs about immediate experiences, say, in my 'sense experiences being usually accurate' (for whatever reason that might be). All that is crucial is that I need to *not* have any beliefs that call those experiences into doubt.

This example highlights the difference between permissive and positive justification of beliefs: the former is one that is central in Sayre-McCord's argument. Permissive being justification we can have because there is no good reason not to hold the beliefs involved, and positive being a justification due to

positive reasons to hold the beliefs; other beliefs or experiences that lead to the inferences in question.

One worry at this point though, is that even if we could come up with some beliefs that are non-inferentially justified and supporting of other inferentially justified beliefs, these would be disappointingly few, and would be nowhere near sufficient to justify our whole system of moral beliefs. The concern to consider is that while we might be able to permissively justify *a few* of our more basic moral beliefs - that these only come to a uselessly small amount and do nothing of significant justificatory force beyond perhaps justifying some very obvious or agreeable moral beliefs. The tiny amount that might be justified on this model will be nowhere near enough to support the rest of our moral belief system (REF). It certainly isn't the sort of theory that's going to help get past any clashes or MSPs in ethics.

In response to this though, one could comfortably maintain that the skeptic (above) is just being too restrictive about what should and shouldn't count as justification. If we are liberal enough with what non-doxastic aspects, and permissively justified beliefs we allow to count, and what inferences we allow to count, then we will have enough to justify the whole system (or at least most of our moral beliefs).

Further to this, one can argue that any system that *doesn't* justify the bulk of our moral thinking must just be wrong in how it is allowing the justification of beliefs (Sayre-McCord, 1996). At the very least, the burden here would be on the skeptic to show that we have moral beliefs that we are certain of, and yet can't be traced back to either inferences from other beliefs, permissively justified beliefs, or to some form of non-doxastic justification. Just pointing to the particular areas of moral disagreement that we do have doesn't work as a criticism of the whole justification system, because this, one might argue, could be corrected by sourcing better 'data'.

This criticism can only really gain traction if the sceptic position could show that lots of our belief justification ends at inferred beliefs as opposed to anything non-inferred.

Another criticism of this path to consider here is that once one makes the distinction between 'x belief being justified' and 'an individual being justified in holding x belief' – then even if we could come up with what was a small number of non-inferentially justified beliefs, they still wouldn't ever get any individual past the issue of regress.

This part of the third criticism concerns the basing requirement and a particular interpretation of it. On this account it demands *doxastic* ascent – that for a person to be justified in holding a belief satisfying the basing requirement that they must have some other beliefs about that belief – such as for the seeing red experience, that the belief was formed under appropriate circumstances/light conditions etc. (or for justice in Rawls' WRE, knowing that there was self-bias removed etc.) Doxastic ascent is to have beliefs about beliefs.

One can respond by maintaining that a person doesn't need to have beliefs purely supporting the notion that her other beliefs are justified: my beliefs X are justified because belief Y. All that matters is that those beliefs are in fact justified, based on the available evidence that a person has. But they don't need to have beliefs about their own belief forming – as long as her beliefs are sensitive to the evidence that they actually do have – which might be non-belief evidence such as an immediate experience (instead of perhaps dogmatic or superstitious beliefs). On this account, perhaps worryingly for the multiple systems issue, beliefs can still be false yet justified if based on some evidence or reason.

Another area one could explore is denying the necessity for doxastic ascent, that is, that we must have beliefs about our beliefs for justification. We might not actually have beliefs about our beliefs and yet still be justified in believing them because they're based in available evidence rather than some dubious

source. And the 'evidence', needn't be in the form of 'beliefs-about-beliefs'; we might employ immediate experiences, strong intuitions, or moral feelings for example.

In response, one could argue that a belief is still only justified if there is some reason, or evidence, that we can appeal to; we have to have a reason to think it's justified. These don't need to be belief-supporting-belief reasons, and they don't have to be infallible, but *some* evidence must be available. A potential problem for Sayre-McCord using only permissively justified beliefs, is that a belief unsupported by other beliefs, or some immediate reason, —whatever is supposed to act as the 'evidence' *"—will be one believed for no reason and so will be unjustified"* (Sayre-McCord, p.157). This criticism uses an assumption from the original regress argument – that for a belief to be justified one must have some reason to think that it is.

There are two routes for justifying our beliefs at this point of the argument: Claim that some beliefs are positively justified by something other than 'other beliefs', perhaps feelings, apprehensions, immediate awareness of something etc. Or, to reject the initial assumption in the regress argument – that a belief is only justified if we have some positive reason to believe it.

Sayre-McCord's paper basically argues for the second option here and rejects the first. The first option is rejected because the reasons that people have for believing x must be *available to them* – But here Sayre maintains that reasons only can count as being appropriately available when they become the contents of beliefs. On this view, the reasons that a person has for believing only enter awareness when that person is in a particular cognitive state (as in when they *believe*). If I don't believe something, then it can't count (to me) as a reason for anything else, in this case, as a reason to justify other beliefs. This is an internalist response in that the justification depends on an *internal* mental state (of believing appropriately). The opposite, externalist route is therefore portrayed as problematic because, it's argued, it maintains either that I can have no reason to believe something justifiably, or that I can have an unavailable reason to justifiably believe something.

But does the rejection of externalism actually work in this case? Could it not be that I have a reason to believe something, that is 'available' to me, yet is just not 'a belief'? For example, I might use my immediate sense of 'seeing red' as an *available* reason to believe that I am colour blind, because I have some other deeper held belief that what I am looking at must not *actually* be red.

I think that one could argue that this type of sense experience falls under the bracket of,

"something so like a belief, for instance, "an awareness that . .", as not to be worth distinguishing from belief in this context" (Sayre-McCord, p.156-157).

But of course the externalist here can claim that the awareness or the immediate experience is not the contents of 'a belief' as such, perhaps due to the immediacy of sense experiences and one not necessarily explicitly acknowledging that they 'believe'. It may in fact be a distinction that is worth making.

In any case, the Sayre-McCord goes for the second option by rejecting that a belief is only justified if we have some positive reason to believe it. Whether the second option of rejecting the necessity of positive ascent is one to the foundationalist entirely depends upon an important distinction between positive and permissive justification. Positive is when I am justified because I have a reason to believe X. Permissive is when I am justified because I have no good reason *not to* believe X.

Therefore, the distinction made between positive and permissive justification allows for one to take the second option of the regress: rejecting the idea that a belief can only be justified with a reason to believe it. It can be permissively justified by having no reason not to believe it.

If we accept that some foundational-*like* beliefs can be permissively justified rather than positively, then we only regress back to those (and not ad

infinitum). On this account, the regress stops there because we don't infer the permissively justified beliefs from any other beliefs - they're not really inferred at all - but rather allowed to function on the basis of an absence of certain counter-beliefs. At the same time, the possibility of those certain counter-beliefs entering our beliefs system at a different time is certainly not ruled out - so the foundational beliefs here are not infallible, but importantly (for stopping the regress) not inferred from other beliefs.

This argument for the relevance of permissively justified beliefs gets over the three key difficulties mentioned earlier in the following ways: Firstly, it stops the infinite regress because the line of justification stops at the permissively justified beliefs - there is a regress but it comes to an appropriate end.

