

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

PETER RAILTON'S MORAL REALISM

Xiaoyu Zhu

A Thesis

in

The Department

of

Philosophy

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts at
Concordia University
Montreal, Quebec, Canada.

August, 2000

© Xiaoyu Zhu, 2000



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

Our file *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-54272-6

Canada

ABSTRACT

Peter Railton's Moral Realism

Xiaoyu Zhu

This study is a critical examination of American philosopher Peter Railton's moral realism. The emphasis is placed on his synthetic naturalistic account. He holds that there are moral propositions which can be true or false; moral facts and properties are natural facts and properties; we can come to know or justify moral judgments through wide reflective equilibrium. I shall try to argue that his moral theory is a generally successful attempt to escape the Moorean charge of 'naturalistic fallacy' when advocating ethical naturalism. I will discuss both his reductionist and non-reductionist accounts and suggest that the latter is a better approach for a moral realist to justify moral beliefs. I will also critically discuss his arguments for instrumental rationality, ideal observer theory, internalism, justification of morality and his objections to non-cognitivism.

ACKNOWLEDGEMENTS

I would like to thank Professor Kai Nielsen for his wisdom, justice, kindness, and generosity. He is such a good teacher that, from him, I have learned so much, not only about philosophy but also about life. During these three years of my study at Concordia, I have always been inspired by his enthusiasm for understanding humanities and his passion for building up a better world. It was such a great pleasure as well as a wonderful experience to have had the opportunity to study under his supervision.

I also would like to thank my husband, Yongqiang, for his understanding and support of my 'unpractical' pursuit of studying philosophy.

TABLE OF CONTENTS

INTRODUCTION.....	1
1. MORAL REALISM.....	1
2. MOORE’S OPEN QUESTION ARGUMENT.....	2
CHAPTER 1: NON-MORAL VALUES.....	6
1. REJECTION OF FACT/VALUE DISTINCTION.....	6
(1) <i>Rational Determinability</i>	7
(2) <i>Value Absolutism</i>	8
(3) <i>The Thin/Thick Moral Concepts</i>	11
2. THE REDUCTION FROM VALUE TO FACT.....	13
(1) <i>Railton’s Account</i>	13
(2) <i>Critiques of Railton’s Account</i>	22
CHAPTER 2: MORAL VALUES.....	35
1. RAILTON’S ACCOUNT.....	35
(1) <i>Moral Facts</i>	35
(2) <i>Social Rationality</i>	35
(3) <i>The Inescapability of Morality</i>	37
2. CRITIQUES OF RAILTON’S ACCOUNT.....	39
(1) <i>The Problem of Railton’s Justification of Morality</i>	39
(2) <i>The Problem of Moral Convergence</i>	42
CHAPTER 3: THE OBJECTIVITY AND NON-HYPOTHETICALNESS OF MORAL JUDGEMENTS.....	46
1. RAILTON’S ACCOUNT.....	46
(1) <i>Belief</i>	46
(2) <i>Action</i>	49
2. THE PROBLEM OF THE ARGUMENT OF NON-HYPOTHETICALNESS.....	52
CHAPTER 4: OBJECTIONS TO NON-COGNITIVISM.....	55
1. WHAT THE NON-COGNITIVIST HELPS US TO SEE.....	55
2. OBJECTIONS TO THE ATTITUDINAL ACCOUNT OF NORMATIVITY.....	56
3. THE ADVANTAGE OF NATURALISTIC COGNITIVISM.....	57
CHAPTER 5: ETHICAL NATURALISM.....	59
1. REDUCTIONISM.....	59
2. THE PROBLEM OF REDUCTIONISM.....	63
3. NON-REDUCTIONISM.....	67
4. THE PROBLEM OF NON-REDUCTIONISM.....	69
CHAPTER 6: REFLECTIVE EQUILIBRIUM.....	72
1. COHERENTISM.....	72
2. FOUNDATIONALISM.....	74
CONCLUSION.....	76
BIBLIOGRAPHY.....	78

PETER RAILTON'S MORAL REALISM

Introduction

1. Moral Realism

Moral realism is concerned with the nature and status of morality and moral claims. A realistic view about morality asserts the existence of moral facts, true moral propositions and objective foundation of moral principles. Perhaps the first and strongest defense of moral realism appears in the writings of Plato, specifically the first passage from Book 6 of the *Republic*. In this passage, Plato explains how the universe is divided into two realms: the visible realm of material things and the intelligible realm of the forms. The field of mathematics inspired Plato's view of the forms. When we look at numbers and mathematical relations, such as $1+1=2$, they seem to be timeless concepts that never change, and apply everywhere in the universe. Humans do not invent numbers, and humans cannot alter them. Plato explained the eternal character of mathematics by stating that they are abstract entities that exist in a spirit-like realm of the forms.

Although Plato's theory of the forms was endorsed by some medieval philosophers, other notions of moral realism emerged at this time, specifically in the theory of natural law. For Aquinas, morality is grounded in principles that are fixed in nature, particularly in natural purposes, and discernible through reason. All human laws are judged in reference to these. Moral realism was pushed to its limits by Eighteenth-century philosophers, such as Samuel Clarke, who developed a rigorous account of moral

realism known as the eternal fitness theory. On this view, morality is founded on eternally fit principles that belong to a spirit-like world of abstract entities – paralleling Plato's world of the forms. In that spirit world, ethical principles exist eternally alongside mathematical truths. Contemporary discussions of moral realism focus on whether moral facts exist independently of people's beliefs and attitudes, and whether moral judgments can in some substantive way be true or false. It can be characterized as a position that opposes moral skepticism, moral relativism, and non-cognitivism.

2. Moore's Open Question Argument

Moral realism in its modern form is exemplified in John Mill's *Utilitarianism*. He argues for a naturalistic account of morality by defining moral good through empirical facts¹. On this view, good is what people actually desire, therefore is identified with natural facts. In this way, moral judgments become factual judgments and the truth or falsity of them can be detected by empirical investigation.

