FOREGROUNDING
BACKGROUND THEORIES
IN WIDE REFLECTIVE EQUILIBRIUM

Olle Blomberg
FOREGROUNDING
BACKGROUND THEORIES
IN WIDE REFLECTIVE EQUILIBRIUM

Författare: Olle Blomberg

Ventilerad 2003-06-06

Handledare: Bo Petersson
This thesis deals with two species of a very well known and popular method of moral justification called the method of reflective equilibrium. These two species are narrow and wide reflective equilibrium (NRE/WRE). More specifically, the thesis looks at the roles so called background theories are expected to play in WRE. The use of background theories distinguishes WRE from NRE. Whether background theories can fulfil all these roles, and whether these roles are needed at all, is discussed. One conclusion is that whether WRE is preferable to NRE or not, depends to large degree on how the key concept of coherence is understood. Given a common understanding of coherence as something thicker than mere logical consistency though, WRE is clearly preferable to NRE as a method of justification.


Keywords: reflective equilibrium, background theories, Norman Daniels, coherence, coherentism, moral justification
Nyckelord: reflektivt ekvilibrium, bakgrundsteorier, Norman Daniels, koherens, koherentism, moraliskt rättfärdigande
# CONTENTS

List of abbreviations............................................................................................................. 6

1. Introduction ......................................................................................................................... 7

2. Wide and narrow reflective equilibrium .............................................................................. 8
   Narrow reflective equilibrium ............................................................................................... 8
   Wide reflective equilibrium ................................................................................................. 9
   Background theories ............................................................................................................ 9
   The concept of coherence ..................................................................................................... 10
   Background theories and coherence .................................................................................... 11
   Noting the difference between wide and narrow reflective equilibrium ......................... 12
   The who and where of reflective equilibrium ...................................................................... 13

3. Reflective equilibrium and coherentism ........................................................................... 14
   Foundationalism and coherentism in ethics ......................................................................... 14
   The position of Norman Daniels ......................................................................................... 15

4. Why wide reflective equilibrium? ....................................................................................... 18
   Additional support .............................................................................................................. 19
   Avoiding vicious circularity ................................................................................................. 23
   Less reliance on considered moral judgements .................................................................. 23
   Revealing structural complexity ......................................................................................... 26
   Making reflective equilibrium normative ......................................................................... 27
   Science piggyback ............................................................................................................... 28

5. Wide and narrow reflective equilibrium ............................................................................. 30
   The structure of morality .................................................................................................... 31
   The entanglement of the moral and the nonmoral ............................................................. 32
   Learning by applying ......................................................................................................... 32
   Varieties of wide and narrow reflective equilibrium ........................................................ 33
6. Conclusions .............................................................................................................. 35

Notes................................................................................................................................ 36

References......................................................................................................................... 38
LIST OF ABBREVIATIONS

RE reflective equilibrium.
NRE narrow RE.
WRE wide RE.
CMJ considered moral judgement.
‘filtering process’ the process of filtering moral judgements, saving the ones that qualify as CMJs.
‘reflective process’ the process of adjusting components in order to reach a state of RE.
CMJ\(_i\) an initial CMJ held before the reflective process starts.
CMJ\(_f\) a CMJ held in RE.
(a) set of particular CMJs.
(b) set of general CMJs (moral principles).
(c) set of relevant background theories.
(a’) set of CMJs constraining the acceptability of theories in (c).
WRE-network the network of components (a), (b), (c), and (a’).
MWRE a WRE-network within which moral theories are justified (WRE refers to MWRE unless otherwise noted).
NWRE a WRE-network within which nonmoral theories (theories in science for example) are justified.
TWRE a WRE-network within in which the totality of our beliefs, both moral and nonmoral, are justified.
1. INTRODUCTION

The subject of this thesis is how moral theories can be justified. More specifically, it deals with two species of a very well known and popular method of moral justification called the method of reflective equilibrium (RE for short). These two species are narrow and wide reflective equilibrium (NRE/WRE). WRE is wide because it treats a wider range of considerations as relevant in moral argumentation and theory construction. In trying to justify a moral theory, according to WRE, we are not limited to appeal to particular and general moral intuitions but can legitimately make use of various relevant background theories about human nature and society, some of which might be purely empirical. NRE on the other hand rely solely on moral intuitions about particular cases or general factors and principles. NRE has been charged by both critics of RE in general and proponents of WRE to be inadequate as a method of justification. The critics of RE in general further argue that extending NRE to WRE do not make any difference, it is still inadequate, and for the same reasons as NRE is. Proponents of WRE, on the other hand, maintain that WRE incorporates elements into the method that make moral justification, and even moral truth, a genuine possibility. Finally, there are some philosophers who believe that NRE is not at all inadequate as a method of justification, and that it is at least as good as, if not better than, WRE.

The main purpose of this thesis is to lay out this debate as clearly as possible¹. To get some structure on the debate, I have chosen to focus on the role of the background theories included in WRE (but excluded in NRE). What reasons are there for involving them? Are these reasons sound? Does the involvement of background theories really make a difference for the method of RE? Can they save RE from the criticism directed at NRE? What kind of background theories can do "the job"? What relations between background beliefs and moral beliefs enable them to do this job? Can purely empirical theories help us decide what values to hold?

The thesis starts off with characterisations of NRE and WRE, as well as some discussion about what background theories are and about the key concept of coherence. I also discuss how NRE and WRE can be distinguished. After that, I discuss the relation between RE and coherentism in epistemology, focusing in particular on Norman Daniels’ (1996) influential version of WRE. The central part of the thesis then follows. Here I present and discuss different reasons given in previous writings on RE for bringing background theories into the method. A section where I first discuss some issues about the role of moral background theories in WRE and then try to assess the adequacy of NRE and WRE as models of moral justification follows. Finally, before finishing off with some conclusions, I discuss what NRE and WRE can deliver in terms of convergence/consensus, justification, and truth.

Given that the title of this thesis is Foregrounding Background Theories in Wide Reflective Equilibrium, a reader might expect to find some in-depth analyses of particular background theories that have been used in applications of WRE. No such analyses will be found however. The thesis rather deals with the roles or functions of background theories at a more abstract level of analysis.
I think one of the most interesting and important tasks for moral philosophy is to work out how to make sense of moral justification and rational moral debate. In my view, the method of reflective equilibrium is an account on the right track. This method is not a "method" in the sense of a procedure or algorithm that can be mechanically applied to get the correct solutions to moral dilemmas. Rather, it is a more abstract model that can help us make sense of ethics as a rational and reasonable enterprise, without grand appeals to dictates of Reason or self-evident Truth. The outlook on morality connected, in my view, to the method of RE rests on a middle ground between moral nihilism and moral absolutism. I find this middle ground very attractive and I hope to roughly map out the terrain in this thesis. This kind of exploration is very theoretical in nature and will not deal with any substantial issues in normative ethics. The status of RE as a method of justification obviously has consequences for how one can go about in trying to solve moral problems. But regardless of these consequences, I find the theoretical questions discussed here tantalising and interesting in their own right, and – due to a peculiar fate of human reason – difficult to dismiss.

2. WIDE AND NARROW REFLECTIVE EQUILIBRIUM

Since Rawls’ discussion of RE in *A Theory of Justice* (1971) the concept of reflective equilibrium has become widespread and used to refer to many different things in different contexts. In this thesis ‘WRE’ and ‘NRE’ primarily refer to the views of John Rawls (1971; 1999) and Norman Daniels (1996; 2003). Since their views are far from unambiguous and clear, I necessarily make some decisions of interpretation in my presentation of them. Although Rawls’ and Daniels’ view on RE might differ in some respects, I will assume in this thesis that they are one and the same. From now on I discuss their view, not their views. I will try to present what I think are the most essential elements of RE as conceived by Rawls and Daniels and clearly state when my conception of RE differ from theirs or others’.

**NARROW REFLECTIVE EQUILIBRIUM**

RE refers both to a method and a state that the method is supposed to help us reach. These are logically distinct but associated ideas. It is possible that a state of RE might be reached by another method than the method of RE but since the state must be reflective it cannot be a state hit upon by mere chance. The method is a certain method of justification or theory acceptance. The state is a state of coherence among moral principles, judgements and beliefs. It is a state where “our principles and judgments coincide” (Rawls 1971, 20) and the justification of the moral theory that results “is a matter of mutual support of many considerations, of everything fitting together into one coherent view” (21). When what is fitting together only include our intuitions about particular moral issues and intuitions on general moral factors and principles, we are using the method of narrow reflective equilibrium (NRE) (Daniels 1996, 67). Moral judgements that do not qualify as considered, for example judgements made in circumstances of emotional instability or ignorance, are disregarded. The judgements that remain after this “filtering process” are our considered moral judgements (CMJs). In the RE literature, CMJs are usually divided into two sets, one set (a) with our more particular CMJs and one set (b) with our CMJs about general factors and principles. After going through the filtering process, we try to bring all members of (a) and (b) into coherence by a process of adjustment and pruning. In this “reflective process”, we not only make adjustments to the CMJs in (a) and (b), but we also inductively construct new principles which generalise and explain more particular CMJs (these principles become members
of (b)). When all inconsistencies have been removed from the union of (a) and (b), we have reached a state of NRE. The set of principles (b) held in this equilibrium state constitutes our moral theory. The principles can then be used to give moral judgement about cases and situations where we find ourselves unconfident about how to react or what to believe. If we later find that we give moral judgements or hold views that conflict with our moral theory, we might either revise our judgements or revise our theory to get us back into equilibrium.

**WIDE REFLECTIVE EQUILIBRIUM**

WRE extends the model above by introducing a third set of beliefs (c) which (a) and (b) should cohere with. This additional component allows us to choose between different coherent sets of principles found in NRE. Essentially, NRE and WRE refer to different conceptualisations of the situation where we choose what moral theory we ought to accept. The difference rests in Rawls’ words on “whether one is to be presented with only those descriptions which more or less match one’s existing judgments except for minor discrepancies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one’s judgments together with all relevant philosophical arguments for them.” (1971, 49). In the latter case, we are dealing with WRE. Exactly how and to what extent these versions of RE differ is not completely clear here. To get to grips with this we must understand what the ‘relevant philosophical arguments’ involved in WRE could be, and how they work. Daniels (1996) tries to explicate the structure of Rawls’ WRE by putting the third set of beliefs (c) into the reflective process. (c) is a set of background theories that form the base from which we can bring arguments to bear on different sets of principles (different ‘possible descriptions’ in the quote from Rawls above). The relevant philosophical arguments “can be construed as inferences from some set of relevant background theories” (22). When we adopt the method of WRE, we try to achieve mutual support and coherence in “an ordered triple of sets of beliefs […], (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories” (22). All three sets are potentially open for revision and adjustment. By making adjustments within these three sets of beliefs, we try to reach the *most* coherent possible triple.

**BACKGROUND THEORIES**

The contents of (c) might be made up of very different kinds of beliefs and theories. Daniels for example mentions “beliefs about […] motivation, moral development, strains of moral commitment, and the limits of ethics; about the nature of persons, about the role and function of ethics in our lives; about the implications of game theory, decision theory, and accounts of rationality for morality; about human psychology, sociology, and political and economic behavior […]” (1996, 6). Petersson (1998, 128) points out that it seems impossible to draw a determinate boundary around the set (c). A partly normative theory about personhood for example, is constrained by some psychological theories and “facts”. These, in turn, are constrained by theories about human biology and physiology. Taking this unknitting to the limit, we might end up expanding (c) to include everything we (think we) know. In order to emphasise this very broad-ranging scope of possible relevant background beliefs, one should perhaps follow the bioethicist David DeGrazia and not think about (c) as simply a “set of relevant background theories” but as encompassing all our relevant tacit “background thinking” (1996, 23).