  Secondly, the permissively justified believes don't have to have the troublesome qualities of infallibility and certainty - they can be traded-off and rejected in light of new 'evidence'. Thirdly, the reasons a person has for accepting the permissively justified beliefs are not unavailable to them because the acceptance depends on the other things that they believe.

The discussion around permissively justified beliefs is hugely relevant to WRE because the vast majority of our nonmoral (certainly non-normative) beliefs are either directly or indirectly and closely linked to, our sense experiences and our observations. And, the evil demon arguments or the brains-in-vats argument aside, the reason we believe the information we get from our senses and our collective observations is precisely that, for the most part, we have no good reason to doubt them; they are permissively justified. Not certain, not infallible, but importantly are not dependant on any other beliefs for their justification.

2. Against the Nondoxastic aspect

Objection:

One initial criticism of the MNC arguments for my version of WRE, comes from Alan Vincellette (1997, p.3): that convergence doesn't make the conclusions of MNC true. That even *if* the system can narrow down the dilemmas and disputes to one epistemologically preferable option – that option might still just be false. The permissively justified beliefs that support the system while avoiding stopping the infinite regress, could just be wrong.

Individuals might simply be wrong about what it is they agree on. To ground our beliefs, we need a criterion, which would "employ something more basic and fundamental but it's the moral theories that are precisely what we disagree on." (Engel, 2012, p.63)

Reply:

In response, firstly and most importantly, it was not intended to show that the products of MNC are true because it's not a theory of truth but a theory of justification – and this same response is used in defence of my WRE. The MNC theory is aimed at those who accept the regress stoppers, both moral and nonmoral, that are taken as permissively justified.

Of course, Engel wants to argue that 'virtually all people' hold those beliefs, and perhaps this is mistaken or fails to take into account other mitigating beliefs. Engel wanted to show by way of consistency that vegetarianism is most consistent with beliefs 'most people' already hold. There can be plenty of debate around his specific example, but what is important here is the justificatory structure he's using, and not the contents of the example Engel happened to use.

If we don't hold the beliefs listed, then of course the conclusion won't be justified. MNC for Engels for example, shows that we are "rationally committed" to vegetarianism, on pain of contradiction, *if* and only if we hold the listed moral and nonmoral beliefs in the example.

Grounding justification in consistency with certain beliefs is arguably a more pragmatic system of justification, than grounding it in moral theories which are more disagreeable. MNC doesn't *certainly* ground the claims, but neither do moral theories. At least we will either feel more certain about the listed beliefs that support the Engel's argument for vegetarianism, or else we will come up with a counter argument that employs a different and more coherent set of beliefs. Either way, we will be more certain about that than we are about, say, whether act utilitarianism or Kant's categorical imperative is a better moral theory (Engel, p.65).

A possible response to the last point echoes Richard Hare's rigging concerns: Vincellette would probably want to say that the fact that we do feel more certain about coherence arguments than moral theories is irrelevant really. We could simply be wrong about the beliefs we render coherent, just as we might be mistaken about moral theories. Saying that Engel's argument for vegetarianism is just aimed at people who accept (and apply) the relevant beliefs in the same way as Engel does, is akin to saying that act utilitarianism is only aimed at people with core utilitarian principles above all others.

I would argue though, that the fact that we feel more certain and are in more agreement with a given coherent set of beliefs, supported by some permissively justified intuitions and empirical beliefs, than we are with most other 'grounding' moral theories is exactly the point. It doesn't certainly solve the multiple systems problem, but it narrows it down to those who accept a certain et of background beliefs. Those may not be the ones that Engel does actually use in favour of giving up meat eating, but the structure is important here: Taking into account counter beliefs we might have, and then working back and

forth to increase coherence (as one does in Rawlsian WRE) is how we increase the coherence, and improve the system of justification.

3. <u>Measuring and Increasing Coherence</u>

Objection:
What does it mean for a belief system to be more coherent than a rival one? The concept of coherence, and measuring it in order to better justify our beliefs is just "*hopelessly vague*" (Douven, p.406). In the case of moral disagreements, and therefore opposing belief sets, there is no obvious way of claiming that one set (or a whole system) is more coherent than any other.

Reply:
There are several ways in which to increase the coherence of a belief system, therefore being able to distinguish between multiple systems as to which is the most coherent. The most telling three measures are, the logical consistency of the system, that is, the one with the least number of contradictory beliefs; the positive connections in a system, meaning the one with the most inter-supporting beliefs; the largest system, providing at that at least the consistency rule is met first.

Consistency

The first measure of logical consistency is the most straightforward, and perhaps obvious way of measuring the coherence between a system of beliefs. This first basic rule has been referred to as 'minimal coherence', evidential consistency, and logical consistency in the literature (Sayre McCord, p.166) (BonJour, p.95). It is essentially the requirement that no belief contradicts or tells directly against another: a system that holds beliefs P & ~P is therefore logically inconsistent and therefore less coherent than a system without such a contradiction.

Two brief examples show obvious logical contradictions:

a)
   1. Tom is taller than George
   2. George is taller than Sam
   3. Sam is taller than Tom

b)
   1. Stealing is always wrong
   2. Some people have more resources than they need, some not enough
   3. It is good to steal from the rich and give to the poor

However, it is crucial to point out at this stage that logical consistency is nowhere near enough to get a fully coherent system of beliefs. This is a two-fold criticism often levelled at coherentists; that firstly, we can have a perfectly logically consistent system of beliefs that is not in any *'appreciable'* or interesting way coherent: Say that I hold, among other things, the belief that there is an man called Oliver eating a ham sandwich in a galaxy 4000 lightyears away from earth – This belief is logically consistent with what I already believe in that I have no beliefs that tell against it, and adding that belief causes no contradiction – yet clearly it is hardly a coherent belief to maintain just on this basis. Secondly is the issue of probabilistic inconsistency: the idea that I may have a set of beliefs that is logically consistent, while being highly improbable given my other beliefs:

1. I believe I will win the national lottery on Friday
2. It is extremely improbable that I will ever win a national lottery

These beliefs are perfectly logically consistent, yet there is an obvious probabilistic inconsistency. This is why it is crucial that the rule of consistency in coherentism should not be taken as the only criteria, and far from it. It is simply a bare minimum for coherence: I cannot have a fully coherent system that contains contradictions, but I can have an incoherent system that contains none. Consistency remains a rule only for bare minimum coherence (Sayre McCord, p.166) (BonJour, p.95) *

Positive connections within the system

As discussed in the third chapter, having more positive connections with inferentially supporting links to other beliefs increases the coherence within that system. Take one of the examples given earlier on:
1. Pain is bad, and should be avoided beyond necessity
2. Animal farming causes pain
3. Eating meat encourages animal farming
4. Eating meat is not necessary or even helpful for survival in developed societies

5. Eating meat is wrong and should be avoided where possible

As well as being logically consistent, in that none in the set tell against the others, this set is also positively connected because beliefs 1-4 suggest belief 5; they make it more likely than the opposite, 'eating meat is good and should be encouraged where possible'. Therefore, increasing the inferential connections within a system of beliefs evidentially makes the system more coherent. Notice that this is just the reverse of probabilistic inconsistency mentioned above – if it makes a system less coherent to hold 'that P' and 'it is improbable that P',

then it stands to reason 'it is probable that P' and 'that P' is increasingly coherent.