At the beginning of the twentieth century, G. E. Moore's 'open question argument' in his book *Principia Ethica* gave ethical naturalism a fatal attack. It convinced many philosophers that ethical naturalism had committed what he called a naturalistic fallacy. According to Moore, when we make a normative judgment, such as 'X is good,' we can always sensibly ask 'Is X good?' no matter which natural property X is considered to be. So the question that 'Is X good?' cannot be closed in the sense that 'Is unmarried male a bachelor?' can, because the semantic meaning of 'X', where 'X' stands

¹ This Moorean view about Mill has been challenged by some philosophers who believe Moore misunderstood Mill's project because Mill was not trying to define 'good' at all. See Kai Nielsen's "Mill's

for a natural property, can never be the same as that of 'good.' Therefore, 'good' can only be defined, if it can be defined at all, by the use of other normative notions. If so, the judgment 'X is good' becomes 'Good is good' which is a tautology or 'Good is what is fitting' which is analytic. One philosophical response to the problem posed by open question argument is non-cognitivism or non-descriptionism, such as C. L. Stevenson's emotivism or R. M. Hare's ethical prescriptivism. They believe moral judgments are neither true nor false in any substantive sense, but are imperatives or emotive expressions which aim to express one's own emotion and to evoke the same feelings from the listeners.

In the last twenty or so years a new type of moral realist has emerged. Peter Railton, David O. Brink, Richard Boyd, and Nicholas Sturgeon are members of this group. They reject straight emotivist/prescriptivist analyses of moral language such as those of C. L. Stevenson, R. M. Hare and an error theory such as that of John Mackie. They think these views do not leave room for the concept of a moral truth that they want to defend. They also reject the realistic polar opposite of expressive theories: the view that there are moral properties essentially different from 'natural' properties – for knowledge about which we must rely on self-evidence or intuitions, as was once defended by many philosophers including G. E. Moore and W. D. Ross. The new moral realists' view is an attempt to be somewhere in between. Some members of the new group, such as Sturgeon, think that, if we make proper adjustments in our theories of knowledge and language, we can justify moral conclusions by an argument that is compelling in the same way as are arguments for theories in the empirical sciences.

Proof of Utility" in *New Dimensions in the Humanities and Social Sciences*, edited by Harry R. Garvin (London: Associated University Presses, 1977), 110-123.

The new moral realists aim to construct accounts of non-moral and moral good to which the open question argument does not apply. Railton's approach eschews making analytic claims. Nor does Railton claim that his proposals about the properties moral predicates designate are analytic of the ordinary meaning of these terms, in the simple sense in which 'is an unmarried male' might be thought to be analytic of 'a bachelor' because he believes that "[a]fter all, it seems clear that most judgments of the goodness of particular things are synthetic statements, not tautologies" (Railton 1989, 152). What he wants to show is that there are reasons for understanding moral terms in a certain way – the same sorts of reasons there are for construing 'scientific explanation,' as permitting an understanding of this conception in its actual use in empirical science. Railton's implication seems to be that terms like 'nonmorally good' or 'morally right' can be construed so as to throw light, if we adopt his proposed revisions, on such facts as how evaluations are made or how moral conceptions are adjusted as a result of empirical facts. Thus, he is a reductionist about the analysis of ethical language². As he explains it, accounts of good that are advanced as *a posteriori* identity claims or reforming definitions do not aim at 'conceptual closure' since they do not profess to disclose analytic truths. Evaluative facts can be shown to be *de facto* identical to natural facts, although there are no analytic statements relating them, roughly in the same way in which we know that water is identical to H₂O through general scientific theory. Therefore, the open question argument cannot be used to refute them.

More specifically, Railton thinks that there are two main *desiderata* for selection of these 'revisionary definitions.' One is to choose them so that there is an appropriate connection between evaluative statements so construed and their commending or

² Railton changed into a non-reductionist after around 1995. I will discuss this change in chapter 5.

prescriptive force. Such a choice of definition would have the effect that, on the suggested understanding, these terms would do what the emotive theory claims that ordinary moral language does – express a favoring or prescriptive attitude – thereby having the motivational force of ordinary moral statements. The second *desideratum* is that the properties involved in a naturalistic construction of evaluative concepts play an explanatory role in an account both of the development of values in individual (for nonmoral value statements) and in the development of certain kind of morality in societies (for the case of moral values). So, he proposes an understanding of nonmoral value statements as part of “a good explanatory account of what is going on in evaluative practices involving claims about a person’s own good” (Railton 1989, 168).

Chapter 1: Non-Moral Values

1. Rejection of Fact/Value Distinction

The fact/value distinction is of great importance in twentieth-century moral philosophy as it was earlier for David Hume. According to proponents of the distinction, no states of affairs in the world can be said to be values, and evaluative judgments are best understood not to be pure statements of fact. One version of the fact/value distinction is John Mackie's error theory which holds that there are no values 'in the world,' for such items were too peculiar to be fitted into any decent metaphysics or epistemology. Existentialist ethics by contrast holds that the non-factuality of value leaves us in a position of radical freedom to choose what normative commitments to have. Perhaps the most famous and influential version of fact/value distinction is that of Hume, who claims that 'ought' conclusions do not follow logically from 'is' statements, which is known as the is/ought gap.

The core argument of Railton's reductionist account is one that aims to harden up values, that is to show values can be identified with or reduced to facts. He rejects the distinction between facts and values³ by pointing out that this distinction rests upon insecure foundations and suggests that we should stop viewing these two realms as categorically distinct. The two lines of arguments that have been most important to the philosophical defense of the fact/value distinction are: the argument from rational determinability and the argument from value absolutism. Railton examines them in turn

³ When Railton later adopts a non-reductionist form of ethical naturalism, he also accepts the is/ought gap. See Railton's "Made in the Shade: Moral Compatibilism and the Aims of Moral Theory," in *On the*

and tries to show neither of them is plausible. The sort of value that Railton is concerned with here is a certain sort of non-moral good which is often simply called by him intrinsic value.

(1) Rational Determinability

The fact/value distinction is importantly supported by the rational determinability thesis which is held by many non-cognitivists. On this view, genuine factual disputes can be resolved by appeal to reason and experience. If two individuals differ over a question of fact, yet each is wholly rational in the formation and assessment of his beliefs, then it should always be possible in principle for experience and logical argument to bring them into accord. But when they disagree on values, the normative disagreement need not be resolved even if they go through all the possible empirical examinations and agree on all the factual arguments.