For the sake of exposition and analysis, one can roughly divide possible background theories along the their general character. I will not present any specific background theory here, but
merely discuss general kinds of theories. According to Daniels (1996), background theories can either be of a purely nonmoral nature, or somehow be connected to CMJs. Nonmoral theories might be philosophical theories or descriptive empirical theories. Moral background theories are theories whose acceptability ‘depend on’, or are ‘constrained by’, some CMJs (23, 26, 29, 49). They ‘must cohere’ with some ‘Level III’ CMJs (61), and they ‘incorporate’ a ‘moral notion’ (23). The set of the constraining CMJs is what Daniels refers to as Level III CMJs. The set of these CMJs are referred as the set (a’). It is clear that this set is different from (a). Daniels writes that “both (a) and (a’) are in the same wide equilibrium with the principles, but they constrain their acceptability in different ways.” (Daniels 1996, 50, emphasis in original). How these different ways differ is unclear though and a topic taken up later in this thesis. (Should (a’) be construed as a subset of (c) [(a’) ⊂ (c)], or rather as an additional fourth component part, equal to (a), (b) and (c), of the WRE-network?) Other than nonmoral theories and theories constrained by CMJs, one could also imagine that there are purely moral background beliefs, or (perhaps) theories containing components which are neither purely descriptive nor purely normative. Exactly what theories belong to which of the above categories may be difficult to determine. Some background theories are largely (but not purely) normative. This seems to be the case with “a theory of the person, a theory of procedural justice” (Daniels 1996, 23), and “the implications of game theory, decision theory, and accounts of rationality for morality” (6). Other background beliefs and theories are at least likely to require some kind of moral stance to make sense. Among these are beliefs about the “strains of moral commitment, and the limits of ethics”, about “the role and function of ethics in our lives”, about “political and economic behavior” (6), about “social stratification, class, and gender”, “theories of moral psychology and development” (25), and “theories about ideology” (Nielsen 1991, 234). Still other beliefs are more clearly empirical, for example beliefs about human behaviour coming from psychology, neuroscience, and biology.

**The concept of coherence**

Now that we know the components of both NRE and WRE, we need to look at the notion of ‘coherence’, a crucial notion for both methods. What exactly do we (or should we) mean by ‘coherence’? What kinds of relations within a set of beliefs contribute to the overall coherence of the set? These are very difficult questions and so far no satisfactorily developed account of what ‘coherence’ means in the context of RE has been given (Daniels 2003, sect. 4.2; Petersson 1998, 130-3). Expressions like removal of “inconsistencies”, “principles and judgments coincid[ing]”, “mutual support of many considerations” and “everything fitting together into one coherent view” has surfaced in this thesis so far to describe this slippery notion. Coherence, for a start, is a property of belief-sets. It is a global property constituted by the relations among the beliefs found in a set. To get a grip on the concept of coherence, we need to specify which kinds of relations between beliefs contribute to their coherence, and how these contributions are made. All coherentists, as far as I know, agree that logical consistency (freedom from contradictions) in a set of beliefs is far from enough (some would not consider it as strictly necessary either since it would imply that hardly anyone hold any justified moral belief). Faced with contradictions, any theory could after all resolve these by incorporating new (possibly completely arbitrary) distinctions and specifications in the belief-set. A thicker characterisation of coherence is plainly needed. Apart from logical relations, different kinds of supporting relations also contribute to overall coherence. These could be inductive relations of support (this is part of my characterisation of RE, where principles are sought which generalise, and hopefully explain, a
range of CMJs), or explanatory relations. More abstract higher-level principles could explain the rationale between different lower-level principles that seem to be in tension. For example, a principle of informed consent for adults and one of best interest for children in medicine could be unified and explained by recognition of the value of personal autonomy (this example is taken from DeGrazia (1996, 16)). Other things being equal, coherence also seem to imply a kind of global interlock between different parts of our theory, or different parts of the moral domain. The stronger the nature of this interlock, the better. And the larger the interlocking whole, the better (Sayre-McCord (1996, 166-7) calls these global properties of belief-sets “connectedness” and “comprehensiveness” – DeGrazia calls them “global illumination” and “power” (1996, 15-6)). Other possible elements of coherence are plausibility – a moral theory should spontaneously “feel” correct – simplicity (other things being equal), and clarity. The concept of coherence then, incorporates both what kinds of relations between beliefs, and what global properties of belief systems carry justificatory force. Exactly how much gets packed into the concept is often only vaguely articulated and how much should be packed into it is a matter on which philosophers disagree. One important issue is whether our favoured specification of coherence will enable background theories to contribute to the coherence of moral theories – and if so, which background theories.

**BACKGROUND THEORIES AND COHERENCE**

In what relations to our CMJs in (a) and (b) can background theories stand? Clearly, nonmoral background theories cannot stand in logical/deductive relations to moral beliefs since this would violate Hume’s law (purely descriptive premisses cannot yield a normative conclusion on their own). Nor can a purely nonmoral background theory be inductively arrived at from a set of CMJs. Conjunctions of nonmoral background beliefs and moral/normative principles though, can clearly stand in logical relations to other moral beliefs. For example, beliefs about the possibility for a person to perform some morally desirable action in conjunction with Kant’s “‘ought’ implies ‘can’” can logically imply some CMJ about whether the person ought to perform the action or not. Beliefs about the likely consequences of performing certain actions, having certain character traits, or organising the basic structure of our society in certain ways, often clearly imply moral conclusions when conjoined with relevant CMJs. These kind of supportive relations between nonmoral and moral beliefs are explanatory. Empirical knowledge can help us explain – or make sense of – some of our CMJs. Beliefs about the physical and psychological needs, powers and limitations of human beings are obvious candidates for this explanatory role. These beliefs tell us something about the conditions of human flourishing and welfare, likely to be of moral relevance in any sensible moral conception. Knowing these things can help us make sense of, or possibly make non-sense of, intuitions about moral responsibility, which character traits are worthy of admiration or to be shunned, what actions ought to be encouraged or discouraged, etc. There is no violation of Hume’s law in these cases, because there are no simple direct relations between nonmoral and moral beliefs, the relations only exist in, and in virtue of, larger systems of beliefs with both moral and nonmoral components (Sayre-McCord 1996). (A puzzling question here is whether there is any substantial difference between (i) the case where some set of nonmoral beliefs in (c) conjoined with some subset of CMJs in (a) stand in an explanatory relation to a different subset of CMJ in (a), and (ii) the case where a background theory in (c) constrained by some subset of CMJs in (a’) stand in an explanatory relation to some subset of CMJs in (a).)
Assuming that we see a role for comprehensiveness and connectedness as properties contributing to coherence, empirical theories can also have an undermining effect on CMJs by knocking down the explanatory support (in the form of nonmoral beliefs previously thought to be true) they rest on. The undermining effect is brought about when the set of beliefs in which the CMJs take part becomes less connected as a result of loss of explanatory support. The clearest example of such an undermining effect is perhaps the changes brought about by evolutionary theory in our view of the moral status of animals (DeGrazia 1996, 26; Rachels 1993, 119). Generally, Darwinism acts as a “universal acid” on traditional views of humanity, life and nature (the analogy is Dennett’s (1995)). Evolutionary theory has completely transformed what we take to be the relation between ourselves and the rest of nature. This is because, in the light of our knowledge about evolution, the biblical story about human origins, which earlier gave explanatory support for mistreating and eating animals, has gone utterly bankrupt. Nowadays, at least partly due to Darwin, such practices can reasonably be questioned. It is not that our new “facts” about evolution proves that such practices are wrong, but by undermining part of their support, the moral conceptions that imply that these practices as permissible becomes slightly less coherent (because their degree of connectedness decrease). Moral conceptions that have had these judgements revised on the other hand, retain, or even strengthen, their degree of coherence (connectedness).

### Noting the Difference Between Wide and Narrow Reflective Equilibrium

Given my characterisations of WRE and NRE it might seem that distinguishing between them is unproblematic: In WRE, we try to get general and particular CMJs into coherence not only with each other (as in NRE) but also with a set of background theories, some of which may be nonmoral. The background theories enable us to choose between different states of (equally internally coherent) NREs. Although the difference between the methods seems straightforward at a glance, I will argue that it is really not that clear-cut.

With the introduction of background theories, empirical “facts” clearly get a part to play in WRE. But do they also have a role to play in NRE? Of course, this really depends on how we want to characterise and describe NRE, but I want to suggest that descriptive premisses do have a role to play in NRE, at least if we want to give the method a charitable interpretation. In trying to check for coherence among general and particular I do not think NRE can do without descriptive premisses. My suggestion is that to check if a general and particular CMJ are contradictory, many times a connecting descriptive premiss is needed. Perhaps a proponent of NRE would answer that the connecting premisses that I claim she needs are already embedded in the CMJs, since she could require that all moral judgements that gets through the filtering process have been formed in light of ‘knowledge of relevant facts’ or something similar. Then the connecting premisses are actually part of what the CMJs are judgements of in the first place. To illustrate what I mean by ‘connecting descriptive premisses’, here is a simple example. If we hold the general CMJ ‘One ought not to kill persons’, as well as the particular ‘It is okay to kill that chimpanzee’, we need to know some facts about chimpanzees (whether they are persons or not) to be able to determine whether we have a contradiction in our NRE-network or not. Empirical facts also underlie the filtering process in NRE. In deciding which moral judgements to count as considered, we rely on “background theories” (roughly speaking) about what constitutes good moral judges and about the influence of various situational factors on people’s judgemental capacities. Possibly, an
One possible difference between WRE and NRE could be the role given to moral background theories. If one views the role of moral background theories as merely that of bringing in more CMJs \([a']\) that can be added to the set of general and particular CMJs \([b \cup (a)]\) within which we try to achieve maximal coherence, then the difference between NRE and WRE, along any claimed benefits of WRE, seems to fade into thin air. The reason is that, using NRE, we could simply make sure that we collect all our CMJs in \((a)\) from start, making the inclusion of any moral background theories completely redundant. If one, in addition, employ a very thin notion of coherence (for example equating it with logical consistency), then nonmoral background theories cannot do much for RE (this could plausibly, I think, be the grounds for Holmgren’s (1987, 1989) view of NRE as superior to WRE). Proponents of WRE might have a different view of the role of background theories and beliefs though. The role of moral background theories might not be to merely bring additional CMJs into the reflective process but to bring in different kinds of CMJs, kinds that are not (cannot be) represented in \((a)\). Daniels, for example, speaks of CMJs as belonging to, or connected to, a certain moral notion, for example to “justice”, “fairness”, “which features of persons are morally central or relevant”, or to “rights and entitlements” (Daniels, 1996:49-50). The domain of morality is perhaps not a unified whole without internal structure, as seems presupposed in NRE, but an arrangement of different subdomains clustered around different moral notions. If one embraces such a conception of the moral, then the use of moral background theories in WRE makes a lot more sense. Coupled with a wide understanding of coherence, utilising anything less than both moral and nonmoral background theories in RE would seem unwarranted.