Size of the system

The most relevant to the topics discussed in this paper, adding nonmoral beliefs to WRE, is the argument that coherence is increased by the size of the system (presuming it already meets the fist or both of the above rules). Assuming we have logical consistency, and perhaps inferential connections, the more beliefs that we add to the system that 'fit in' or 'hang together' well with what we already hold, will increase the level of coherence – this is expressed explicitly in BonJour's account:

 "3. The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the **number** and strength of such connections." (BonJour, p.98)

This is also discussed in the Sayre McCord paper as increasing the level of "*comprehensiveness*".

Objection:

The idea of increasing the size, the comprehensiveness, of a system if beliefs to increase degree of coherence raises a significant problem. It is clear that as the system grows larger in size, the probability that there are going to be more inferential connections discovered will *necessarily* increase simply because there are more possible connections that *could be* made within the system. One would therefore expect the size of a belief system and the number of connections to correlate: the larger the system, the larger the degree of coherence - being just down to its "sheer size" (Stanford, 2017).

But on the surface this seems highly problematic because it would entail that simply the more belief I add to my system, the better it will be. But it is not the case that larger systems of belief are better - I may hold 100 times as many various moral and nonmoral beliefs as you do; but they may all be highly dubious! I may have just sourced my beliefs badly somehow, been subject to particular and unusual circumstances that gave me some bizarre but vast system of beliefs - Consider that I may have a highly warped moral sense (due to a terrible upbringing or decades of indoctrination for example) yet I may to a great deal of moral thinking and reflection, so may end up with more intuitions, more connections to facts I belief and therefore a much larger system. Yet I am still unlikely to have a highly coherent and highly justified set of beliefs and principles.

Reply:
One possible response, noted explicitly in the Stanford encyclopaedic account of this point, is that we might measure the *density* of inferential connections rather than the sheer number of this connections, explained as follows:

"*the inferential density of the system…is obtained by dividing the number of inferential connections by the number of beliefs in the system*" (Stanford.com, 2017)

This is the idea that it is not just adding more of any beliefs that increase the coherence, but adding ones that have positive inferential connections - this is meant to get over the complaint the one could keep adding random beliefs to the system, simply to get the numbers up to increase the chances of getting more positive connections. By dividing the total number by the number of actual connections reflects better the actual level of coherence we have.

However, another, and I think an even stronger, and more obvious response to the issue given above is an answer that pays more attention to the first 'minimal condition' of coherence that we gave; the consistency aspect. By just

blindly increasing the sheer size of the system by adding any sets of beliefs that are numerous and only give *more* beliefs, while increasing the chances of getting more inferential connections and therefore positive support, we would surely also increase the chances of getting contradictions too. Just adding more of *any* beliefs increases the chance of contradictions just as much as it increases the chance of connections. This is not what is implied by the condition of comprehensiveness. There has to be a reason, namely one of reliability, to add certain types of belief to the system in the first place – not just to add beliefs in randomly.

Consider the model that I'm arguing for in this paper, the main aspect that is intended to improve the WRE to make it a viable coherence theory is the adding of nonmoral beliefs to the system. But the adding of nonmoral beliefs have to be, in a sense, inter-coherent between each other as well as with the other aspects of WRE like the sure-intuitions and moral 'deep-theories' (CMJs and background theories). One can't increase the coherence of moral beliefs by simply adding *any* possible nonmoral beliefs (in the hope that some might just happen to add inferential support) because this approach is very likely to create a number of inconsistencies *between* all of the nonmoral beliefs that I hold. What is crucial, and obvious with nonmoral beliefs, that that they are properly based in evidence or reliable testimony: if I have a larger set of nonmoral beliefs that is logically consistent and inferentially connected, then this means I have done (or learnt of) more observations and gained more empirical information that is relevant and that I have a good reason to believe. If I only have a larger set that is unconnected and inconsistent, which is the aspect that the above criticism takes issue with, then it will not increase the coherence.

Moreover, the general point here is that you need all three of the points that increase coherence; consistency, connectedness, and size. And in that order of priority. They have to be understood together. If we just increase the size, having no reliability on the method we use to provide that increase (i.e. the reliability of learning facts by observing and applying scientific method vs. simply adding potential nonmoral beliefs), then we will not have a more coherent system than one with less.

WRE with added nonmoral beliefs adds and emphasises the role of those nonmoral beliefs because we have established and reliable means of acquiring them – we can use our observations of the world, and increase the weight of our nonmoral beliefs by doing more relevant observation. Moreover, unlike the moral intuitions, moral principles, and background theories – that are limited in how many we can apply that are relevant to one individual moral claim, we can always do more to get better nonmoral data, and more nonmoral data to apply to the problem. Once again, coming back to the Meena and Malcom example from Van Roojen's text (2015), If Meena and Malcom have the same facts and same background theories about abortion yet they still disagree about it while having equal justification for their views, either person can go and acquire more facts that are relevant to try and make the problem more tractable or to tip the argument in their favour. If Meena is pro-choice and Malcom is pro-life, but they have the same fact base, Meena can go and get a better fact-base than Malcom's, or reflect more so as to have more relevant intuitions than Malcom does: It is always therefore possible in theory for one person to be able to better justify their argument than the other; by getting *more* inferentially connected beliefs.

4. <u>WRE still rigged? (elaborate intuitionism?)</u>

Objection:

The WRE model must either be rigged to fit intuitions, or reduced to the 'facts', if we add the role of nonmoral beliefs. If facts have too strong a role, in that, given they are established by observational evidence, then these 'facts'

have the epistemic privilege the coherentist is trying to avoid. But on the other hand, if the coherentist can reject nonmoral information in light of certain intuitions, then our subjective moral feelings rule again and we have form of intuitionism in disguise.

To illustrate, if the Meena and Malcom abortion dilemma could potentially be solved by seeing which belief system is more coherent with the facts, then it all still ultimately depends on those facts. But then facts have the privilege, this means we are left with a version of reductive naturalism, because we are prioritising the nonmoral over the moral by default (if only in cases of moral disagreement or dilemma).

Reply: The nonmoral category does not get epistemic privilege within the system. They can be more core to the whole system, or more peripheral to the whole system, and indeed it is probably the case that established nonmoral beliefs often tend to be more core to the system given their link with observation, calculable probabilities etc. But, in coherentist spirit, nothing -not even they- are immune from revision and constraint from the other parts of the WRE system.

Take the Mark and Julie example. Mark and Julie are brother and sister, and they decide to sleep together. However, they agree to take every possible contraceptive precaution making it impossible for the encounter to result in a pregnancy. They agree that it will only happen once, and that they will never tell anybody about it (so as not to propagate the behaviour) (Haidt, 2001, p.1024). For arguments sake, one can alter Haidt's thought-experiment to cover more bases: Either Mark, or Julie are infertile – making it completely impossible for a pregnancy to occur, to cover for even the small chance that contraception might leave. They only 'agree' to do it once and not tell, but they might easily renege on this agreement, it may become habit and/or they will encourage others and spread the behaviour – what's to stop this? Perhaps Mark and Julie are in possession of two memory-erasing tablets that cannot fail to work. They will enjoy the encounter in the moment, then entirely forget it,

making it impossible for them to propagate or normalise their behaviour at all – also freeing them from the risk of encouraging a repeat.