This seems to say that facts are hard and values are soft, because empirical evidence can bring people who disagree on a factual issue to arrive at an agreement while values are in a completely different situation where they cannot be determined by reasons and experience at all. But is this true? Railton points out, "This defense of the fact/value distinction appears to presuppose that there are canons of induction so powerful that experience would produce convergence on matters of fact among all epistemic agents, no matter what their starting points." But, he continues, "the history of efforts to discover a 'logic of induction' has convinced many philosophers that no such canons exist: our

Relevance of Metaethics, edited by Jocelyne Couture and Kai Nielsen (Calgary: University of Calgary Press), 104.

beliefs about the world are underdetermined even by all possible evidence” (Railton 1986a, 6).

Moreover, Railton suggests that deductive logic does not tell us what it is rational to believe either. It only tells us which propositions follow from other propositions. He gives an example to illustrate the point as follows:

- (1) I believe that p.
- (2) P implies q.
- (3) I rationally ought to believe that q.

The conclusion (3) would follow only if we add the premise:

(0) One always rationally ought to believe the logical implications of one’s current beliefs.

Obviously, (0) is not a principle of deductive logic at all. For example, if I believe that all animals are mortal and that my cat is an animal, I rationally ought to believe that my cat is mortal only if I believe that one always rationally ought to believe the logical implications of my current beliefs. However, Railton suggests that even rational agents need not always go where the application of logic would lead them. So, if facts are supposed to be distinguished from values in virtue of what reason requires us to believe, then even the existence of an inductive logic would not secure the distinction. He claims further that the question ‘What ought one rationally to believe?’ will always be a practical question rather than a theoretical one.

(2) Value Absolutism

The second assumption that the fact/value distinction rests upon is value absolutism which is the claim that, in order for something to be intrinsically valuable for

a particular person, it must induce a resonance in any arbitrarily different rational being, that is to say, it must be something that everyone would find to be intrinsically valuable. This is the premise that Railton doubts. He argues that internalism will be problematic when combined with value absolutism while it is not a problem when appealed alone. The central claim of internalism is that, in order for a normative judgment to apply to an individual, he must have some reason – not necessarily overriding – to comply with it. If he is a Humean, he would have some desire to do what the norms require him to for the requirements are necessary means to his ends. The normative judgment is just telling him a way to realize his existing ends. If he is a Kantian, he would believe that meeting the requirements by norms is a part of his being a rational agent. On the internalistic view, a reason will fail to be a reason for an individual if it does not motivate him in any possible way.

If internalism is true, then normative judgements should be able to motivate the person to whom the judgements apply. But if value absolutism is true as well, then the value judgments could not be factual, because according to instrumentalism⁴, there are no substantive ends or activities that all rational beings as such have a reason to pursue regardless of their contingent desires. For example, someone may believe that ‘doing business’ is valuable while I may not find it so. Therefore his value judgment need not apply to me, because his life goal of becoming as wealthy as possible is different from mine, say, that of becoming a good writer. It is obvious that the combination of internalism, instrumentalism and value absolutism implies a distinction between facts and values. Railton points out that, at this point, one might want to protest instrumentalism

⁴ Railton believes in instrumentalism. I will discuss the problem of instrumental rationality in next section of this chapter.

and attempt to argue that some substantive ends or actions are indeed mandated by rationality. But he believes that the problem arises from value absolutism rather than instrumentalism.

He argues that good is a relative notion rather than an absolute one. When we make a claim 'X is good,' we always mean that X is good for a certain subject. There is nothing good for all subjects. For instance, what is good for human beings does not have to be good for all rational beings; because humans are made up of "carbon-based stuff rather than silicon-based stuff, they are the result of a particular evolutionary history, they have various individual developmental histories, and so on" (Railton 1986a, 10). This view is similar to what we think about nutritiveness, for nothing is nutritious for all possible organisms and there is only relational nutritiveness. For example, S is a nutrient for organisms of Type T. "Similarly," Railton argues, "we might say that although there is no such thing as absolute goodness – that which is good in and of itself, irrespective of what or whom it might be good for or the good of – there may be relational goodness. Moreover, among the relational goods for a given being, some may be intrinsic, others, instrumental" (*ibid.*). It is important to see that relationalism of Railton's is distinct from relativism. Although a relational conception of value denied the existence of absolute good, it may yield an objectively determinate two-place predicate 'X is part of Y's good.' Thus we split value absolutism from internalism. It therefore removes an essential premise from the argument favoring for the fact/value distinction.

(3) The Thin/Thick Moral Concepts

Railton rejects the fact/value distinction by arguing against the two important premises that the distinction rests upon. However, his arguments are different from the one that some philosophers (e.g. Philippa Foot, Bernard Williams, Hilary Putnam, etc.) use when they object to the dichotomy of fact and value. This is the 'thin/thick moral concepts argument' that advocates the entanglement of fact and value (Williams 1985, 120-155).

The opponents of the dichotomy argue that it is impossible to separate factual components from normative components for moral concepts, such as 'cruel,' 'insensitive,' 'dishonest,' 'treacherous,' and so on. For instance, we cannot separate the factual content of the moral concept 'courageous' from its normative content. When we make a judgement that someone is courageous, we have to refer to his deeds. So the descriptive element is entangled with prescriptive element in one thick moral concept.

The proponents of the fact/value dichotomy (i.e. non-cognitivists) may respond that the word 'courageous' may have some factual content, but this can always be distinguished, at least conceptually, from the evaluative content. The former contains his factual deeds and the latter contains our pro-attitudes. Suppose, for instance, Tom was a soldier in a war. He saw his friend Richard trapped in a burning tank. There was no way for Richard to get out of the tank if no one came to help. In a few minutes, the tank would explode. The situation was so dangerous that nobody would dare to help Richard but Tom. He climbed onto the top of the tank and dragged Richard out.