I think the most important difference between NRE and WRE lies in that in WRE, we systematically use background theories to compare and rank different moral theories, while in NRE, descriptive premisses are simply used and assumed when they are needed in checking for consistency in \((a)\) and \((b)\). The background theories provides a vantage point from which comparisons between different equally internally coherent NREs can be made, and we are to consider “all possible” such belief-sets and “all relevant philosophical arguments” for or against the adoption of these sets of beliefs (this is an ideal, not a requirement). Anyhow, at this point it should be noted that the differences between NRE and WRE is not nearly as clear-cut as is sometimes assumed in the RE literature.

**The who and where of reflective equilibrium**

Both Rawls and Daniels seem to treat RE as an account of personal (or "doxastic") justification; the reflective process is supposed to be carried out by one particular person, making adjustments in her own sets of belief (Daniels 1996, 22; Rawls 1971, 50). But in a different passage, Daniels states clearly that he is giving an account of impersonal (or "propositional") justification (Daniels
1996, 41n1). On the former treatment, the issue is when a person $P$ is justified in holding a belief $B$. On the latter one, the issue is when a belief $B$ is justified. One can also construe RE as a form of interpersonal justification, where the issue is when a person $P$ has justified a belief $B$ to some other person (or group) $Q$. $P$ and $Q$ could be persons involved in a disagreement over how to handle some moral problem or people living in a community in which some formerly agreed upon norms are being questioned and defended. This last kind of justification seems, at times, to be what Rawls was aiming at when he developed and used the method of RE (see his remarks on justification in (1971, §87 – especially 580-1)). The issue here is both whose CMJs and beliefs make up (a), (b), and (c), as well as where the reflective process unfolds (in a person’s mind, between minds, or in an impersonal realm of pure concepts). The fact that both proponents and critics of RE has been unclear about the ‘who and where of RE’, has I believe messed up the debate somewhat. I also believe that one of these three conceptions of RE and moral justification is preferable to the other two. But more on that later on.

3. REFLECTIVE EQUILIBRIUM AND COHERENTISM

Another issue, which needs to be brought up, is the connection between the method of RE and coherentist moral epistemology. First, I will describe briefly the meaning of coherentism and foundationalism in moral epistemology and discuss how closely tied Daniels’ version of RE is to coherentism. My opinion (with Sayre-McCord 1996, 143; DePaul 1986, 60; Ebertz 1993, 200) is that RE in general is not so closely tied to coherentism as is sometimes stated. (One might use the method of RE because one thinks it is an effective way of discovering justified moral beliefs, without holding that their justification turn on coherence-contributing relations to other beliefs.) Furthermore, I will argue that Daniels can be interpreted as rejecting a strong form of coherentism, acknowledging that some (nonmoral) beliefs that are part of the WRE-network are epistemically privileged.

FOUNDATIONALISM AND COHERENTISM IN ETHICS

Foundationalism and coherentism in ethics are views on what the structure of moral justification (should) look like. According to most accounts of foundationalism, for a moral belief to be justified it must either be directly justified or be indirectly justified by being derivable from a directly justified moral belief. There is no other way for a belief to be justified. That a belief is directly justified does not necessarily mean that it is infallible or unrevisable, merely that its justification does not depend on how the belief is related to other beliefs one hold. The beliefs that are justified directly form a set of foundational beliefs on which all other justified nonfoundational beliefs depend. Using the terminology of RE, one version of foundationalism in ethics would treat CMJs as providing fallible foundations for a moral theory (a classical form of intuitionism). In the case of moral justification, there is also another kind of foundationalism that does not treat any moral beliefs as directly justified, while still giving certain beliefs a special epistemic privilege. Some moral beliefs might be justified purely in virtue of their relations to some nonmoral beliefs. If we can derive some moral beliefs from nonmoral background theories, these can be used as a foundation for our moral theory, indirectly justifying all other nonfoundational moral beliefs. Whether to call this last position foundationalist or coherentist is a matter of whether we focus on a totality of beliefs, moral as well as nonmoral, or on only the total set of moral beliefs. Focusing on the set of all beliefs, both moral and nonmoral, the position is coherentist in the sense that all justified beliefs are justified in virtue of their relations to other
beliefs. If the focus is only on the set of all moral beliefs though, then the position foundationalist in the sense that there are some justified moral beliefs that are not justified in virtue of their relations to other moral beliefs, but purely in virtue of their relations to nonmoral beliefs.

Coherentism is the view that beliefs are justified in virtue of how well they cohere with all other beliefs we hold. A strong version of coherentism would claim that all justified beliefs are justified in virtue of the beliefs’ relations to other beliefs. This, it is claimed, is the only way in which a belief can be justified. A weaker version would simply claim that the members of at least one subset \( U \) of the set of all justified beliefs \( S \) are justified in virtue of the members’ relations to other beliefs within \( U \). The other justified beliefs, the members of the complement of \( U (S – U) \), are justified directly. There can be no logical relations between \( U \) and \( S – U \), since that would make the weak coherentism collapse into foundationalism. Various supporting relations between the two subsets are possible though. Since coherence is not an all or nothing parameter, where the thresholds for necessary and sufficient coherence are set needs to be specified. According to most versions of coherentism, we need to be able to speak of degrees of coherence, of a ‘best fit’. For a belief to be justified it should be part of the set of beliefs, taken from a range of alternative sets, which is the most coherent one. On a weak version of coherentism in ethics then, for a moral belief to be justified, it must either cohere with the most coherent set of moral beliefs we can think of which is logically unrelated to any foundational belief, or be logically related to some foundational belief. On a strong version of coherentism, cohering with the most coherent set of moral beliefs we can think of is the only way for a moral belief to be justified.

**The position of Norman Daniels**

According to Daniels, adjustments are clearly permissible in all components during the reflective process: particular CMJs (a), general CMJs (b), and background theories (c). Hence, he seems to embrace non-foundationalism and the claim that “no judgement is immune to revision” (Petersson 2000, 36). Perhaps he also embraces the stronger claim “no judgement in itself has more (epistemic) weight than another” as Petersson (36) believes. How to best interpret Daniels is a very problematic issue though. Although support for Petersson’s strong coherentist reading can be found in Daniels (1996), I make a different reading which I believe has overall textual support in its favour. Daniels clearly thinks of WRE as (in his own words) “a coherentist model of justification” (1996, 60). In light of other remarks made by Daniels, this label should be understood as implying two things: (I) cohering with the most coherent possible WRE-network from some range of alternatives is sufficient for a moral belief to be justified, and (II) whatever (if any) epistemically privileged beliefs we may find in the WRE-network, these will not be able to provide a foundation for an adequate moral theory. The ‘coherentist’-label does not imply that there are no epistemically privileged beliefs in the WRE-network. Furthermore, one could argue that the claim that “no judgement is immune to revision” should be qualified in ways which would make the elements of coherentism left in Daniels’ WRE even fewer (which would bring it in line with Petersson’s own favoured version of WRE). By presenting this interpretation I hope not only to present my favoured interpretation of the most influential exponent of WRE, but also to demonstrate that whatever connection exists between WRE and coherentism, it is far from straightforward.
In several places in the papers collected in *Justice and Justification* (1996) Daniels expresses himself in ways that lead me to interpret him as indeed giving empirical facts (or at least observation reports) more epistemic weight than CMJs. For example, on page 26, he expresses concern for the fact that if disagreements about background theories rest on CMJs in (a’), then it will not be clear how the introduction of background theories can facilitate either justification or convergence/consensus. This concern seems to rest on the assumption that CMJs are somehow less hospitable to widespread agreement than observation reports, and disagreements about background theories are more tractable than disagreements about CMJs (25). A CMJ is according to Daniels “in many ways far more like a ‘theoretical’ than an ‘observation’ statement” (30).

While he is carefully agnostic about whether there are any moral truths, he alludes that possible agreement on background theories might be “found because some of the background theories are, roughly speaking, true – at least with regard to certain important features” (32). Furthermore, on page 28 when he discusses theory-based revisions of CMJs that lead to adjustments in our background theories, he always assumes that it will be the CMJ-components in (a’) belonging to these theories that get revised, not the “empirical facts” (look up page 60-61 and 70-71 for similar expressions):

“In seeking wide reflective equilibrium, we are constantly making plausibility judgments about which of our considered moral judgments we should revise in the light of theoretical considerations at all levels. No one type of considered moral judgment is held immune to revision.” (28, my emphasis).

“No considered moral judgments at any level are taken to be unrevisable, that is, strongly foundational…” (60, first emphasis is mine)

Daniels also writes that “nonmoral deep theories [i.e. background theories]... in general underdetermine moral theory selection” (1996, 61, similar remark on 29-30) but never points out the reverse, that moral principles and judgements “underdetermine” the choice of nonmoral background theories. Presumably, this is because he does not think that moral theories could ever determine nonmoral theories in a way that nonmoral theories could determine moral theories (note the “in general” in the quote above).

Daniels does not, it seems to me, exclude the possibility of nonmoral background theories determining moral theories, but merely deny that they, as matter of contingent fact, actually do. Furthermore, he at one time talks about a “nonmoral wide reflective equilibrium” (32), suggesting that although observation reports and facts are reversible within this nonmoral WRE, they may perhaps not be reversible within the moral WRE (this interpretation is due to Petersson 2000, 41-2). Finally, in a recently written encyclopaedia entry on reflective equilibrium, Daniels makes the following comment on Rawls’ use of the contract in *A Theory of Justice*:

“If Rawls were trying to justify the structure of the contract by appeal to theories that themselves were completely non-moral, then he would be offering the kind of independent justification for the principles that would characterize them as foundational (Daniels 1996, Timmons 1987), so the claim that the background theories are themselves moral is part of the rationale for concluding that Rawls is clearly rejecting foundationalism.” (Daniels 2003, sect. 3.2.1, third paragraph)

For this comment to make sense, Daniels must hold that if the structure of the contract was justified using only nonmoral background theories, then Rawls could not plausibly be interpreted as rejecting foundationalism. Judging from this, again, Daniels thinks nonmoral background
theories have a higher epistemic status than CMJs, although he does not think that a moral conception can be supported solely by nonmoral background theories.

It is worth noting that the examples of background theories that Daniels brings up in discussing WRE – all of them taken from Rawls’ political theory – are usually not nonmoral ones. Nonmoral background theories are collected in a category separate from the moral ones by Daniels. They end up among “level IV” background theories, which are used to assess the stability of conceptions of justice (or more generally, morality). Moral background theories are on the other hand “level III” theories, more directly involved in justification.

In light of the remarks by Daniels I have brought up I think the most reasonable interpretation of Daniels is that he does not embrace the strong claim that “no judgement in itself has more (epistemic) weight than another”. He clearly embraces the claim that “no judgement is immune to revision”, although not necessarily the claim that “no judgement is immune to revision in a moral WRE”. Admittedly, the question about whether Daniels is a strong coherentist or not really depends on whether we are only looking at moral theory construction and justification or on his view of theory construction and justification in general (both scientific and moral). Daniels is clearly a coherentist about the epistemology of science. The reason that Daniels believes observation reports are less open to revision (although not immune to it) in a nonmoral WRE (NWRE) than CMJs are in a moral WRE (MWRE) is that we have a plausible (neurological/psychological) story about the reliability of those reports, while such a story is at present nowhere to be found about CMJs (1996, 32, 2003, sect. 4.1). The reason is not then, that scientists’ observation reports “in themselves” have more epistemic weight than the CMJs of moral agents. If WRE is taken as radically wide, to include the totality of our beliefs (Petersson 1998, 128-130), then Daniels could be interpreted as embracing a strong coherentist account of justification. Whatever Daniels’ view really is, I think (with Petersson 2000) that the most plausible version of WRE is one in which purely empirical components of the moral WRE-network are insulated from revision pressures within that network. On this version, instead of there being one WRE encompassing the totality of our beliefs, nonmoral as well as moral (a TWRE), there are two WREs, one for moral justification (MWRE) and one for nonmoral justification (NWRE).