At this point, the reductive naturalist would be stuck with morally permitting the situation, because if morality can just be reduced the 'facts of the matter' then all of the above 'facts of the matter' would provide no argument against Mark and Julie's sexual encounter. Whatever fact-based concerns one has with Mark and Julie, they can always be countered with over-riding opposing facts. Yet we still obviously have the overwhelmingly strong intuition that Mark and Julie still ought not to have sex – despite the facts. But to the reductive naturalist, this intuition is ineffective.

However, the coherentist can avoid this dilemma by *not* giving the nonmoral category epistemological privilege over anything else. While the nonmoral beliefs in the WRE do play a hard-to-alter role in the justification process, the justification is not reduced to them, and nothing remains unrevisable within the system. Therefore, the overwhelmingly *strong* intuition that Mark and Julie ought not to have sex whatever the precautions are, can indeed override the empirical beliefs that counter it. We can reject the 'argument from the facts', based on the argument from the intuition, because this particular intuition is just so heavily weighted that it, in this case, trumps the facts.

Objection:
But, if intuitions that happen to be strong enough can still potentially override the nonmoral information, the coherentist argument risks becoming an intuitionist once again. Because unlike nonmoral beliefs, background theories, or moral principles, the CMJs (sure-felt moral intuitions) can have exponential weight. There are certainly some limits to how weighted background theories and moral principles can be, and even nonmoral, evidence based information can only have so much force in the WRE process – in that there seems to be some limit to how much *more* evidence for a 'fact' makes a difference to their weight within our moral reasoning: For example, the nonmoral belief that a

'foetus cannot feel pain' can only be strengthened so far by relevant evidence, because once the evidence in support of a fact is overwhelming – adding yet even more doesn't make much of a difference to how we regard it.

The same doesn't seem to be the case for our moral intuitions however, even Rawlsian CMJs. The reason is because these are still *felt* intuitions, and unlike the effective limit that adding more and more evidence can have on the weight of our nonmoral beliefs, there is no obvious limit for how strongly we can *feel* something is either wrong or right. And how strongly we feel something, is the only way that we can weigh the CMJs in the first place. Referring this back to Haidt's example, it's just that we feel so strongly against incest that the moral intuition can override the argument from the facts – but it is still just a feeling *of ours*, however strong. What if a feeling, emotion or attitude (that's actually the result of one's particular circumstance) is a more sinister one than merely the feeling that Mark and Julie ought not to have sex? Letting the moral intuitions have any role in potentially overriding facts puts the process at dangerous risk of some peculiarly bad intuition having too powerful a role. But this is now what the coherentist has committed to accepting, on pain of denying the moral intuitions a role in potentially rejecting nonmoral beliefs – having to permit Mark and Julie's encounter, and thus effectively accepting reductive naturalism.

Reply:
It is fundamentally mistaken to think that intuitions always have the potential to be more strongly felt than relevant facts. They may do at times, and that is a fundamental part of coherence justification; to deny epistemological privilege to any one category. Moral intuitions have a proportionate role in almost all of our moral decision making. The reason it may seem as though the they always have the potential for a stronger role within the system is just that they are harder to quantify than the supports that exist for nonmoral beliefs: It's relatively straight-forward to note and quantify the evidence that exists on animal farming statistics, or foetal development cycles, or historical facts about the second world war, for example. But it's much harder to quantify how *much*

I feel about abortion, or to what degree of strength my attitude and feeling about eating meat is.

Chapter 5

Conclusions

This paper set out to question the coherentist approach to moral epistemology, specifically how well moral beliefs may or may not hope to be justified within such a system. I used John Rawls's wide reflective equilibrium model as an initial template, but I found there to be some serious issues with the justificatory structure of WRE, the core problem was pinned down to the fact that the only inputs for Rawlsian WRE were the CMJs, or rather, strongly-felt moral intuitions – and even after consideration of the responses to this, the moral principles and background theories still seemed 'rigged' to fit the CMJs. The challenge therefore, was to see if there was any other class of belief, besides the CMJs, that could ground the justification and provide some genuinely objective and independent constraint. My solution was the additional consideration of nonmoral beliefs within the WRE process. I argued that, for various reasons the important role of the nonmoral was left out of the original WRE model, but, that if those reasons for avoiding the nonmoral could be overcome, that their role was absolutely essential to WRE being able to provide positive moral justification. I demonstrated the role of, and the relationship between, the moral and nonmoral beliefs largely by exploring and discussing a selection of case illustrations; the vegetarian argument in Engels modest non doxastic coherentism paper, the use of atomic weapons in Anscombe's *Mr. Truman's degree,* and the abortion discussion in Van Roojen's text, among others. I then proceeded to engage with some of the problems encountered by introducing nonmoral beliefs into WRE, and attempted to provide some answers to these concerns in the previous chapter. Ultimately, I maintain that by introducing nonmoral beliefs to the coherentist justificatory model of WRE, that one can more independently and objectively check and constrain the role of moral intuitions. I therefore maintain that a coherent balance between nonmoral beliefs, moral intuitions, general moral theory, and

moral principles can lead to a genuine progress of moral justification. Chapters 2 and 3 of my paper form the main argument that I make, with the 4ᵗʰ Chapter considering the relevant responses and replies to the argument made.

***

My second chapter functioned mainly to actually explain the process of WRE, to point out the initial concern (Hare), explain the standard response to this concern of an independence constraint (Daniels) (Dworkin), and the reason why I argue that a problem still remains to be solved within WRE.

I began the chapter by explaining the WRE structure and how it functions within Rawls' work, the core idea being that our moral judgements and principles become more justified when individual intuition-based 'CMJs' are brought into a coherent balance with both general moral principles and crucially, our relevant moral background theories.

The initial and most obvious concern with this standard version of WRE was a charge of circularity and a resulting strong Multiple Systems Problem. It first seems that the three key 'levels of justification' in the system are far too dependent on the CMJs, as these are the only actual input into the system; the worry being therefore that if we just happen to have very skewed or biased/influenced strong moral intuitions (CMJs) then the other two parts of the system will be built around cohering with those, and any moral judgments and principles we end up 'justifying' will only be so relative to our particular individual intuitions; the 'elaborate intuitionism' of my title.

I found that the standard response to this concern was that the Rawlsian WRE structure is just not understood properly. That there is in fact an independent check, via the background theories, on the initial moral intuitions that stop WRE from merely systemising those initial intuitions – therefore giving the model real justificatory force, rather than amounting to 'rigged' intuitionism-in-disguise. As discussed, this 'independence constraint' is not independent of the

moral beliefs of that individual, rather the two sets of intuitions that support either the moral principles or the background theories must be disjoint, distinct and independent of each other. The form of independence here can be misunderstood as meaning 'from an external independent source' rather than 'independent of each other' in this case.

However, as explored towards the end of my second chapter, the concerns of intuitionism and of rigging still remained, even after considering the 'independence constraint' as described by Norman Daniels (1996). The condition for the two sets of intuitions supporting the different parts of the system (principles and background theories) simply being independent *of each other* proves to be a red herring – because both sets, however independent from each other, are still not independent of the individual person's intuitions; the fact that they might be 'disjoint' is irrelevant. If all of one's moral intuitions happen to be influenced by their class, gender, ethnicity or some other accidental circumstance, then what I can justify with them will always be relative to my own subjective experience. There therefore remains a substantial problem with using moral intuitions to regulate other moral intuitions, however disjoint they happen to be.