Non-cognitivists claim that if you approve Tom's behavior, you will make a judgment that he was courageous. But if you don't, you will make a judgment that he was reckless. Cognitivists would respond that the courageousness or the recklessness of Tom's action has nothing to do with commentators' attitudes, rather it has everything to do with the facts of the situation at the moment. If there were no chance at all to save Richard, Tom should be considered as reckless and unreasonable. But if there were a reasonable chance, even a small reasonable chance, Tom was courageous. Non-cognitivists still can ask 'How much chance is counted as a reasonable chance? Where should we draw the line? Isn't it defined by people's different attitudes?' Cognitivists may reply that it is not a matter of attitude but a matter of fact. For instance, according to one society's convention, social members agree on the fact that if there is a certain range of probability of success, then Tom's action was worthwhile. Non-cognitivists could still ask, 'Is social convention only a matter of fact? Isn't it a combination of empirical facts and people's emotions and attitudes?'

Some non-cognitivists, such as Allan Gibbard, do not reject the thin/thick moral concepts argument. Gibbard will readily propose a descriptive account of what moral thick concepts distinctively come to and will not deny that statements containing them are true or false. He will also allow that these thick concepts can enter into explanations of facts like a person stealing – explained by the fact that he is dishonest. However, it is the 'thin' concepts, such as 'moral obligation,' 'duty,' 'morally wrong,' and 'intrinsically good' that are in dispute. It is not clear that they are descriptive. Here, he thinks, an expressivist analysis is more adequate.

2. The Reduction from Value To Fact

(1) Railton's Account

i. A Person's Good

On Railton's account, the notion of good is relational rather than absolute in the sense that it must be related to a certain subject. This thesis was originally proposed by Thomas Hobbes who discussed relational good in *Leviathan* (Hobbes 1651, 120). He held that to call something good is always to speak of someone's good, and that the only sense in which something can be good for someone is that he desires it. In other words, people's preferences are indicative of what is desirable. It is this further claim that Railton believes is mistaken.

Indeed, we do not always most want what is good for us. Hobbes confuses what is desired and what is desirable. A person's good is desirable, but it may not be actually desired by that person.

Now let us discuss two examples provided by Railton. The first one is Sheila's job changing story. She was a successful journalist working for a southern American newspaper. What she always wanted was an even more successful career. One day, she received a job offer from a big newspaper *The Daily Planet* in New York. Her first impulse was to accept it. But she started asking herself if it was really good for her. If she moved to New York and worked for a much more demanding position, she would have very little spare time to spend with her family and would lose the easy access to nature, the familiar environment and many good friends whom she knew for a long time. So, for Sheila, is to do what she had always wanted really good for her? (*ibid.*, 12)

Another example is about Beth's choices. Beth was a happy accountant who always wanted to become a novel writer. Though she had some successful experience of writing at school, her novels had never been accepted by any publisher. She believed it was because she did not have enough time to devote to writing. One day she quit her job and started writing at home very seriously. To her surprise, she found it so hard to concentrate on work when she finally got enough time and she also got a different impression for writing from before. Writing novels became so boring and tiring that she would not want to continue. The actual experience of being a full-time writer helped Beth to realize that it was not a good career for her at all. (*ibid.*)

In these two cases we question whether our good coincides with what we most desire. Why was Sheila anxious about her choice? What did she need to know when making important decisions? What information did Beth not know when she just quit her accountant job to start a writing career? All the worries that Sheila had and the mistakes that Beth made indicate that what is desired does not have to be what is desirable in the sense that what we most desire does not have to coincide with what is good⁵ for us. Moreover, what is good for us is psychologically very much related to the fact whether our desires are satisfied. To have a desire is, among other things, to care whether or not it is satisfied. For an individual to deem something a goal or a value of his own involves the idea on his part that it is an appropriate object of desire or pursuit⁶. The notion of appropriateness at work is internal, but may concern the desires he would want to be

⁵ 'Good' refers to 'objective good' or 'objective interest' that Railton has in mind.

⁶ Railton believes that an object will be an 'appropriate' object of one's desire or pursuit if it is not impossible to be achieved in real life. Appropriateness is a naturalistic notion, which consists in a certain match between an agent's motivational system and his capacities and circumstances.

effective in his actual life were he to contemplate that life with full awareness of the facts and full rationality in deliberation.

ii. A Person's Intrinsic Good

From Beth's case we can see that what she actually desires is not what makes for her good. Her good is better identified with her higher-order desire⁷ which is the desire that she has when she has full information about what it is like to be a writer. The higher-order desires of the sort that are involved in embracing a desire are more responsive to changes in belief, and so not only do they become more closely tied to our identity, but also they become the basis of the idea of value. Railton uses this higher-order desire to define a person's good by claiming that "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" (*ibid.*, 16). The higher-order desire is in this sense an objectified desire. The fact that Beth was at that moment so constituted that she desired something which, had she better knowledge of it, she would wish she had never sought, implies that the desire she originally embraced was defective. The objectified desire remedies this kind of defect and better matches the agent's interest. Railton calls the interests reflected by a person's objectified desire 'objectified subjective interest' (Railton 1986b, 142) or simply 'objective interest' (*ibid.*, 143).

This kind of program leads Railton to suggest an analysis of one's non-moral intrinsic value in terms of its being the object of a person's 'informed desires' or of being

what is in one's 'objective interest.' The identification of non-moral values with natural properties can be seen as of the same type as an ordinary identification of water with H₂O, both of which could be modified on the basis of empirical discoveries.

The higher-order a desire is, the closer it is to an individual's intrinsic good or objective interest. So "an individual's intrinsic good," according to Railton, "consists in attainment of what he would in idealized circumstances want to want for its own sake – or, more accurately, to pursue for its own sake (for wanting is only one way of pursuing) – were he to assume the place of his actual self" (Railton 1986a, 17).

Sheila's intrinsic good is what she would want were she to be in an ideal situation where she had full and vivid information about her life in the new working position. Beth would know what is good for her were she to have actual experience of being a full-time writer. Generally speaking, for a certain agent S, the full factual and nomological information about her physical and psychological constitution, capacities, circumstances, history, and so on will make S become S+, who has complete and vivid knowledge of herself and her environment and whose instrumental rationality is in no way defective. Then what S+ would want her non-idealized self S to want were she to find herself in the actual condition and circumstances of S is the intrinsic good for S in her place. Railton's account of an individual's intrinsic good leaves open the possibility that the intrinsic good of different persons may differ, therefore it is not compatible with value absolutism.