It should be noted that regardless of these problems of interpretation, there are elements of WRE that does not seem to be part of the WRE-network, and hence not revisable in it. First, beliefs about the setup of the “filtering process” and its rationale are not fixed with the method of WRE. Secondly, beliefs about the importance of consistency and coherence seem to be presupposed by the method, and not open to revision. We can call these unrevisable elements meta-beliefs and meta-principles. If the principles that govern WRE is not accepted by some of the actors involved in the reflective process, then those actors cannot be obliged to accept whatever fruits the process bear (unless it can be shown that they must accept the meta-principles). WRE “assumes that persons are rational and will be persuaded by sound arguments […].” (Daniels 1996, 70). I am treating WRE as a method for moral inquiry here, not as a general position in epistemology. If one views WRE as a general epistemological theory, equating our network of judgements, principles and background theories with what Quine called our “web of belief”, encompassing all
our beliefs, then perhaps there are no unrevisable meta-principles to fall back on. But this is not, I believe, Rawls and Daniels’ conception of WRE.

4. WHY WIDE REFLECTIVE EQUILIBRIUM?
Daniels claims that using the method of wide instead of narrow RE "increases our ability to choose rationally among competing moral conceptions" (1996, 26). I will try to present and scrutinise arguments made to support this claim. How can the introduction of background theories make Daniels’ claim true?

A lot of philosophy papers have been written on reflective equilibrium since Rawls introduced the term to moral philosophers in (1971), but only a few of these deal with the differences between WRE and NRE, or the role played by background theories in WRE. John Rawls explicitly distinguishes wide from narrow RE in “The Independence of Moral Theory” (1975), and Norman Daniels’ early papers collected in (1996) deal extensively with the method of WRE and reasons for adopting it. I have also relied on the elaboration and defence of WRE presented by Kai Nielsen in his monograph After the Denial of the Tradition (1991). Questions brought up by reading Margaret Holmgren’s defence of NRE in “The Wide and Narrow of Reflective Equilibrium” (1989) impacted greatly on my thinking about NRE and WRE, which should be evident throughout this thesis. Naturally, I have not merely relied on proponents of RE, but also read some attacks and responses from critics (Brandt 1979; Hare 1975; Haslett 1987; Sencerz 1985). By going through this literature and thinking about these issues myself I found the following six claims about background theories which provide reasons for including background theories in the RE-network (detailed accounts of these claims and reasons appear further on):

(1) Additional support. Background theories give additional support or ‘justificatory force’ to the moral conception/theory given in the state of WRE.

(2) Avoiding vicious circularity. With the help of background theories, WRE can handle the so called ‘circularity objection’ to NRE. While justification is still circular in WRE, it is not so in a vicious way.

(3) Less reliance on CMJs. Many have criticised the method of NRE for being too conservative, placing too much weight on moral intuitions. By introducing background theories into RE, we are not entirely at the mercy of our unreliable CMJs.

(4) Revealing structural complexity. By applying the framework of WRE with its component parts to our and others’ actual moral conceptions, we get a clearer picture of how these conceptions hang together, in what ways they differ and what unites them.

(5) Making RE normative. Sometimes it is claimed that RE can only serve as a framework for describing moral theories or conceptions, not as a source of recommendations about which moral theories should be accepted. Background theories enable RE to give such recommendations, i.e. making RE serve a normative purpose.
(6) **Science piggyback.** If the model of justification inherent in the method of WRE has succeeded in accounting for objective scientific truth, perhaps this model can do similar work in the domain of ethics and show us whether objective moral truths exist.

(1) – (6) is far from entirely independent of each other, but rather highly overlapping and interrelated. The claims in (1), (2) and (3) are closely related to each other, as are the claims in (4), (5) and (6). In the case of claims (1), (2) and (3), these respond to different objections to NRE. (1) respond to the objection that the moral theory coming out of NRE is too underdetermined. (2) respond to the objection that justification by the method of NRE is viciously circular. (3) respond to the objection that NRE makes moral theory construction and justification to reliant on CMJs. Both (5) and (6) build on, but are both distinct from, the claim in (4).

**Additional support**

According to the methodology of NRE, CMJs in (a) and (b) get their justification from cohering with other CMJs in these two sets. None of our general CMJs (moral principles), nor any of our particular CMJs has a privileged status when it comes to making adjustments in (a) and (b) to achieve coherence. Without any further constraints, this process of adjusting components to achieve equilibrium can develop along many paths, leading to many different equally coherent NREs. Few would take coherence of such a narrow scope as constitutive of justification. Since there is no scale other than their internal coherence by which we can rank the different NREs, there is no method for rationally choosing among them (if they are equally coherent). To achieve a ranking we need a vantage point outside the competing moral conceptions themselves.

Such an exterior vantage point is provided by various relevant background theories according to Daniels (1996). As I have already mentioned, this vantage point does not constitute a foundation for a moral theory in Daniels view, it is simply an additional component of the RE-network. As such it constrains what moral theory we ought to accept. The background theories support inferences that can be used to choose between different NREs (Daniels calls a NRE a “level I partial equilibrium” in this context (23)). By supporting or possibly destabilising NREs from outside, background theories provide independent criteria for evaluating which of the NREs are preferable from the point of view of justification (providing that the background theories satisfy the “independence constraint”, see more on that below).

How one understands justificatory support and what one believe gives such support depends on (first) one’s conception of coherence, and, (secondly) one’s view of the connection between the justification of nonmoral judgements and the justification of moral judgements. If one has a thin conception of coherence as mere logical consistency, then only elements standing in logical relations to the CMJs of our original NRE can give additional justificatory support. This will clearly stop purely nonmoral elements from contributing to moral justification. With a thicker conception of coherence on the other hand, this remains possible. Two views on the connection between moral justification and nonmoral justification have already been mentioned. According to one view, these kinds of justification are really different perspectives on one and the same TWRE-network of beliefs, in which we try to justify both our moral and our nonmoral beliefs. According to the other view, while nothing stops nonmoral components from playing a part in an MWRE-network, these components themselves are justified in a separate NWRE-network,
immune to revision pressures from the components of the MWRE-network. On this view, nonmoral components have a privileged status in the MWRE-network (which they take part in, but are not justified through).

Given these starting points there seem to be three possible versions of claim (1) (where holding one version does not exclude the possibility of holding others):

(i) The inclusion of moral background theories in RE make our moral theory more inclusive and less likely to merely be the outcome of accidental generalisations from a too narrow set of CMJs in (a) and (b). (Daniels 1996 on the “independence constraint”, for example 22-3, 49-50; Holmgren 1989, sect. III.)

(ii) Background theories connect moral and scientific theory construction, which means more constraints on both moral and scientific theory, making our choice of moral theory less underdetermined. (Daniels in Petersson’s (2000) interpretation; White 1998, especially 652-4.)

(iii) A different kind of support for our moral theory is given by in the inclusion of relevant and appropriate background theories. (Daniels (1996) when read as a weak coherentist; Holmgren 1989, sect. II; Peterson 1998, 2000.)

Juxtaposing these three versions of claim (1) with the distinctions made above between different views of coherence and the connection between moral and nonmoral justification, we get the following table:

<table>
<thead>
<tr>
<th>MWRE / NWRE</th>
<th>TWRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thin coherence</td>
<td>(i)</td>
</tr>
<tr>
<td>Thick coherence</td>
<td>(iii), (i)</td>
</tr>
</tbody>
</table>

Minimally, (i) - (iii) view the introduction of background theories as a means of achieving ‘maximal coherence’ (Daniels 1996; Sayre-McCord 1996). On a thin conception of coherence, ‘maximal coherence’ means maximal coherence among CMJs, nothing else ((i) becomes the only possible interpretation of claim (1)). Additional CMJs brought in by background theories further constrain the construction of moral theories simply due to logical relations between different general and particular CMJs. On a thicker conception of coherence, the search for ‘maximal coherence’ is interpreted as (ii) or (iii), depending on whether or not one thinks that purely empirical background beliefs are revisable and justified within the same TWRE through which our moral theory is justified. Roughly speaking, (iii) corresponds to my interpretation of Daniels, while (ii) corresponds to Petersson’s interpretation. In (ii) and (iii), various supporting relations between CMJs and nonmoral beliefs create additional support.
Background theories that come with moral components [(a')] could give moral theories additional support by making theory construction less likely to amount to nothing more than mere generalisation of, and abstraction from, an unnecessarily limited set of CMJs. With Hume’s law\(^\text{10}\) in mind, one might think that this is the only way in which background theories can support CMJs, ruling out (ii) and (iii).

I will now introduce what Daniels calls the independence constraint. This is a constraint on which background theories can provide justificatory support for a moral theory. Generally, the constraint ensures that the grounds for the support given by (a) and (c) to our set of moral principles (b) are independent from each other. One way in which this constraint is cashed out is by requiring that (a’) must “be to some significant degree disjoint” (1996, 23) from (a). This would normally be satisfied for example when the moral notions of (a) and (a’) are different. If the background theories in question are nonmoral, without constraints from (a’), I take it that the independence constraint is automatically satisfied according to Daniels. The ‘notions’ nonmoral background theories trade in are after all completely different from moral ones. If we can introduce moral background theories that satisfy the independence constraint, then perhaps we can boost the justificatory support for our moral principles in (b). That, at least, is the claim of (i).

Several commentators of Daniels however, have pointed out something strange about the independence constraint (DeGrazia 1996, 26; DePaul 1993, 20-1; Haslett 1987, 307-8; Holmgren 1989, 58-9). Why shouldn’t we simply include all our CMJs in (a) from the start, including the ones to be found in (a’)? Not to take all those CMJs into account from the beginning seem to be contrary to the method of both NRE and WRE, indeed contrary to honest moral inquiry. As Haslett puts it:

“Deliberately not to take some of our considered moral into account in arriving at a narrow reflective equilibrium would appear to have no other purpose than that of keeping them held back in reserve so that they can then be used for the first time in arriving at a wide reflective equilibrium, thus allowing us to claim that the independence constraint has been satisfied.” (Haslett 1987, 308)

How can this objection be met? I think the objection is based on a view of the moral domain that Daniels does not share (see NOTING THE DIFFERENCE BETWEEN WIDE AND NARROW REFLECTIVE EQUILIBRIUM in section 2). Considering how moral theorising proceeds, I think Daniels has something like the following in mind. When we engage in moral theory construction we are usually interested in a limited range of issues connected to a certain moral notion, say “justice” or “what we owe to each other” (some theories, utilitarian ones for example, do try to cover the whole domain of morality though). This is clearly how Rawls proceeds in the justification of his theory of justice (which Daniels throughout his writings on WRE uses as an example of WRE in action). The general idea is that we collect all our CMJs regarding this notion and how it applies to different situations and then we try to systematise these judgements into general principles through the reflective process. The independence constraint requires that whatever CMJs are part of the background theories that supports certain partial equilibria (different NREs) should be part of other subdomains of morality, based on different moral notions. Textual support for this interpretation is to be found in several places\(^\text{11}\).
The first (i) version of the claim for additional support then, seems to assume certain structural features of the moral domain, or otherwise the adoption of WRE rather than NRE does not make any sense. If the moral domain is uniform or hierarchically structured (for example with the notion of utility overriding all others) then WRE does not seem to have any advantages over NRE. (In fact, it seems to be worse off since NRE incorporates a better strategy for avoiding accidental generalisations of CMJs, see Holmgren 1987).