The end of the second chapter demonstrates this remaining problem with WRE by showing how equally acceptable yet different moral background theories, such as a deep theory of fairness that takes market dessert into account rather than just equality, can be employed into WRE by appealing to alternative moral intuitions that other individuals might happen to have. Moreover, the possibility of this difference in deeper intuitions didn't appear to be a far-fetched suggestion, it only takes someone to have slightly different moral intuitions to John Rawls's own to diverge from the key things the structure was intended to justify (namely, Rawls's two principles). By showing how one can plausibly switch the 'relevant deep theories' from the background theories in WRE to yield different justifications, I show that there is a string multiple systems problem – even considering the independence constraint.

Additionally, by showing how different deep-theories and broad moral beliefs can be considered within the WRE structure, aside from revealing a MSP problem, I also illustrate how WRE can be considered quite separately from Rawls's contained arguments for his 'principles of justice'.

The upshot of chapter 2 is that using one's 'deeper held' moral beliefs to regulate the other moral beliefs in WRE, is still very dependent on the particular intuitions one just happens to have. Given that it's accepted that various accidental factors strongly influence those moral intuitions, this version of WRE is able to justify practically any moral belief system – provided the right intuitions are there to back it up, which leads to a strong MSP (and therefore subjectivism and/or strong moral relativism)

My third chapter began by re-iterating the problem that WRE was still left with by the end of chapter 2; that using some intuitive moral beliefs to regulate others doesn't actually add anything interesting in the way of positive justificatory force, because it can be used to justify practically anything given the corresponding moral intuitions, hence a strong MSP still remains (even when the standard responses such as the independence constraint are considered). At best, Rawlsian WRE *might* be able to rule out some very obviously bad moral judgements, at least for most people (given that some background theories are very widely acceptable) – but it certainly struggles to provide a positive justification for anything else besides the very obvious.

After concreting the core issue with Rawlsian WRE, I moved on to arguing my proposed solution to the problem; adding a key role for non-moral beliefs in the WRE process. My argument is essentially that the one aspect that seems to be left out of WRE is a role for our non-moral or 'empirical' beliefs; beliefs we hold in virtue of the way the external world is. As chapter 2 concluded, the whole reason why the independence constraint ultimately failed is because the

only thing regulating moral intuitions were other moral intuitions – but this WRE model left out nonmoral beliefs entirely.

I suggested that once the relevant nonmoral beliefs are considered and brought into the equilibrium process with the other parts of WRE, that this drastically weakens the effect of the MSP, providing an opportunity to positively progress moral reasoning by weighing up the level of coherence between potential rival belief sets. Further, the independent-check on our moral intuitions on this model, is genuinely independent of all of our moral intuitions which therefore negates the charge of 'rigging' or elaborate intuitionism.

I first discussed the potential role of non-doxastic assent in WRE – that is, how immediate moral experiences, present mental states, and perceptions – as well as our informed beliefs, can have an important role in WRE. This preamble was important as it sets the scope for what can actually count towards justifications.

I then moved on to discussing a variety of cases from various different texts, to show how the nonmoral is absolutely essential to a coherentist moral justification. Mylan Engel's argument for vegetarianism and therefore why eating meat is morally wrong, directly and purposely appeals to both the moral (a world with less pain and suffering is better than a world with more pain and suffering) and the nonmoral (animal agriculture causes intense pain and suffering) for its moral judgement, and holds that the conclusion is most coherent one. Elizabeth Anscombe's famous argument against Harry Truman's honorary degree again crucially appeals to both the moral (choosing to kill the innocent as a means to an end is wrong because it is murder) and the nonmoral (the atomic bombs dropped on Japan under Truman's orders knowingly killed many innocent people as a means to an end).

Further on, the abortion case from Van Roojen's Meena and Malcolm example, as well as detailing the role of the nonmoral in moral reasoning in that case, also highlighted the problem of empirical underdetermination which is a potential worry for the version of WRE I am arguing for. I argued a two-

part answer to that worry: given that I am still implying a constructivist rather than realist account of moral justification, the fact that the MSP is not completely solved when some facts are introduced, firstly does not mean that all the facts have been considered, even if all the known ones have been (which is highly unlikely in practice) and secondly – does not mean that the MSP can't at the very least be significantly weakened by narrowing down the possible 'multiple systems' available by introducing the relevant facts. The first part of that response was stronger than the second as it suggests that the consideration of relevant facts can never really be an exhausted process in the way that considering only moral beliefs can be. The second part of the answer points to a worst-case for my WRE model, that rather than completely getting past the MSP, it only minimises the possible 'multiple'.

Although both Rawls and my own versions of WRE are both prescriptive, in the sense that they are models for how we should/can do better moral reasoning rather than descriptions of how we actually do make moral judgements - the discussion of the different applied ethics cases I think shows that besides providing better prescriptive moral justification, my version also provides a more accurate descriptive account as it seems obvious that the specific nonmoral beliefs discussed were essential to the justifications in all of the cases explored.

Finally, after considering the benefits of including the nonmoral in WRE, I discuss the reason why this might not have been included by most other commentators in the first place and how those further problems might be overcome. Two key reasons for avoiding the nonmoral emerged here: Firstly, the prevalence of Hume's 'is/ought gap', that is, the argument that one can't logically deduce an 'ought' statement from an 'is' statement. Secondly, the risk of providing the nonmoral beliefs with epistemic priority and therefore becoming a reductive naturalist model rather than a coherentist one.

I have argued that my WRE model can avoid Hume's famous problem where reductive naturalism cannot. I did not suggest that moral justification happens by making logical deductions from 'is' beliefs to 'oughts' – because in that case

any moral beliefs would play no substantive role at all – once you had the correct facts, 'morals' would just follow as a property of those facts. On my account though, you still need both is's and ought's for the system to function, what I reject is that one can be reduced to the other. The nonmoral beliefs play a justificatory role in how coherently they 'hang together' with the potential moral belief-options we have. For instance, take the atomic weapons example: the conclusion that 'dropping atomic bombs on japan was wrong' was not logically deduced from the nonmoral belief that the action 'killed many innocent people' – the additional moral belief that 'killing the innocent as a means to an end is wrong' was vital, but my point is that the nonmoral belief was also vital. The moral and nonmoral aspects have to sit together coherently – and if the nonmoral changes, then so does the level of justification: if the action 'killed mainly military targets' or 'killed nobody', then my conclusion that the action was morally wrong quite clearly becomes less justified. The is/ought gap is only a problem if one tries to reduce one to the other, it does not demonstrate that nonmoral beliefs are irrelevant to moral justification. What is important, and what my altered WRE model satisfies, is the condition that both the moral and the nonmoral play a substantive, but non-prioritised, role in the justification.

Leading on from this, it was therefore important to consider how exactly a coherence relationship might exist between moral and nonmoral beliefs. Towards the end of chapter 3 I discussed 'the three Cs' idea from Sayre-McCord's paper as conditions for measuring the coherence of the two types of belief; consistency, connectedness, and comprehensiveness. Particularly, of the three, I found that adding nonmoral beliefs would very clearly increase coherence by comprehensiveness; by increasing the size – Adding nonmoral beliefs to the moral consideration greatly increases the total amount of beliefs being considered, and provided they are consistent with the moral beliefs, they increase the potential total amount of overall coherence to be achieved. The other side of this point is that adding nonmoral beliefs and increasing the total amount considered, increases the risk that one might in fact come across inconsistencies and have to then revise the moral beliefs. But this is precisely

why we have a better justification when we *don't* find inconsistencies even when the nonmoral beliefs are added.