⁷ The more information one has concerning the desired object, the higher-order one's desire is.

iii. Reduction Basis

Besides a person's good and a person's intrinsic good, Railton introduces another important conception: reduction basis. On his view, what makes a certain end or activity be part of an individual's good is not that his idealized self would, as a matter of fact, want his actual self to pursue it, but the existence of the reduction basis of the person and his environment. In other words, the existence of the reduction basis offers a reason to explain why an individual has a certain desire for its own sake when he is in idealized circumstances. For example, what makes Beth want to quit writing is the feature of her reduction basis that she lacks the creativity, imagination, and certain personal traits that a successful writing career requires. Just as secondary qualities have their reduction bases: the perceiver is so constituted that this sort of object in this sort of context will excite that sort of sensation, Railton believes an individual's intrinsic good has its reduction basis, too. It is a complex set of facts including an agent's capacities and circumstances that constitutes the agents' intrinsic good. The reduction basis gives expression to an idea of appropriateness or fitness of an end for an agent. Fitness consists in a certain match between an agent's motivational system, on the one hand, and his capacities and circumstances, on the other, when all are accurately represented and adequately appreciated. The reduction basis for Beth's desire – which would include her ability, personality, and circumstances – constitutes the fact that a life as a writer would not be a good life for her. The apparent advantage of this strategy is that truth-values of facts do not fluctuate with either our particular desires or decisions or that of others, so that this toehold can support us even when we find ourselves in a world with people whose beliefs or ends differ from our own.

The following examples can illustrate the same point. When we build a roof, the size of rafters is determined by the reduction basis which are constituted by a set of facts, such as, the size and material of the roof, the style and the budget of the house, etc. If it ought to be 2×8's , then it will collapse when we build it as 2×6's. Thus the facts about the house and other conditions give rise to the criteria of the intrinsic good for the house, which the builders must try to approximate. Similarly, when we try to make a ship, the reduction basis of the ship consists in a set of mechanical facts that will affect the ship's performance in the sea known as a ship's sea-worthiness. A sea-worthy ship must be built in accord with physical and mechanical principles but does not have to be properly blessed or beautifully painted. These examples have further shown the connection between the non-moral values and their reduction bases.

iv. The Normativity of Goodness (Wants/Interests Mechanism)

Can intrinsic goods determined by reduction bases guide our actions? Railton's answer is yes. He believes one's objectified subjective interests can also play an explanatory role in the evolution of one's desires. They guide us by the positive or negative reinforcement of our previous desires. According to an influential psychological theory, Railton remarks, pleasure is the primary reinforcer of desires. What we now want is fixed by our experiences with the now desired sort of thing as having been pleasant in the past. Railton concedes, of course, that it is obvious that people want other things besides pleasure or happiness. But in various conditions, experiences of happiness

reinforce the desires that tend, in those conditions, to produce happiness. Thus, happiness can enter into an explanatory theory of the individual's desires.

Lonnie's story illustrates the point. Lonnie was traveling in a foreign country and feeling miserable. He was homesick and had a problem with his stomach. He didn't know what his malaise was but found himself having a craving for the familiar: a glass of milk. So he drank some milk. Soon afterwards, he felt much worse. Then he bought a bottle of '7-Up' and drank it too. This time, he felt much better almost at once although he was still unable to identify the source of his malaise. As a matter of fact, Lonnie's malaise was dehydration which can be worsened by hard to digest milk and cured by clear fluids or soda drink. Railton remarks, "The reduction basis of his objectified interest includes facts about Lonnie's circumstances and constitution, which determine, among other things, his existing tastes and his ability to acquire certain new tastes, the consequences of continued dehydration, the effects and availability of various sorts of liquids, and so on. This reduction basis is the constellation of primary qualities that make it be the case that the Lonnie has a certain objective interest" (Railton 1986b, 143). When next time Lonnie traveled, he would have some conscious or unconscious tendency to seek out 7-Up and avoid milk.

This reinforcing process is what Railton calls the wants/interests mechanism, which permits individuals to achieve self-conscious and unselfconscious learning about their interests through experience. This process is possible because it is, as a matter of fact, impossible for a person to have the experience of happiness and not be drawn to the prospect of it. So, 'pleasant' must have normative force, just as do 'good' and 'valuable.' It is a psychological truth that humans are constantly pursuing their physical and

psychological well-being and avoiding physical and psychological ill-being. Although they do not have a reliable way to know their reduction bases directly, they can nonetheless get to know them gradually through constant trials and experiments. The feedback of the experiments often motivates them to keep going in the direction that positive reinforcements construe. Railton concedes, "There is no guarantee that the desires 'learned' through such feedback will accurately or completely reflect an individual's good. It is entirely possible, and hardly infrequent, that an individual lives out the course of a normal life without ever recognizing or adjusting to some of his most fundamental interests" (Railton 1986b, 145). Thus the truth-condition of the claim that such-and-such is good for a given individual is directly given by the existence of this constellation of features, without detour through idealized desires.

A certain type of wants/interests mechanism, as Railton suggests, would cultivate a corresponding type of individual rationality. Our tendency through experience to develop a certain type of habit and strategy may co-operate with the wants/interests mechanism to provide the basis for an extended form of criterial explanation⁸, in which an individual's rationality is assessed not relative to his occurrent beliefs and desires, but relative to his objective interests. Therefore, if an agent acts according to her individual rationality, she is likely to obtain her objective interests and intrinsic goods. In this way, an individual's objective good is normative for her in the sense that it guides her, if she is reasonable, to pursue certain things and to eschew certain other things. The wants/interests can operate with not only the objective interests classified as needs, but

⁸ According to Railton, 'criterial explanation' is the criterion that conveys explanatory information. "[W]e explain why something happened by reference to a relevant criterion, given the existence of a process that in effect selects for (or against) phenomena that more (or less) closely approximate this criterion." See Railton 1986b, 148.

with respect to any interest – even interests related to an individuals’ particular aptitudes or social role – whose frustration is attended even indirectly by consciously or unconsciously unsatisfactory results for him.