(ii) According to the second (ii) version of claim (1), the inclusion of background theories means that we not only have to respect a larger number of constraints set up by additional CMJs when constructing moral theories, but also constraints set up by purely empirical background beliefs. Vice versa, CMJs put constraints on the relevant background theories that are included, since no judgement in the TWRE-network, moral or nonmoral, is immune to revision according to the TWRE-view of justification. Arguing for the inclusion of background theories with (ii) does not necessarily mean a commitment to the view that “no judgement in itself has more (epistemic) weight than another” (Petersson 2000, 36), only that our nonmoral beliefs are not completely insulated from revision pressures from CMJs. Morton White for example, holds that revision of empirical judgements are permissible in the light of moral judgements, although such revisions would presumably rarely occur (White 1998, 654)\(^1\). Sometimes, this also seems to be Daniels view (see endnote 9).

(iii) One might be slightly disturbed by the fact that according to (ii), it is permissible to revise purely empirical judgements in light of our moral views, in case such revisions lead to a higher degree of coherence in our overall view of the world. (Perhaps this disturbing fact is what has lead Daniels to be so vague and ambiguous about what the relation between moral and nonmoral justification is in his account of WRE.) One might find the view that nonmoral background theories and beliefs do have an important role to play in the justification of moral theories attractive while rejecting the claim that the reverse holds. This is claim (1) according to the third (iii) interpretation. This is, I believe, the view that Daniels (1996) is most consistently leaning towards, although, as I have already noted, his writings are ambiguous. This is also the view of Petersson (1998, 2000), as well as my own view. All beliefs might be revisable, but not in all contexts. And the contexts of moral and nonmoral justification are different.

It might be argued that since scientific theories are always underdetermined by observation reports, it ought to be permissible to revise empirical theory-laden “facts” in light of moral considerations as long as the new revised “facts” answer equally well to all observation reports. While this might indeed be permissible, moral considerations seem to be completely irrelevant to what we ought to believe to be true about the empirical world as explored by scientific inquiry.

Most people, including myself, think that there is a difference in epistemic status between empirical and moral judgements. Empirical judgements seem to be more reliable, or at least we have some idea of what it means for such judgements to be reliable (which is not true of CMJs, see Science Piggyback below). The hope behind (iii) then, is that not only do moral theories need to pass through additional checkpoints but, that but that moral theories need to pass through
checkpoints of a different kind, a kind which is more stable, secure and likely to receive widespread acceptance. Various supporting relations can infuse our moral beliefs with the (alleged) superior epistemic status of empirical judgements.

**AVOIDING VICIOUS CIRCULARITY**

Another reason for introducing background theories, which is closely tied to the discussion of (1b) above, is to counter an objection directed at NRE called the ‘circularity objection’. The objection is that the justification of NRE is circular in nature. The general CMJs of (b) that are supposed to give support to our particular CMJs in (a) are themselves justified solely by those very CMJs themselves. Daniels, who is sceptical about NRE as a method of justification, believes that “there must be more to moral justification of both judgments and principles than such simple coherence considerations” (1996, 21). In order to move beyond this objectionable circularity of justification, Daniels believes that we should introduce background theories that fulfil the independence constraint. I have already brought up a serious objection to this constraint above in my discussion of **additional support** and this objection obviously holds in this context as well. Artificially separating the totality of our CMJs into two independent sets (a) and (a’) does not seem to enable us to avoid the circularity objection raised above.

To put the questions surrounding the independence constraint to rest, defenders of WRE must explain the nature of the set (a’) and how it differs from (a). This difference can perhaps be explained entirely by the idea of different moral notions. But it is still far from clear what it means for (a’) to constrain the acceptability of (c). To avoid the criticism of DePaul and others, this constraining relation, whatever its nature, cannot be the same kind of relation as the one between the CMJs of (a) and (b), and the beliefs in (c) (following DePaul 1993, 20-1). If (c) is constrained by (a’) simply because (c) must cohere with this set of CMJs to bring our network of beliefs into coherence, then (c) is also constrained by (a) in exactly the same way. One important feature of WRE we need to understand then, is the relation between (a’) and (c).

However, proponents of WRE do not believe the circularity objection to be as serious as the critics of RE claim. Neither Daniels nor Nielsen denies that there is a circular structure to justification in WRE, but circularity goes with any coherentist account of justification. There are better and worse kinds of circularity though. Daniels and Nielsen are groping for a circle as wide as possible (Daniels 1996, 59; Nielsen 1993, 203).

**LESS RELIANCE ON CONSIDERED MORAL JUDGEMENTS**

A well-known objection against NRE is that it needs to be coupled with foundationalist moral epistemology in the form of intuitionism (Brandt 1979, 16-23; Daniels, 1996:27; Hare 1975, 82-5). This is seen as a consequence of NRE’s exclusive reliance on CMJs. Proponents of WRE claim that WRE does not require the same strong reliance on CMJs as NRE though. If this were indeed true, then it would constitute a strong reason for adopting the method of WRE rather than NRE. Daniels states that "[w]ide reflective equilibrium keeps us from taking considered moral judgments at face value" (1996, 28). Daniels agrees that RE in its narrow version demands too much faith in our considered moral judgements and thinks it should properly be called intuitionism. In a similar manner, Kai Nielsen characterises NRE as “the inadequate method used by contemporary intuitionists” (1993, 199), but then argues extensively for WRE as the most
viable model of justification available. Is it fair to say that with NRE follows intuitionism? And if so, does the transition to WRE really make the difference Daniels and Nielsen claim it does?

But first: Why should labelling RE as intuitionism be considered as an objection? The reason that intuitionism, the view that we are justified in trusting our CMJs at face value, raises so many voices of objection is that there are many possible debunking explanations for us holding the CMJs we happen to hold. CMJs might after all merely reflect “class or cultural background, self-interest, or historical accident” (Daniels 1996, 21). In light of this, many philosophers conclude that CMJs do not have the required credibility for being the building material of moral philosophy. If the initial CMJs do not have any credibility, why would a reflective process that leads to coherence among those CMJs ever give us a justified moral theory? (According to the principle of “garbage in, garbage out”.) In Brandt’s words, “there is no reason to think that this claim is true unless some of the beliefs are initially credible – and not merely initially believed – for some reason other than their coherence, say because they state facts of observation” (Brandt 1979, 20). Similarly, according to Hare, the mistake of what he calls subjectivism (which he attributes to Rawls) “is to make the truth of the theory depend on agreement with people’s opinions.” (1975, 83). How should we go about then? According to Brandt, what “we should aim to do is step outside our own tradition somehow, see it from the outside, and evaluate it […]” (1979, 22).

It is useful to remind ourselves of the distinction between personal, interpersonal, and impersonal justification here. Both Hare and Brandt seem to think that the only kind of justification worth wanting and having is impersonal justification. This kind of justification is formulated only as an unattainable ideal in WRE. But it is recognised – and this is crucial – that this is an ideal we should strive towards. Using WRE, “one seeks the conception, or plurality of conceptions, that would survive the rational consideration of feasible conceptions and reasonable arguments for them.” (Rawls 1999, 289, my emphasis; see also Rawls 1971, 49; Nielsen 1991, 236; Beauchamp & Childress 2001, 399). Trying (without ever actually succeeding) to find this final conception(s) by expanding the horizon of our tradition is part of the method of WRE. In other words, it is an important meta-principle defining the method. This aim to expand, criticise and transform our tradition – to learn collaboratively at the collective level – makes sense if we conceive of justification as occurring between persons, where ‘traditions’ and ‘people’s opinions’ actually meet and mingle. I believe thinking of WRE as a method for interpersonal justification makes the most sense. The desire we have to justify our actions plainly originates in the fact that we live our lives interpersonally, cooperating with other people and making joint decisions. Using WRE as a model of personal justification might sometimes be useful, but our understanding of this use would probably be parasitic on our understanding of interpersonal justification. Impersonal justification finally makes the most sense if we conceive it as a form of interpersonal justification in a maximally general context, where we are required to make our actions acceptable to everyone (this is in line, I believe, with Nielsen’s (1993:242-3) and DeGrazia’s (1996:17-8, including n.7) views of WRE). Hare’s and Brandt’s objections are adequate and serious against WRE if conceived as a method for personal justification only. It is obvious that at least Hare, in labelling Rawls’ view “subjectivism”, interprets him as giving an account of personal justification. Rawls and Daniels are far from innocent in spreading this, in my view,
inadequate construal of WRE though (see section 2, THE WHO AND WHERE OF REFLECTIVE EQUILIBRIUM).

Personally, I think there is something very dubious behind the objection voiced by Hare and Brandt (among others). The way I think Hare and Brandt understand ‘people’s opinions’ and ‘tradition’ is misleading, because they construe them as rigid entities. This construal makes it look suspect to rely on these opinions and traditions, as well as making it difficult to imagine how to overcome them. But I believe opinions and traditions are more dynamic and fluid than this picture suggests (although they are unquestionably subject to considerable inertia). Every effort to “step outside our own tradition” inevitably has to take off from that very tradition, and as we take the step, we transform our tradition to engulf whatever new insights (or neglects) we set our feet upon. How does this bear on the credibility of CMJs? First of all, not taking our tradition seriously (in a broad inclusive sense) will never be an option. And since we are facing moral dilemmas and problems coming from a tradition (again, broadly conceived), we have to recognise that the very tools of moral inquiry – ideas, principles and theories – are part of what defines moral problems and dilemmas in the first place. The reason we should take CMJs seriously is that we recognise them as moral judgements, our or somebody else’s. It seems to me that if ethicists do not take these judgements seriously, they are not doing ethics at all. Moral inquiry has to be a reflexive enterprise since ideas about morality form an integral part of the social ecology (ecologies?) within which moral inquiry should help human beings flourish (and to some extent, other animals too)\(^{15}\). If our CMJs, instead of being judgements about the properties of an inquiry-independent moral reality, are (partly) constitutive of the moral reality we are exploring (inventing?), then the case against their credibility becomes far from conclusive. We have no reason to immediately discredit prejudices about moral matters (in the sense of previously made judgements entrenched in – and by – tradition) on the grounds that they have no contact with reality, because they are part of reality (at least the reality moral philosophers take an interest in). Daniels suggests a defence against the ‘no credibility’ objection in a similar vein I think, when he argues that the objection relies on an implausible analogy between CMJs and observation reports in science. The analogy is misleading because “[r]ightness and wrongness, or justice and injustice, are unlikely to play a role analogous to that played by observational properties in the causal-reliability stories we tell ourselves concerning observation reports.” (1996, 31).