The charge of reductive naturalism, which appears to be Daniels's main reason for avoiding the nonmoral role, is unwarranted because my version of WRE isn't intended to be reduced to or entirely dependent upon the nonmoral, and it does not get a special priority over the moral beliefs. Despite the role as an independence constraint, the nonmoral beliefs can be rejected or revised in light of strong moral arguments, just as the moral 'deep theories' can be – which is impossible on a reductive naturalist account. My WRE model therefore retains the coherentist essential of not granting priority to any one class of beliefs. This is most clearly demonstrated with the incest hypothetical discussed in chapter 4, where the moral intuitions feel far stronger than the nonmoral fact-set that would otherwise permit the scenario on a reductive account.

If both the is/ought gap and the worry of reductionism are successfully avoided, then the case illustrations in chapter 3 show precisely how nonmoral beliefs fit into a coherence-based justificatory framework. And given the discussion on coherence measures, the chapter shows how adding this nonmoral role can increase the amount of coherence, and how this ultimately leads to a better moral justification than one which entirely excludes the role of the nonmoral beliefs.

<div align="center">***</div>

In my fourth chapter I attempted to tackle five of the stronger problems with WRE and of coherentist theories in general. Problems of: infinite regress, of using non doxastic assent, with measuring coherence, remaining MSP, and finally the penultimate question of whether the system is still 'rigged'.

The issue of an infinite regress of justification is the notion that unless the system becomes circular, the chain of belief justification won't come to an end. Unless one can accept the circularity option, 'regress stoppers' must be found.

The key for the coherentist is to find something that can stop the regress, while at the same time does not become foundational. My answer to this problem was found by distilling and reapplying some of Sayre-McCord's arguments regarding permissively justified beliefs. The idea here was that some of our beliefs within WRE can be justified by nondoxastic assent: through immediate experience, strong intuitions, self-evidence, because we have no reason not to believe them… essentially anything that is *not* another belief. The reason this option works well with my model is that empirical, nonmoral, beliefs mainly come through observation and ultimately our (or somebody's) senses – and this input is belief forming, but not a belief itself, and arguably not dependent upon any other beliefs for justification. Equally, the strong intuitions that we apply in WRE in the form of CMJs are initially justified, or given weight, because of a strong feeling we have – and a strong feeling is not a belief, it is a perception of sorts that seems to some degree self-evident. Therefore, CMJs and empirical beliefs are candidates for regress stoppers as they don't necessarily depend upon other beliefs for their initial justification. But crucially, for the coherentist aspect, my 'regress stoppers' are not infallible, unrevisable, certain, or fixed in anyway – on the contrary, they can can force revisions or rejections on each other. For this reason, and for the reason that I have two classes of regress stoppers in my WRE (CMJs *and* empirical beliefs), the model does not become a foundationalist one. The regress stopping beliefs can still be traded away or rejected because of their incoherence, and other beliefs can lend extra support to them, but they are not *dependent* upon beliefs for their initial justification, so they can stop the regress while avoiding becoming foundational.

Linking to the above response to the issue of infinite regress, was the concern of using any nondoxastic assent within a coherentist theory in the first place. This was Vincellette's complaint that the beliefs we use to cohere with, and justify the other beliefs in the system, could just be wrong. My response essentially accepts this point, because I'm providing a coherence-justification theory, not a coherence-truth theory. It may well be the case that the permissively justified beliefs in my WRE system such as strongly felt intuitions (CMJs), or empirical beliefs derived from our sense experiences, could indeed

just be wrong. That is the nature of permissively justified beliefs, they are permitted because we have a reason to assent to them and no immediate reason to reject them – but they are not concrete, as I discussed above, they can be revised or rejected. And the chances that the 'wrong' ones do eventually get rejected in light of new beliefs, is increased by the expansion of the whole system; that is, by doing more moral reflection, and acquiring more or better empirical beliefs.

<div align="center">***</div>

The core worry I still have with my altered WRE model, really takes shape at the end of my fourth chapter. The chapter ends with what is essentially the ultimate question of this paper, is WRE (even after reconsidering the role of the nonmoral) still 'rigged'? If so, then we perhaps have only a greatly elaborated intuitionist account of moral justification – if not, then the notion of moral progression through having better justified beliefs has some hope on the coherence based WRE model. The largest remaining concern that I had with my altered version of WRE, was regarding the weighting of the CMJs that are used in the process. It is a fundamental aspect of my particular model that there can be trading-off, constraining, causes for revision, and rejection between both the moral beliefs and nonmoral beliefs in the process. This is how I avoid granting any class of belief epistemic priority, therefore avoiding my model becoming a foundationalist or reductive naturalist one. It is integral to my argument logic that both moral and nonmoral classes can have provisionally equal roles.

But in the process of achieving 'equilibrium' - a coherent balance - it is obviously required that the relevant beliefs are weighted in some way, so as to decide what is being supported and what is being questioned throughout the process. A better or more extensive set of facts may cause me to alter my initial intuitive moral beliefs (i.e. Engel's vegetarian argument, Ch. 3), or a strong intuition may overcome my available fact set (i.e. incest hypothetical, Ch.3). The problem here is that the only thing we possess to weight the intuitions is in form of a feeling or emotion; my intuition informs me by my feeling good,

or comfortable, or somehow drawn towards one thing – or conversely, repelled, uncomfortable, or even angry/upset about some other thing. But this method doesn't seem to have any scale or comprehensible degrees – I might feel infinitely confident or positive about *x,* or infinitely repelled and put-off by *x.* There doesn't seem to be any limit to how one weights the intuition-based aspect, and moreover the way we *do* weight our intuitions is most often down to particular circumstances and other accidental features. This is especially problematic for my theory because of the comparison with how one weights nonmoral beliefs (which are so crucial to my version). Nonmoral beliefs are, predominantly, weighted by the empirical evidence we have for them – but in practice this is obviously always going to be a finite amount of support: for example, although I might be able to infinitely collect more facts surrounding, say abortion, a) I can't ever actually do this in practice – my fact base is always practically finite, and b) there is a limit on the effect that adding more evidence for nonmoral beliefs actually has. To expand on the second point about adding more evidence: Once I get to a level I might call 'beyond reasonable doubt' about some nonmoral belief, the adding of further supporting evidence has little or no effect on how I would weight it as part of a wider reflection about something; once I know beyond reasonable doubt that the heart and brain of a foetus starts to develop between 4 and 6 weeks in all historical human cases observed, observing one more case, or indeed one hundred more cases that confirm this doesn't add any substantial (and certainly not proportional) weight to that belief. It is only in a strictly theoretical sense that this might technically add weight, but *I* won't count it any differently in WRE. Therefore, for both practical and essentially probabilistic reasons, there seem to be limits to how we weight the nonmoral beliefs in WRE, where there seem to be no limits or intelligible degrees on how we weight the relevant moral beliefs. In short, the worry is that our moral intuitions will too often have the upper-hand over the facts, because we can give them infinite weight in our reflection, simply by 'feeling' more strongly – the facts require more to gain weight, and have a definite limit on that weight.