The normativity of the goodness also presents itself in the recommending or normative force of some factual claims. Discourse about a person’s good also has a prescriptive side, for to make or accept a judgment that something is good for someone is in some sense to recommend that thing to him. In the case of building a roof, when we say the size of one rafter ought to be 2×8’s, we say something prescriptive as well as descriptive. “One can,” according to Railton, “use a description to recommend, or use a prescription to convey information” (Railton 1989, 152). Discourse about a person’s good presupposes an inherent connection between the content of the description and the force of the prescription. The proponents of non-cognitivism have sought to deny that discourse about goodness is genuinely referential, despite the descriptive character of its surface grammar. They seek to remove any mystery that there might be about the ‘two-sidedness’ of discourse about goodness by developing an interpretation of it as essentially prescriptive, but not fundamentally descriptive. But no one has as yet offered an expressivist account that satisfactorily preserves all the cognitive features of discourse about value. Railton’s cognitivist approach moves in the direction that is the opposite of expressivism. It treats the cognitive character of value discourse – its descriptive side – as essential to it, and then seeks to account for the prescriptive force of value judgments as arising from the substantive content of such judgments. It attempts to locate value properties among features of the world that are accessible to us through ordinary

experience and that play a role in empirical explanations. It therefore treats value properties as natural properties, and so is a form of naturalism about value.

(2) Critiques of Railton's Account

i. The Problem of Instrumental Rationality

Since Railton attempts an empiricist approach to ethical naturalism, 'reason' in his account is 'instrumental reason' in roughly Hume's sense rather than 'practical reason' in Kant's sense. Perhaps Railton's Humean view is like this: instrumental rationality is unproblematic. It is rationality about what means will reach what ends when we quarrel about both facts and ultimate goals. Christine Korsgaard criticizes Humean instrumental reason as a reason that needs a complement, because "the instrumental principle cannot stand alone. Unless something attaches normativity to our ends, there can be no requirement to take the means to them" (Korsgaard 1997, 251). On her view, the problem of the empiricist account can be summarized as follows: if instrumental reason is the only reason to be used, it cannot be normative; but if some kind of normativity is attached to instrumental reason, the empiricist account will transform itself into its opponents' account. I find Korsgaard's criticism of empiricist moral theories very plausible. If we apply it to Railton's account, we will see that he has smuggled 'prudential reason' into his theory when he claims that instrumental reason is the only reason that is needed.

According to Korsgaard, there are three kinds of principles proposed as requirements of practical reason. They are the instrumental principle, the principle of

prudence and moral principle. The instrumental principle requires one to take means to her existing ends; the principle of prudence requires one to pursue her overall good which is sometimes identified with her self-interest; moral principle requires one to promote others' interests. She remarks, "empiricist philosophers and their social scientific followers have obscured the difference between the instrumental principle and the principle of prudence by making the handy but unwarranted assumption that a person's overall good is what he 'really' wants. Prudent action is then just a matter of taking the means to your true end; and the instrumental principle is the only non-moral imperative we need" (Korsgaard, 218). It seems to me that the confusion that Korsgaard mentions also appears in Railton's moral realist account. Though he does not identify a person's overall good as his immediate or local desires, he does believe what a person really wants under idealized circumstances is his overall objective non-moral good. However, Korsgaard believes that one's prudential good cannot be recognized by appeal to the instrumental principle only, because instrumental reason cannot tell us which end to choose, but only tells us the fact that a person is caused to act by the recognition that an action will promote her existing end. It does not tell us, as the principle of prudence does, to pursue our own overall good. The person himself, the one whose behavior is in this way predicted, is not guided by any dictate of reason as far as instrumental reason is concerned. Let us examine what kind of reason Railton appeals to when he develops a conception of one's objective good from what one wants under idealized circumstances.

As I quoted before, he says that "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely

free of cognitive error or lapses of instrumental rationality” (Railton 1986a, 16). For instance, Beth realized that full-time writing was not a good life for her and Lonnie formed a habit of drinking 7-Up and avoiding milk when travelling. Both cases are designed to show that, when people have obtained better knowledge about themselves and their circumstances, their desires are very likely to be coincident with their objective interests. But to arrive at this new level of knowing what is good for them, do Beth and Lonnie *only* use instrumental reason? As far as I can see, it’s not the case. In Railton’s stories, Beth came to conclude that constant frustration in her writing career began to threaten her psychological health and Lonnie came to believe that drinking milk will worsen his stomach malaise. Apparently, their concerns about their psychological or physical well-being play an important role in the process of their choosing or abandoning goals. In other words, the premises of the stories are: Beth has already chosen psychological well-being as an ultimate end and Lonnie has already chosen physical well-being as an ultimate end. Obviously, they cannot choose these ends as ultimate and overriding by appeal to instrumental reason, because instrumental reason can only tell Beth and Lonnie to choose a better means to realize their already existing ends. However, Railton does not explain how and why they choose well-being as ultimate ends. But his permission for his protagonists to adopt certain values before they use instrumental reason seems to suggest that Korsgaard’s criticism is right on the mark. I think Korsgaard is right in believing that instrumental reason cannot by itself give us a reason to do anything because it tells us only to take the means to our ends and it can operate only in conjunction with some view about how our ends are determined, about what they are.