Now, Brandt and Hare might reply that this is not enough. When we make moral judgements, they might reply, we do not merely make claims about what is right and wrong within (or between) groups of people or traditions. Moral talk is talk not about what we think is interpersonally justifiable within a tradition, but about what is impersonally justifiable universally. Perhaps belief in this kind of justification is mistaken and deceptive, but still, this is the nature of moral talk (an ‘error theory’ of moral semantics). But it is at least possible to doubt the existence of a universal moral language that necessarily expresses claims of impersonal justification (perhaps this is only a property of moral language in a certain Judeo-Christian moral tradition). Perhaps moral talk without such claims wouldn’t be moral talk at all according to Hare and Brandt. Well, then so be it. We are not obliged to accept Hare and Brandt’s circumscription of morality. If one accepts their strict conception of morality and justification, then I think one is bound to be disappointed with moral philosophy and in danger of becoming a thoroughgoing
moral sceptic. However, since our attitudes toward morals and their justification are likely to be subject to change, this scepticism need not be our final fate.

What can we say about the difference between NRE and WRE in light of my defence of CMJs? This ought to be clear I think. Given the aim to include as many NREs as possible when choosing what NRE to accept in WRE, and to consider every possible argument for or against all those partial equilibria before making our choice (keeping in mind that this is an ideal), I think the CMJs held in a state of WRE can reasonably claim to be more than merely a reflection of “class or cultural background, self-interest, or historical accident” (Daniels 1996, 21). With the method of NRE on the other hand, one is never confronted with rationally choosing between different equally coherent NREs, and it seems that we simply end up with more or less what we started with in terms of credibility (so if “garbage in”, then “garbage out”).

Finally, let us momentarily assume that the ‘no credibility’ objection is sound in the sense that CMJs need to show that they reliably track objective moral truth (thus, we go along with the requirement that justified CMJs must be justified impersonally). Can background theories in WRE neutralise the threat of the ‘no credibility’ argument? Margaret Holmgren (1989) has argued, correctly in my view, that an affirmative answer to this question must rest on the plausibility of the three versions of claim (1). In both NRE and WRE we have to give our CMJs at least prima facie credibility. Without credible CMJs, we have no reason to accept moral principles that systematise them in NRE. Neither does the method of WRE make sense without prima facie credibility of our CMJs. Nonmoral background theories could perhaps give support on their own to moral principles in WRE, but proponents of WRE usually believe this to be impossible. Anyhow, proponents of WRE definitely do not claim that the method requires such foundations to work as a method of justification. The conclusion drawn by Holmgren is correct – CMJs must at least be given prima facie credibility.

**REVEALING STRUCTURAL COMPLEXITY**

In *The Independence of Moral Theory* (1975) Rawls characterise moral theory, the central subject of moral philosophy, in the following way:

“Moral theory is the study of substantive moral conceptions, that is, the study of how the basic notions of the right, the good, and moral worth may be arranged to form different moral structures. Moral theory tries to identify the chief similarities and differences between these structures and to characterize the way in which they are related to our moral sensibilities and natural attitudes, and to determine the conditions they must satisfy if they are to play their expected role in human life.” (Rawls 1999, 286).

Rawls here describes the business of moral philosophers not as a normative enterprise, but as a purely descriptive one, a kind of abstract moral anthropology. The role of the philosopher here is not the one of the situated moral agent, attempting to adopt moral beliefs which are justified, but one of a third person, the observing moral theorist (Rawls 1999, 288; also Holmgren 1987, 108). Rawls further suggests that it is the framework of WRE we should use to explore the structures of various moral conceptions (1999, 289). Like Rawls, Daniels clearly recognises WRE both as a method of justification and a framework for doing moral theory in the Rawlsian sense. He refers to the latter as the “modest role” of WRE, and the former as the “daring role” of WRE (1996, 60).
WRE in its modest role is “analytic and explicative”, while it in its daring role “serves as a basis for a coherence account of justification” (60).

The claim (4) then, is that when doing moral theory (in Rawls’ sense), WRE provides a much better framework than NRE for studying the structure of the basic moral notions embodied in moral conceptions and for drawing out the connections between these conceptions and our psychological and sociological circumstances. This does not seem surprising considering the fact that using the method of NRE, there does not seem to be any explicit checking for stability involved, except perhaps, as embedded in the filtering process. If we want to describe how different moral theories are related to our “moral sensibilities and natural attitudes” then theories of those sensibilities and attitudes are of course needed.

An important point here, already brought up in section 2, is that Rawls and Daniels assume that the there is a structure of different moral notions in our moral conceptions (there is the CMJs in (a), and various CMJs related to different moral notions in (a’)). If there is no such structure, if there are no different kinds of moral judgements, no different moral notions, only different individual moral judgements, then the advantage of using WRE as a framework for doing Rawlsian moral theory is not so evident. But using the simple framework of NRE, one could argue, we would be blind to any internal heterogeneity in our moral conceptions. Additionally, any connections between nonmoral background theories and moral conceptions would be kept out of our sight.

Daniels’ suggestion that WRE could fruitfully be used for describing different moral conceptions is uncontroversial I think. What exactly would we get out of the descriptions we end up with? Daniels reasons that we can get three things. First of all, (α) WRE helps us sort out which moral disagreements are truly moral and which ones are - at the “bottom” - nonmoral disagreements about ”facts” (with moral repercussions). WRE can also be used as a tool to understand (β) why people treat certain arguments and considerations as relevant or important, and other not. And finally, (γ) if we can trace a substantive part of our moral disagreements back to disagreements about facts (in background theories), it might be the case that these disagreements are more tractable than disagreements about moral principles and judgements.

**MAKING REFLECTIVE EQUILIBRIUM NORMATIVE**

According to Daniels, NRE, unlike WRE, is limited to a modest role. NRE cannot help us in choosing between different moral conceptions, or in other words, it cannot justify a moral conception. While NRE is accurately accused of being merely descriptive, in the sense of being a tool for describing rather than a tool for justifying, WRE is more properly interpreted as normative or justificatory (Daniels 1996, ch. 4, 2003, sect. 3.1). For Daniels, the modest role of WRE is really only a stepping-stone towards its more daring role of a justificatory device. In WRE, moral conceptions are not only being refined with narrow coherence constraints and then described, but they are compared and appraised with the help of background theories in order for us to choose between them. Assuming that people are rational and willing to listen to arguments, this will make them choose, out of many plausible options, the best (most coherent) moral theory. But clearly, to be able to make an informed choice between different moral conceptions, we need to explicate the available options. So WRE needs to fulfil its modest role in order to be able to
play the more daring one. While the power to inspect and compare different moral conceptions from the viewpoint of different background theories gives WRE more justificatory power than NRE, it is going to far to portray NRE as only capable of describing moral conceptions.

NRE should not be depicted as merely descriptive because constraints of coherence are normative, and – if they are themselves reasonable – gives the method some justificatory force (unless one claims that the coherence constraints are descriptions of constraints of cognitive processes given by nature, but this is absurd). The filter used in the filtering process also embodies strong normative elements, and this again gives the method some justificatory force. I agree with Daniels though, that NRE is vulnerable to the ‘no credibility’ argument in a way that WRE is not. The narrow scope of NRE limits the critical potential of the method. Without a story about the reliability of our CMJs NRE cannot be accepted as a satisfying method of justification. In NRE, we are not invited to (indeed cannot) choose between different (within the framework of NRE) equally coherent packages of beliefs. We have to trust luck that we walk the right path in the reflective process. We have to take the right steps of adjustment to end up with the best coherent package of beliefs. To be this lucky, we better find ourselves with a set of initial CMJs that can get us there.

**SCIENCE PIGGYBACK**

While Daniels denies that any moral theory is derivable from or reducible to science (Daniels 1996, 23, 29-30, 49, 61), his account of moral justification in several places points out, and draws on, an analogy between ethics and science (22, 24-5, 33, 37-39, 74-7). By drawing this analogy, Daniels wants to show where moral theory might end up if we embrace the coherence account of justification embodied in WRE. In the domain of science, this account has led to an equilibrium containing a causal story telling us why our observational data has credibility (a neurological/psychological story) as well as a convergence from different starting points on a widespread consensus about scientific theory and method. Of course, there are plenty of disagreements in science but there are at least often agreement on what in principle it would take to settle these (first-order) disagreements. Although CMJs are different from observations in science in some respects, in Daniels view (and in the view of practically every philosopher of science alive today), they are more similar than one might think. Observations, like CMJs, are “theory-laden” and open to re-interpretation and the credibility we grant them are not due to some inherent quality in the observations themselves but due to a casual story (itself accepted in light of coherence considerations) about how they can give us information about the world. Since the domains of ethics and science are at least prima facie similar and the method of WRE has worked towards convergence and consensus in science, Daniels suggests it is well worth trying to implement this method in ethics as well. By this, Daniels hopes that the kind of progress demonstrated in science might be “imitated” in ethics. According to the science piggyback claim (5) then, if we use a method analogous to the one used in science (WRE), progress analogous to the one witnessed in science might result.

With the ethics-science analogy in mind, Daniels speculative suggestion is that if moral philosophy now goes into WRE mode, future moral philosophers might get a convincing story telling us that our CMJs in RE are indeed credible. Assuming we accept the presuppositions of the ‘no credibility’ objection (outlined in LESS RELIANCE ON CONSIDERED MORAL JUDGEMENTS), hoping for, and trying to tell, this kind of story is all we can do in order to hold off this objection.
We need an account of how CMJs could provide us with (revisable) information about moral properties, or at least argue that we have reasons at this point not to rule out the possibility of such an account. Daniels thinks we have such reasons, and although he concedes that the ‘no credibility’ argument puts the burden of proof on the WRE-theorist, he does not find the argument conclusive. According to Daniels, it is possible that when we have constructed a moral theory set in a state of WRE, then the kind of story about credibility we need will be available (1996, 31). However this story goes, we will have to wait to hear it until we have a moral theory in a state of WRE. But “being in a state of WRE” is not a Boolean variable, it is a matter or more or less. We may ask Daniels what degree of coherence we need to achieve before our credibility story emerges (minimally, it must be coherent to a degree beyond all our currently available moral theories, since Daniels admits that credibility stories about CMJs are lacking). Although this unclarity is not fatal, Daniels “burden-of-proof” response is far from satisfactory. Such a response can after all be used to argue for all sorts of weird beliefs. For example, a moral foundationalist could, using the “burden-of-proof” response, argue that a foundationalist method is promising since the arguments against foundationalism are not entirely conclusive, although no adequate foundations for a moral theory have been found so far, we have no reason to doubt that some might be forthcoming.

Another strategy against Daniels’ science piggyback suggestion found in the literature (Haslett 1987, 308) is to argue that in the case of empirical observations there exist an independent reality, a noumenal realm, which puts some limits to the theory-ladenness and plasticity of observations (and observational reports). But since we have no reason to believe that such a constraining inquiry-independent moral reality exists, we have no reason to trust our CMJs either. Though I share the intuition behind this argument, I think it either begs the question or is in need of heavy back-up. The argument either (I) merely reiterates what Daniels himself has asserted, that we have a story in a state of “nonmoral WRE” about the credibility of observations but no such story about CMJs, or (II) constitutes an argument based on a foundationalist epistemology of science not shared by Daniels. I do not think (I) is very interesting for obvious reasons, and (II) is in need of arguments for favouring a foundationalist epistemology instead of Daniels’ radically coherentist account of nonmoral justification. If one is a coherentist about nonmoral justification, one should only take ones intuitions about independent realities seriously if these intuitions cohere well with all our other beliefs (for example beliefs about the workings of perception). So either way, without arguments favouring foundationalism, we are back where Daniels left us.