All that being said, perhaps in future work, I would explore a response along the lines of ideal WRE remaining just that – an ideal. Perhaps there are more

sophisticated ways we can give degrees to our intuitions, and perhaps feeling too strongly about anything is just an error that we must stoically overcome. Moreover, perhaps there can be infinite support added to nonmoral beliefs by gaining evermore evidence for them, and the fact that '*I* won't count it any differently in WRE' is again, just error or lack of capacity in some way. Still, until more is done to argue this, there does remain some concern with my altered, with added nonmorals, WRE.

<div align="center">***</div>

The relationship between facts and morality is interesting to say the least. I have argued in this paper for a coherentist justification of moral beliefs, that takes into account the nonmoral as well as the moral intuitions for the support or revision of our moral judgements; it essentially amounts to a version of Rawls's WRE with an extra role for the nonmoral beliefs. In a wider context, this model is most interesting because of the gaps it intends to bridge: The foundationalism vs. coherentism gap is perhaps bridged by the (almost quasi-foundational) permissive justification theory that I used to stop the regress, rather than taking the route of justifying circularity (as some classic coherentist theories do). The gap between intuitionist ethics and deontological ethics is somewhat bridged by taking into account both moral intuitions, and principles that are informed be deeper/broader moral theories. And finally, Hume's gap between facts and values, the is's and the oughts, is bridged by showing how both types of belief need to hang in a coherent balance in order to have positive moral justification.

<u>Bibliography</u>

- Anscombe, G. E. M. (1956). *Mr. Truman's degree.* Oxonian Press.

- BonJour, L. (1985). *The structure of empirical knowledge.* Harvard University Press

- Brandt, R. (1990). The Science of Man and Wide Reflective Equilibrium. *Ethics*, *100*(2), 259-278.

- Daniels N. (1989) *Reading Rawls,* California, Stanford University Press

- Daniels, N. (1996). *Justice and justification.* Cambridge [England]: Cambridge University Press.

- Daniels, N., (1979), 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy*, 76(5): 256–82

- Dworkin R. in Daniels N. (1989) *Reading Rawls,* California, Stanford University Press

- Engel Jr, M. (2012). Coherentism and the Epistemic Justification of Moral Beliefs: A Case Study in How to Do Practical Ethics Without Appeal to a Moral Theory. *The Southern Journal of Philosophy*, *50*(1), 50-74.

- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, *108*(4), 814.

- Hare R. in Daniels N. (1989) *Reading Rawls,* California, Stanford University Press

- Lyons D. in Daniels N. (1989) *Reading Rawls,* California, Stanford University Press

- Rawls, J (1951) *Outline of a Decision Procedure for Ethics,* The Philosophical Review, Vol. 60, No. 2 (Apr., 1951), pp. 177-197

- Rawls, J (1974) *Reply to Alexander and Musgrave,* The Quarterly Journal of Economics, Vol. 88, No. 4 (Nov., 1974), pp. 633-655

- Rawls, J, (1975)"*A Kantian Conception of Equality,*" Cambridge Review 94–9 - Reprinted as "*A Well-Ordered Society,*" in Philosophy, Politics, and Society, Vol.5. Oxford, Blackwell, (1979) pp. 6–20

- Rawls, J. (1999). *A theory of justice.* Cambridge, Mass.: Belknap Press of Harvard University Press.

- Rawls, J. (2009). *A theory of justice.* Harvard university press.

- Sayre-McCord, G. (1996). Coherentist epistemology and moral theory.

- Scanlon, T. M. (1973). Rawls' theory of justice. *University of Pennsylvania Law Review*, *121*(5), 1020-1069.

- Stanford.org, *Coherence Justificaiton.* (2011). *Stanford Encyclopedia of Philosophy.* Retrieved 3 june 2017, from http://plato.stanford.edu

- Stanford.org, *Reflective Equilibrium.* (2011). *Stanford Encyclopedia of Philosophy.* Retrieved 1 May 2016, from http://plato.stanford.edu/entries/reflective-equilibrium/#CriRefEqu

- Timmons, M. (1991). On the epistemic status of CMJs. *The Southern Journal of Philosophy*, *29*(S1), 97-129.

- Van Roojen, M. (2015). *Metaethics: a contemporary introduction.* Routledge.
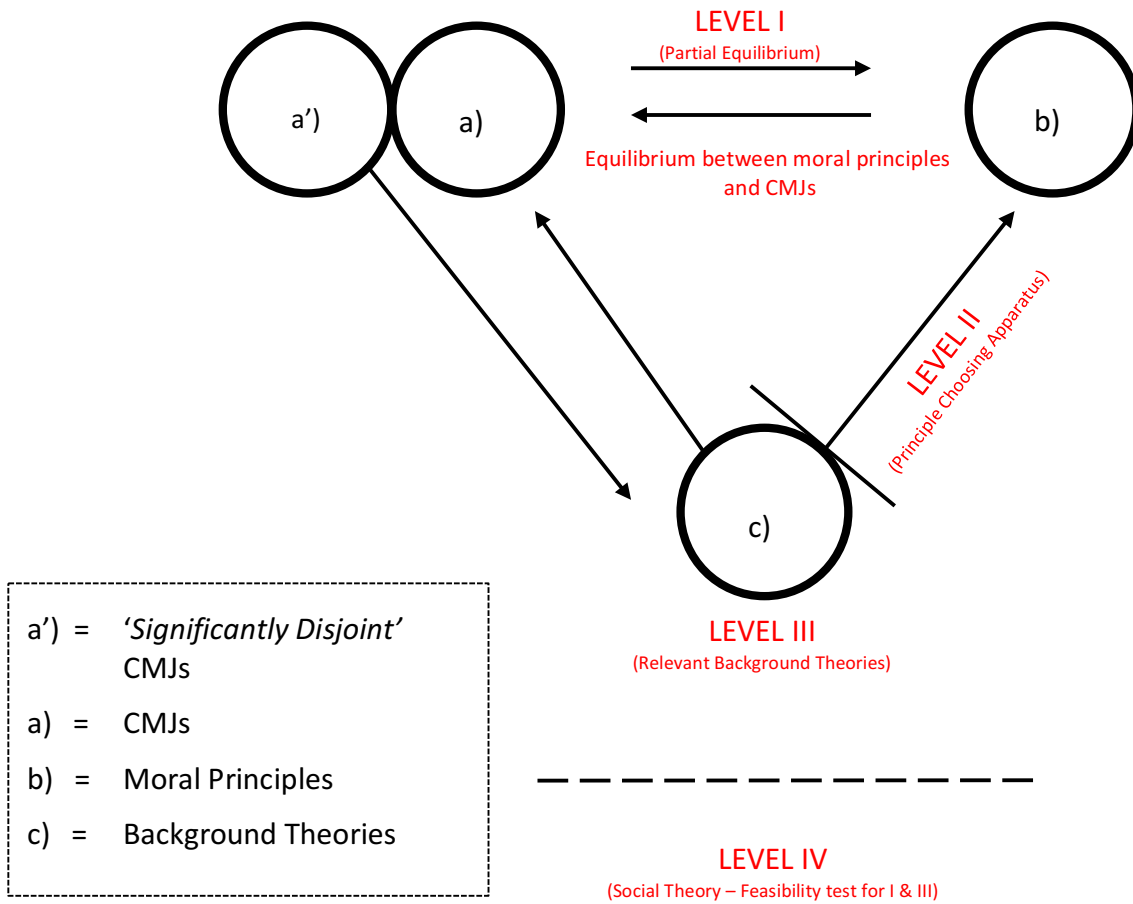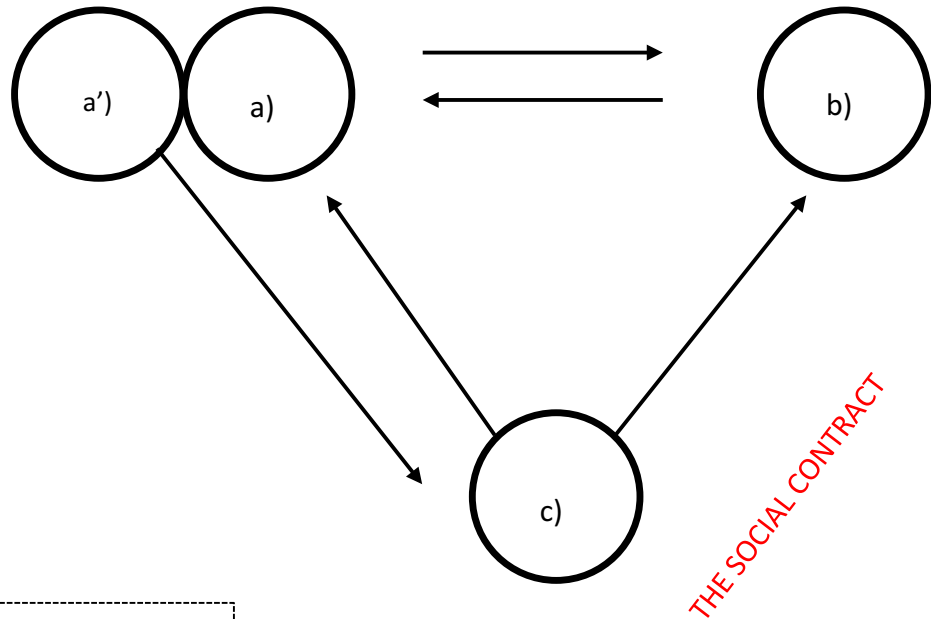
Figure 1