Railton could respond to Korsgaard's charge by pointing out that desiring one's overall well-being is no more a choice of reason than an impulse of psychological inclination. Anyone who has ever experienced pleasure and pain in the past will certainly want to duplicate the pleasant experience and to avoid painful experience in the future. When an individual is provided with full and vivid information, he would calculate about how much pleasure and pain he is going to experience in each possible way of living his life. Then, his psychological inclinations will direct him to the life in which he can experience the most pleasure and the least pain. Railton's strategy is to replace prudential reason with hedonist psychology. However, first of all, psychological hedonism is a controversial thesis that appears to ignore various complications as well as the influence on values of parental teaching and example. If we explain 'non-moral good' as being just 'pleasant,' it would seem that pleasure must be the only thing that is intrinsically good, which, as the long history of debates about hedonism manifests, is doubtful or at least thoroughly problematic. Moreover, we may doubt Railton's hedonist argument by considering why reinforcement by pleasure might not produce a strong desire for adventure in a person, a desire so strong that she regularly sets aside some pleasures in order to gain more opportunity for adventure. The hedonist may want to show that, for this person, adventure can produce much more pleasure than her lying on the beach can. But the attempt of ranking all human activities according to their ability to produce pleasure brings enormous difficulties. Finally, it is possible for a person not to care about her overall happiness or well-being at all. For her, the full presentation of the information can hardly induce a certain desire in her aiming to promote her long-term physical and psychological well-being. She is so careless about her life that she only tries to satisfy her

immediate and local desires whatever they are. Therefore, to maximize pleasure is not always the goal with superior normative force. Thus, it is not clear, from pleasure having primacy in a psychological explanation, that it always has superior normative force. The argument to the hedonist's analysis of the non-moral good, then, basically only shows that the connection of happiness with motivation is very deeply rooted. Without an ultimate goal of wanting one's overall good, an individual can desire something that is quite different from her objective interest even under idealized circumstances. The instrumental principle says nothing about our ends, so it is completely unequipped to say either that we ought to desire our overall good or that we ought to prefer it to more immediate or local satisfactions.

Railton can still reply that, following Hume, prudence is not a rational requirement but a virtue. Hume believes that our tendency to act prudently is not the result of our rational nature but rather of the original instincts which nature has implanted in us. It is not a requirement of reason that we should have concern for our future, but that it is, given the beings that we are, natural to have such a concern. The criteria against which we pick our ends is not a principle of reason, but can be virtue or some other normative principles. Korsgaard reveals the nature of this sort of option by pointing out that "although Hume denies that prudence is a rational requirement, he certainly does think it is a virtue" (Korsgaard, 232). But in admitting it as a virtue, "[w]e will have rescued the instrumental requirement for Hume, but only at the cost of showing that the word 'virtue' simply does the work in his account of action that the word 'reason' does in his supposed opponent's account" (*ibid.*, 233). That is to say, Railton will contradict

himself and advocate his opponent's account if he accepts desiring one's overall good as a virtue of an agent.

Moreover, instrumental reason is more than having rational means-end beliefs. Any account of instrumental rationality must cope with interactions of reasoning agents one with another. It must cope with uncertainty. We need something like Frank Ramsey's theory to capture these complexities (Ramsey 1926). He treats rationality as a kind of formal coherence. His axioms require more than sheer logical consistency in belief. They require that one's preferences form an ordering. Instrumental reason seems just a matter of beliefs about consequences, and of doing the thing whose consequences one most wants. The two examples – prisoner's dilemma and Russian roulette – have shown that things are not so straightforward. In both cases the facts are clear enough, which are given by stipulation and known to the agents. But the puzzles of instrumental rationality lie not just in the facts. Agents seek their goals in interaction with other agents who have different goals. Agents pursue their goals knowing they are in certain respects ignorant. Instrumental rationality includes coping with both these features of life, and there are disputes about how it makes sense to do so. In a broad sense, Ramsey's decision theory treats rationality as a kind of coherence, but this coherence is neither just logic narrowly construed nor just straightforwardly choosing what one most wants.

ii. The Problem of Ideal Observer Theory

The structure of Railton's moral realism is similar to that of Ideal Observer theories canonically proposed by Roderick Firth and revised by Richard Brandt. Railton

has simply restricted the scope of the Ideal Observer's evaluative authority and made each of us his own Ideal Observer. I believe his theory shares the same difficulty that Firth's and Brandt's theories have confronted.

The proponents of the Ideal Observer Theory hold that a morally right action is what can be approved by an ideal observer, who is omniscient with respect to non-ethical facts, omnipercipient, disinterested, dispassionate, consistent, and is normal in other respects. Although Railton does not require his agent to be disinterested or dispassionate, he does think an agent needs full information and imagination in order to desire what is objectively good for her. So Railton's account is more like that of Richard Brandt's who revised Firth's theory by replacing 'ideal observer' with 'fully rational person.' According to Brandt, the essential differences between the conception of a 'fully rational person' and an 'omniscient, omnipercipient, disinterested, dispassionate, but otherwise normal person' are as follows: (1) A 'fully rational person' is defined as one who makes full use of all available information, but not as a person who is omniscient or omnipercipient. (2) A fully rational person is one whose desires have been fumigated (or enlarged) by cognitive psychotherapy; the alternative conception lays down no such requirement. (3) It is not a conceptual necessity that a fully rational person be either disinterested or benevolent; it is an empirical question whether or to what extent he is. (4) One might question how nearly 'normal' a person who is omniscient and disinterested and dispassionate might be in other respects. (Brandt 1979, 225)

Brandt's account could be put as follows. First, we characterize the intrinsic desires that it would be rational for a person to have in the sense that they purport to achieve the intrinsic good for the person. They are the ones that the person would have

after repeated representation of all relevant, scientifically available information, in an ideally vivid way, at appropriate times. Brandt calls this vivid, repeated representation of facts *cognitive psychotherapy*. Second, we define the rationality of an act as a means, or its 'instrumental rationality.' Take some system of intrinsic desires as given, an act is instrumentally rational as a means to the fulfillment of those desires if and only if the following holds: if a person had those intrinsic desires, and he had all relevant scientifically available information "present to awareness, vividly, at the focus of attention, or with an equal share of attention" (*ibid.*, 11), then he would be willing to perform that act. An act, then, is fully rational if and only if it is instrumentally rational as a means to the fulfillment of the intrinsic desires it would be rational for the agent to have.

First, we may ask if it is true that any desire that survives cognitive psychotherapy is the one that is rational for the agent to have. Some desire that is so deeply ingrained in early childhood as to be inextinguishable by cognitive psychotherapy will count as rational on his definition but still it might be irrational. Brandt calls this kind of desire a "mistake" which is related to the "exaggerated valances produced by early deprivation" (*ibid.*, 122). The example Brandt provides shows that if a person experiences a material shortage in his early childhood, he would want to accumulate economical means as much as possible in his adult life.