But the proponent of WRE, even if accepting the fall of the CMJ-observation analogy, might still argue that there are other commonalities between ethics and science that she can use as a base for the science piggyback claim. There is according to Daniels “a strong similarity in the way coherence constraints on theory acceptance (or justification) operate in the two domains [ethics and science] …” (1996, 33). So, even if CMJs are not credible in the way observations are, the kind of connections we seek to establish between the beliefs in our moral and nonmoral WRE-networks are similar according to Daniels, as are the rules or criteria which govern when and how adjustments should be made in order to achieve greater coherence. A central question here is what the precise coherence constraints are in science and ethics (see section 3). Furthermore: why should there be a “strong similarity” between science and ethics in this respect? Theoretical virtues like simplicity and connectedness are reasonable in scientific inquiry because the world
explored by these disciplines are one and the same. We are there seeking a covering, unifying and consistent account of this one world. But, barring an independent moral reality, what can provide a rationale for employing the same constraints in ethics? Philosophers with anti-theoretical inclinations doubt that such a rationale can be provided, arguing that moral philosophy should not be modelled on science (DeGrazia 1996, 1816). This fascinating issue is unfortunately beyond the scope of this thesis and I will not discuss it further. Personally, I definitely believe there is room for theory building in ethics. But I must confess that I am not sure how to answer the anti-theoretical challenge without becoming a moral realist (in the sense of positing a moral reality that is independent of moral inquiry), something I would rather avoid.

I have simply assumed that Daniels’ claim that a method of justification very similar to WRE operates in science. Perhaps this does not square well with the actual practice of science. Perhaps the method used usually resembles NRE rather than WRE, drawing support and evidence from a very limited base, isolated from wider considerations of coherence by narrow disciplinary boundaries (set up by office walls and technical terminologies). But it seems correct that justification based on a wider scope of considerations is at least desirable in science. Postponing an answer to the anti-theoretical challenge until later, I believe thick coherence requirements are general reasonable requirements in any kind of inquiry, not only in science. If this is true, then the suggestion that we should try to piggyback our way to progress on an analogy with science do not really make sense unless we hope to establish the truth of moral realism. But for this establishment to succeed, the WRE-realist needs a story about the credibility of our CMJs. I believe it is extremely unlikely that such a story is forthcoming and I suggest we should settle for being reasonable and doing good within a world without independent moral properties.

5. WIDE AND NARROW REFLECTIVE EQUILIBRIUM — AN ASSESSMENT

We now have some idea about what background theories are asked to do in RE. Background theories should provide additional justificatory support, interrelate CMJs belonging to different “subdomains of morality”, reduce our reliance on CMJs, enlarge the “circle of justification”, empower us to criticise and revise our own traditional moral views, reveal the structural complexity inherent in moral conceptions, and generally provide leverage for choosing between different NREs. Whether we allow background theories to perform these tasks depend on what kind of relations we deem to be possible between background theories and CMJs, and whether we think relations of these types contribute to the coherence of belief-sets. I have already discussed the possible relations between background beliefs and CMJs in section 2, BACKGROUND THEORIES AND COHERENCE, but some puzzling issues concerning moral background theories and their place in WRE remain to be discussed.

I will first briefly discuss these issues in THE STRUCTURE OF MORALITY and THE ENTANGLEMENT OF THE MORAL AND THE NONMORAL, and point out some limitations of my treatment of them. I then try to assess NRE and WRE in light of the considerations I have brought up in this thesis. The distinction between NRE and WRE is, as I argued in section 2, not always clear-cut.
Depending on the details of how we characterise the methods the distinction (and our choice of method) will be perfectly clear or collapse altogether (in which case the choice of method is illusory). This discussion is taken up again in Variety of Wide and Narrow Reflective Equilibrium. Finally, I discuss what the possible variants of NRE and WRE can promise or provide in terms of convergence/consensus, justification and truth. I argue that WRE, unlike NRE is likely, but not guaranteed, to lead to convergence/consensus, at least if we follow the common ways of distinguishing WRE from NRE. WRE furthermore, again unlike NRE, can give us a form of (interpersonal) justification. Finally, I argue that talk about the outcome of either WRE or NRE in terms of moral truth are best dropped and abandoned.

But first, how should we understand the place of moral background theories in WRE? This issue is clearly important since whether such moral background theories have a place in WRE or not seems to be an issue on which proponents of WRE and NRE are divided. To understand the role of moral background theories two issues that have surfaced earlier in the thesis must be brought up again. One is the issue of different moral notions and the structure of morality: What coherence-contributing relations are possible between different kinds of moral notions? How are (a) and (a’) related? The other issue is the nature of the relation between the CMJs constraining some background theories (a’) and the nonmoral components of those theories (c). What do Daniels mean when he says that some CMJs (a’) “may play an ineliminable role constraining the acceptance of background (level III) theories” in WRE? (Daniels, 1996:29-30) How are (a’) and (c) related? While, I do not have much to say about these difficult questions, I at least present a few tentative suggestions below.

THE STRUCTURE OF MORALITY

First, concerning the relation between moral notions, the examples of different notions given by Daniels are “justice”, “fairness”, “which features of persons are morally central or relevant”, and “rights and entitlements” (Daniels, 1996:49-50). Presumably, Daniels and Rawls do not believe judgements falling under these various notions can all be simply collected in (a). This would be desirable if possible though, since it would make whatever set of principles (b) we end up with much more comprehensive and powerful (in the sense of being better at generating plausible CMJs and decisions in new problematic situations). If our ambition is to grasp a large part of morality with our moral theory, then perhaps the three sets of CMJs (a), (b) and (a’) are far too limiting. We would possibly need more sets to capture the fragmented heterogeneous structure of morality, say at least (a’’), (a’’’), and (a’’’’) to capture the notions mentioned by Daniels. The virtue of Rawls’ theory of justice however, is that despite the (by Rawls) assumed heterogeneous structure of morality it is able to draw to together many different clusters of CMJs around the contract apparatus, making the theory highly connected (and connectedness is usually a part of the concept of coherence (see section 2, The Concept of Coherence)). Of course, different dimensions of the concept of coherence might end up in tension here. Although trying to massage all the CMJs of all notions under some kind of supra-notion (say, utility) might be desirable because we end with a very simple and manageable moral theory, it will suffer from implausible output in some situations. If what I imagine here is on the right track, then an important role of some background theories will be their ability to bring different subdomains of morality into contact with each other. I am painfully aware of the fact that this is terribly vague and only suggestive at best, but it is the only way in which I can make some sense of it. A further question
to ask, which I am not even trying to answer, is if there is some kind of general property of these background theories that allow them to play this integrating role.

**The Entanglement of the Moral and the Nonmoral**

What about the relation between (a’) and (c)? If my characterisation of the structure of the moral domain is on the right track then the relations between CMJs and nonmoral background beliefs could be constrained by this structure. Because of the structure, some kinds of CMJs can only be related to some limited domains of facts. For example, judgements about “which features of persons are morally central or relevant” are (unsurprisingly) not relevantly related to any fact, but only to facts about beings capable of being persons. But why not introduce only the nonmoral beliefs into our WRE, leaving the constraining CMJs in (a’) (or (a’’), (a’’’), etc) aside? Perhaps certain background theories are only true if certain CMJs are held. This could be the case with psychological, political or economic theories since people plainly think and behave differently depending on what CMJs they hold. This is uncontroversial but completely beside the point. In this case the CMJs are not normative but “merely” constitutive parts of the (moral) agents the theories trying describe and predict. To accept the theory is obviously not to accept the CMJs that are part of what the theory is describing. The only alternative left, as far as I can see, is that some descriptions cannot be given without simultaneously holding an accompanying CMJ, and vice versa – that some CMJs cannot be held without a simultaneous commitment to some description. Fully elaborating this idea (similar to the idea of ‘thick concepts’) and judging whether positing such “CMJ/fact-twins” make sense are beyond the scope this thesis.

**Learning by Applying**

I have in this thesis only mentioned some examples of background theories taken from the literature, and only on an abstract level explained what kind of relations between background theories and CMJs proponents of WRE take to be relevant. But much more could be said about this I think. The abstract formal elements of coherence I have presented only provide some basic scaffolding for grasping what background theories can do in WRE. While ‘connectedness’, ‘comprehensiveness’, and ‘explanatory relations’ give some intuitive sense of the nature of coherence, these concepts need to be filled out by case studies of WRE in practice to be thoroughly understood. We need to look closer at the particular background theories themselves. Given the nature of the method of WRE, the abstract account I have given can only go so far. This follows from the fact that the relevance of a background theory in WRE is partly determined by the content of the CMJs in the WRE-network itself (Sayre-McCord, 1996:151). For example, consider the theoretical virtue (element of coherence) proposed by DeGrazia (1996:2) called ‘argumentative support’. If a theory has argumentative support then all demarcations drawn by CMJs (for example that we should be treated with respect but not them) are backed up by adequate reasons and arguments. But – and this illustrates my point about the limits of my abstract exploration – what counts as an adequate reason or an adequate argument is to a large part determined by the contents of other background theories in (c) and the constraints set by other CMJs (Daniels, 1996:24). (This means that what I have said about meta-beliefs and meta-principles, that they are fixed outside the WRE-network, is a truth with modification. Although they are properly understood as unrevisable within the network, they have some considerable degree of “semantic plasticity”. The contents of the meta-beliefs can be fleshed out and determined only when examined in the context of an actual WRE-network.)
Varieties of Wide and Narrow Reflective Equilibrium

The two critical steps of the method of RE, whether wide or narrow, are the process of filtration and the process of reflection. If followed faithfully, these two steps are supposed to make one adopt moral theories and beliefs that are justified. Now, the defining aspects of these two steps are the filter and the concept of coherence. Many discussions about RE, including this one, spend much time on structurally describing the method and the various kinds components, steps, and operations of it, without much discussion about either the filter or the concept of coherence. But if we want to assess the methods, I think it will be necessary at some point to finally take a closer look at these two aspects.

NRE and WRE can – depending on what filter we use and what we pack into the concept of coherence – either become almost indistinguishable twins or two complete strangers. Suppose we are using the method NRE. Let us assume the filter we use is very strict, only letting those judgements pass which has been formed under consideration of all relevant facts by someone with certain requisite moral sensibilities and past experiences. The concept of coherence we employ in appraising different moral theories is thick. It includes notions like connectedness, comprehensiveness, clarity and simplicity. Using an extremely absorbing filter and a thick concept of coherence makes NRE virtually equivalent to WRE. But this does not mean that NRE and WRE would be equally adequate here. WRE might be a better description of the coherentist account of justification we are interested in – when background beliefs are put in a separate distinct set in the WRE-network their role might be more visible and clear than when embedded in the filter of the filtering process. Having more relaxed requirements on what should count as CMJs is furthermore preferable to stricter requirements. Stricter requirements will make the method less useful since fewer people are likely to agree to use the method.

On a much more sparse conception of NRE on the other hand, were the filter is more relaxed and what count as coherence-contributing relations between beliefs are few other than relations that uphold logical consistency, we get a completely different picture. This thin conception of coherence would exclude any interesting role for nonmoral beliefs to play in moral justification. This version of NRE would plainly make the method inadequate for justification. There are of course more possibilities. Margaret Holmgren’s (1989) view of NRE has a stricter filtering process but a minimal conception of coherence, and Michael DePaul (1993) has completely got rid of the filtering process as a distinct step in his version of WRE, relying instead on a thick conception of coherence.