LEVEL I
(Partial Equilibrium)

Equilibrium between moral principles and CMJs

a')

a)

b)

LEVEL II
(Principle Choosing Apparatus)

c)

LEVEL III
(Relevant Background Theories)

a') = *'Significantly Disjoint'* CMJs

a) = CMJs

b) = Moral Principles

c) = Background Theories

LEVEL IV
(Social Theory – Feasibility test for I & III)

Figure 2

a') = *'Significantly Disjoint'* CMJs

a) = CMJs

b) = Moral Principles

c) = Background Theories

THE SOCIAL CONTRACT

Theories of the person – The role of morality in society – Well-ordered society ETC.

Figure 3

LEVEL I
(Partial Equilibrium)

CMJs

a)

a')

b)

d)

c)

RELEVENT MORAL
BACKGROUND THEORIES
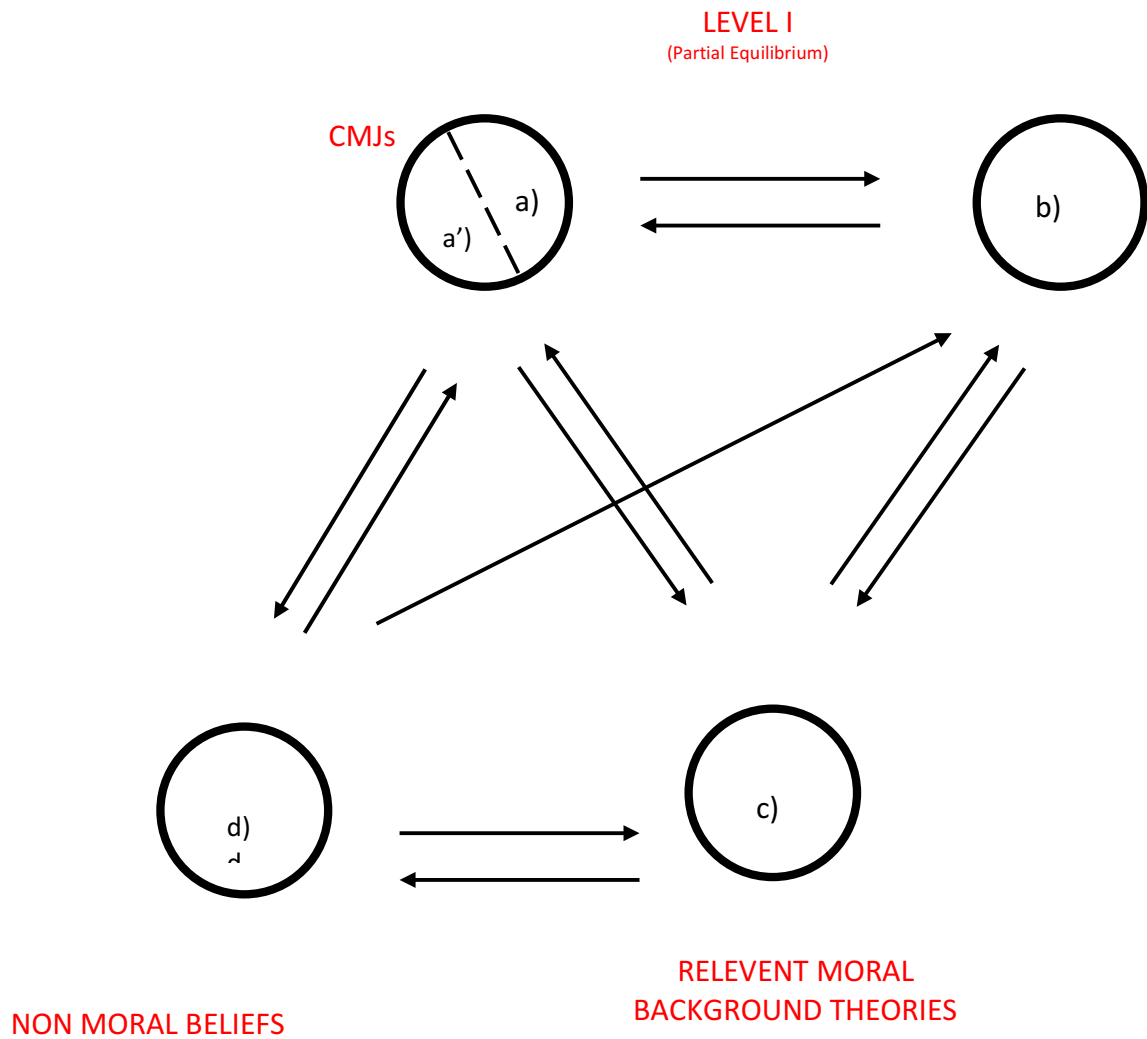
NON MORAL BELIEFS

a) = CMJs

a') = *'significantly disjoint'*
CMJs

b) = General Moral Principles

c) = Moral Background
Theories

d) = Non-moral beliefs