Justification

When one allows relations of explanatory support into the concept of coherence, allowing the nonmoral beliefs to contribute both to the connectedness and to the comprehensiveness of moral conceptions, then the method of RE should, I think – to avoid confusion – be referred to as WRE. With the introduction of background theories comes the ability of arguing for and against states of NRE from outside those states themselves, i.e. the possibility of rationally choosing the moral theory to accept from among a range of options (all undifferentiated in NRE since their internal coherence is equal). This rational choice situation is what makes WRE adequate as a method of justification in my view. What about a proponent of WRE who only recognise the contributions of moral background theories (someone who only accepts the first (i) version of claim (1))? Given that the internal structure of the moral domain required for this version of WRE does exist,
perhaps the method would be adequate as a method of moral justification. But as Daniels himself points out, it is not clear how the inclusion of these moral background theories would help us much, unless we can provide some reasons why these background theories would be in less need of justification than the theory we are trying to construct (1996, 26). Of course, whether WRE has the capacity to justify moral beliefs and theories or not depend on what one means by justification. As I have already stated, I am suspicious about a strict requirement for impersonal justification in ethics. I recognise it only as a honourable ideal, as does the method of WRE. For Daniels, justification and theory acceptance is synonymous. To justify something is to show that it is rational to accept this something in light of the existence of all other coherent alternatives.

**Convergence/consensus**

Even if the only justification good enough for us is impersonal justification and we acknowledge that we cannot get this with the method of WRE (which is true) then perhaps the use of WRE can at least lead to a convergence of different moral outlooks toward a large moral consensus. Daniels thinks that this is a plausible consequence of using WRE. Using WRE will reveal the structural complexity of moral conceptions. This new knowledge will help us see where we agree and disagree and sometimes show how disagreements can be settled. But remember that the claim is not that given this knowledge of the structural complexity of moral conceptions, people’s moral views will start to converge and eventually reach consensus. The claim is rather that given our newfound insights in moral anthropology, and our willingness to follow the method WRE (we want nonarbitrary convergence), we are likely to start converging on a moral consensus. That some such converging tendencies will result if a lot of apparently moral disagreements are traced back to underlying disagreements about facts (which can be settled at least in principle) seems reasonable. Whether most or all moral disagreements can be traced back to nonmoral ones is a question that cannot be answered before extensive use have been made of the method of WRE. But I believe, as does Daniels, that whatever moral theories WRE lead to, they will to a significant degree be underdetermined by nonmoral beliefs. When moral disagreements (in the foreground) are traced back to moral background theories, it is less clear that there should be some kind of convergence as a result.

Even if I am slightly reluctant to concede that WRE will lead to convergence/consensus, it is an even less likely outcome of NRE (assuming that nonmoral beliefs do not play a part in NRE). Without the knowledge about the structure of moral conceptions facilitated by WRE, moral disagreement will not be traced back anywhere. The moral theories resulting from the method of NRE will not only be underdetermined by nonmoral beliefs but will be completely unconstrained by such beliefs. Furthermore, given the lack of constraints, widespread adoption of NRE as a method for constructing and refining moral theories could just as well lead to a divergence of moral views and increase in disagreement.

**Truth**

Daniels suggests that if convergence comes as a result of WRE, this is evidence for the existence of some moral truths (1996, 35). While it is plausible I think to acknowledge that such an evidential relation exists, I believe that the burden is on Daniels to show what ‘moral truth’ means, why it would make a difference if it existed, and what it is. Isn’t it more plausible, less metaphysically suspect and more comprehensible, to interpret whatever convergence WRE leads to as reflections of universal human sensibilities and appetites?
In this thesis I have presented some of the motivation behind making the transition from NRE to WRE by including background theories. Many of these are formulated in the literature as arguments defending RE from the weaknesses of NRE. I have also problematised the distinction between the two methods by pointing out that NRE many times cannot do without some connecting descriptive “background beliefs”, since the degree of coherence among our general and particular CMJs cannot be adequately assessed otherwise. A proponent of NRE might also design the filtering process in such a way that all facts with relevance for what CMJs (a) should contain are “automatically” taken into consideration. Construing the RE in this way though makes the background beliefs recede to far off into the background. An advantage of giving background theories and beliefs a recognised explicit place in our method is that the relevance of them for our moral views, and the way they get their relevance, are illuminated.

Generally, I think the reasons I have found for including nonmoral background theories in RE are persuasive. Nonmoral background theories and beliefs can explain and illuminate our CMJs. Our CMJs in turn, point out which nonmoral beliefs are morally relevant. There is no dubious crossings of the “is”/”ought” gap involved. Including nonmoral background theories can give additional justificatory support to moral theories. While it is clear that the additional CMJs in (a’) brought in with the background theories do give additional support, it is not clear why these CMJs shouldn’t belong to the original set of CMJs (a). I have suggested, correctly I believe, that the reason proponents of WRE think that the CMJs in (a’) are unavailable in NRE is that they have a view of ethics as more fragmented, containing different kinds of moral judgements, while defenders of NRE seem to view ethics as more uniform. Whether WRE or NRE is the best method for taking care of the CMJs connected to moral background theories will be unclear until we know more about how best to conceptualise the moral domain. Another major problem hit upon is how to characterise the relation between the moral and the nonmoral components of background theories constrained by CMJs. Here, I had no real answers or solutions, but it is at least clear that the relation must be different from the relation between nonmoral background theories and the CMJs in (a). Still another problematic and important conceptual issue is how to understand coherence. I think seeing the different ways of understanding this concept is important for understanding the debate between some critics of WRE and defenders of NRE. Whether one opts for WRE or NRE as a method of justification seems to be guided by what one means by coherence (or what one should mean). A thick view of coherence makes it natural to embrace WRE while a thin view of coherence makes NRE the most likely method of choice.

Some of my responses to criticism of RE have been dependent on the fact that I think that it is most fruitful to understand justification as something occurring between persons. I believe impersonal justification is strictly speaking impossible and personal justification only makes sense in light of our understanding of interpersonal justification. Arguing for (or perhaps merely presupposing?) this particular understanding justification formed the backbone of a response to the perhaps most serious objection against RE, namely the ‘no credibility’ objection. My scepticism of talk of moral truth and inquiry-independent moral properties has made me reluctant to accept Daniels’ own main defence against (or rather, escape from) this objection, which rests
on the burden-of-proof strategy. Whether one believes NRE or WRE to be viable models of moral justification probably depends on one's stance in such metaethical debates. I have barely touched on how these debates impact on NRE and WRE (although, I have argued that NRE seems to require moral realism to make sense, while WRE does not).

I hope by now that it is clear that my sympathies lie with WRE, rather than NRE. Of course, as I have duly acknowledged, there are problematic and difficult issue left to sort out. But this is enough for now.

NOTES

1 I would like to thank Bo Petersson, my supervisor, for his contagious enthusiasm and support, and Martin Andersson, my opponent, for helpful suggestions and criticism.

2 What criteria we use at this first stage will determine which judgements qualify as considered and hence have a significant influence on what we will get out of the method as a whole. How to construct this filter without building in obvious biases while at the same time avoid that appalling moral views comes out as being justified by some groups of people is a difficult problem that I completely ignore in this thesis. For discussion about this "filtering process" and some problems associated with it, see (Sencerz, 1985).

3 Rawls does not explicitly use the term wide and narrow RE here, but does so in The Independence of Moral Theory (1999/75:289), and makes the distinction along the same line as in A Theory of Justice.

4 Holmgren (1987, 112) for example argues that a moral judgement made when a moral agent is "ignorant of relevant facts" should not qualify as a CMJ.

5 In my characterisation of foundationalism and coherentism I am largely following Timmons (1987).

6 Nielsen also imply that there are several equilibria involved in WRE when he writes that "[t]he overarching aim is to get the best fit possible of all these diverse elements [of WRE], elements which are themselves the best warrented elements in their respecitve domains." (1993:322).

7 To assess the stability of a conception of justice (or morality) is to check whether a conception of justice (or morality) governing a community would cause persons developing in the community to embrace and reproduce the conception, thereby securing the conception’s hold of the community. To do this assessment, so called level IV theories are used (this could for example be various social theories and theories about moral development and psychology).

8 Similarly, Nielsen divides background theories into “(c) a cluster of background theories including most centrally moral theories and social theories, among them social theories that are quite definitely empirical theories about our social world and how we function in it, and (d) an empirically based, broadly scientific conception and account of human nature.” (Nielsen, 1991:200).

9 I recently emailed Daniels and asked him whether, according to him, nonmoral background theories were revisable in WRE and this was his brief reply:

   “I imagine that the process for revising empirical beliefs and theories that play a role as background theories will involve standard measures; I am not imagining that we would on the whole think our description of the
world should be revised to fit our views of what we think is moral. But this is not a foundationalist view of empirical beliefs or theories themselves; nor do I think empirical theories stand in a foundational relationship to normative ones. What is true about the empirical world, however, does often constrain what we ought to do, if only because it affects what we can do and what kinds of systems are stable. Sorry to be so brief.” (Email correspondence, May 26, 2003, my emphasis)

Judging from this reply, Daniels does not think that empirical and moral beliefs are on equal footing in WRE, but he seems unwilling to exclude the possibility that empirical beliefs could be justifiably revised in rare circumstances (note the “on the whole” that I emphasised). What these circumstances could be we are left to figure out for ourselves though.

10 This “law” says that ethical conclusions cannot be drawn from only nonmoral premisses.

11 “The background theories might...not incorporate the same type of moral notions as are employed by the principles and those considered judgments relevant to “testing” the principles” (Daniels 1996, 23), “initial level I considered judgments about justice...against initial level III judgments about fair procedures or about which features of persons are morally central or relevant” (30), “if the background theories in (c) incorporate different moral notions (say, fairness and certain claims about the person) from those (say, rights and entitlements) employed by the principles in (b) and judgments in (a).” (49-50).

12 Although White does not explicitly discuss the method of RE in (1998), but rather Quine’s view of moral values, I straightforwardly take his discussion as one about RE. In a later work, A Philosophy of Culture: The Scope of Holistic Pragmatism (Princeton University Press, 2002), he discusses Quine’s holism and Rawls’ RE approach together. Furthermore, Rawls himself refers to the views of both Quine and White when characterising his own view in A Theory of Justice (1971, 579n33).

13 Daniels uses the expression “a circle of justificatory support wider than simple matching of [(b) and (a) in NRE]” (1996, 59). Nielsen declares that it is important “that we have a large circle involving many considerations rather than a small one.” (1993, 203).

14 I interpret the objection voiced by Hare, Brandt, and others as an objection to relying exclusively on any kind of CMJ. As my opponent Martin Andersson pointed out, one could also interpret the ‘no credibility’ objection more narrowly as an objection to relying exclusively on particular CMJs only.

15 I very much like the idea of considering ethics as part of the study of human ecology. See Owen Flanagan’s “Ethics Naturalized: Ethics as Human Ecology” in his Self Expressions: Mind, Morals, and the Meaning of Life (Oxford: Oxford University Press, 1996) for some sketchy (but good) ideas regarding this.

16 DeGrazia, who disagrees with the anti-theorists, refers to Bernard Williams, Richard Rorty, and Alasdair MacIntyre).

17 One can of course argue for the coherence constraints on moral or practical grounds (Daniels, 1996:74-5, Holmgren, 1987, Sayre-McCord, 1996:143-4), but then these grounds can be questioned and there can be no appeal to coherence for justifying these grounds themselves.
REFERENCES