Dear Abby: My husband grew up fatherless during the depression. Now, at age 50, his net worth is around the half-million dollar mark. He is a professor with tenure, and has an excellent retirement and insurance program. Yet he buys second-hand clothes, day-old bread, and refuses to spend any money on a decent car, vacation or travel.

The reason? He wants to be sure he has enough money for his old age. What could be the matter with him?

HIS WIFE. (*ibid.*, 123)

To be sure, this man's desire is irrational. It is expected to be extinguished with vivid and repeated awareness of certain facts, such as saving money as much as possible will not promote his well-being and living in this way actually reduces the quality of his life. Suppose, though, he has confronted all these facts, repeatedly and vividly, he nevertheless finds his desire still in him and says that 'I realize all that. But I just don't feel that I have enough money for my old age.' If so, this person would be counted as rational in his preferences on Brandt's account. But, as Brandt himself has realized, it is a mistake. Therefore, it is possible that an irrational desire could survive cognitive psychotherapy.

Secondly, one problem for any 'full awareness' account such as Brandt's is that rationality, in the ordinary sense, often consists not of using full information, but of making the best use of limited information. Allan Gibbard has pointed out that "[a]cting in full awareness of all relevant facts suggests not rationality, but something more like 'advisability.' Whereas rationality is a matter of making use of the information one has, advice can draw on information the advisee lacks" (Gibbard 1990, 18). For example, a detective investigating a murder case: what she could get from the murder scene was far from full information, but just a truck's tire mark for about half a meter long in the mud near the dead body. What she could rationally do is to look for the truck that left the tire mark in the scene and hope to get some information from the truck driver. But, when the truck and the driver were both found, the detective was convinced that they had nothing to do with the murder and that looking for them was actually of no use for the investigation. She knew that what she could rationally do at the very beginning could be

very different from what she could rationally do if she had more, let alone full, information concerning the truck and the murder. Suppose that now a person came to tell her what he saw through the window when the murder happened, what she could rationally do now would be what the new information suggests her do. Rationality, we may conclude, is related to advisability, but the connection is this: in the special case in which an individual knows all that bears on her choice, what it is rational for her to do is what it is advisable for her to do. Otherwise the two may differ, and full information pertains to advisability.

Finally, we may doubt the recommending force possessed by the full information. The word 'rational,' in the sense we are after, has an automatically recommending force. But should it possess such a force? For a traumatized veteran, the full vivid information about war will only cause him horror and anxiety. For a mother who loses her son in a brutal battle, the detailed information concerning how her son was suffering to death will make the sad news more cruel and unacceptable to her. A similar story might be told for an egoist. Suppose an egoist achieves happiness by keeping his mind off the joys and sorrows of others, but he thinks that if he fully realized what suffering he could alleviate and what joy he could spread by a life of self-sacrifice in the service of others, he would sadly forsake his life of self-centered enjoyment, and take on an irksome burden of service to humanity. Must he find that a recommendation for self-sacrifice? Why is it not instead a reason to shield himself from the facts? He is convinced that the life he would want to lead, if he fully realized what was at stake for others, is painful in comparison to the life he will lead with his egoism fostered by ignorance. If he is also convinced that the altruistic life is the one he would choose with full knowledge, why should that commend

the altruistic life to him? Further: Would a given moral code be justified to an agent if she knew that she would support it if she were fully informed and either completely disinterested or benevolent or both? Would that fact necessarily work as a recommendation to her to support such a moral system of that type? Not necessarily.

The three points above show that we may not be reliable formers of rational desires even if we get the facts straight and have them clearly and vividly in view. The part of Railton's theory based upon Ideal Observer Theory is therefore problematic.

iii. The Problem of Internalism

One difficulty encountered by Railton's account stems from its commitment to internalist moral psychology. According to internalism, in order for a normative judgment to apply to an individual, she must have some motivation (not necessarily overriding) to comply with it. A person's objective intrinsic good is what an individual would desire for its own sake under idealized circumstances, which is closely connected to her reduction basis. A person's reduction basis is constituted not only by her capability and environment, but also by her motivational system including her personal traits, such as being careless, sensitive, courageous, optimistic, and so on, which help to characterize who she is. In this section, I will try to argue that the inclusion of a motivational system into reduction basis can cause a serious problem for Railton's account, which is rooted in the internalist moral psychology that he appeals to. Let us first look at an example.

Bob is a very sensitive and thoughtful person and his wife Cindy is just the opposite. Bob has been thinking if it is a good idea to become more like her. But he is not

so sure that becoming an insensitive and unthoughtful person is good for him. He has also convinced himself that if he really becomes a person like Cindy, he would not worry about the same question. It is not because there is no downside in being Cindy's kind of person, but rather, a person like Cindy would not have that question before her mind. Suppose that he can swallow a pill to change his personality and that he wants himself to change, Bob can still meaningfully question the benefit of the change because he is wondering about being what kind of person is better for him. Bob reasonably decides that, because of his traits, he is not a proper judge of his own good. He may have the ability to recognize that his problematic traits impair his judgment, while lacking the ability to prevent those traits from influencing his assessments of what to desire. It seems to me that he is asking a Moorean question that 'X is something that I would want myself to want, were I fully informed and rational, but is it good for me as something I ought to have?' Given his concern about the sort of person he is, he may reasonable wonder whether what someone like himself would want for himself if fully informed and rational is indeed good for him. Whether the kind of person we are now influences our decision making in a way that begs the question that most deeply lies behind our queries about what to desire – 'what sort of person shall I try to become?'

It is obvious if we make a person's own motivational system a part of the very criterion that determines her good, her personality traits will affect the character and the outcome of her deliberation. Suppose that Hitler would want what he wants even as his fully informed and rational self, we cannot therefore conclude that what he wants is good for him. Similarly, a desire that a cruel person's reduction basis gives rise to is unlikely to be the same thing that a loving person's reduction basis does. Railton's account does not

have the resource to tell which desire under idealized circumstances is good for the person himself and which desire is not. It endorses any desire that is given rise to by any reduction basis as a desire that purports to achieve the agent's objective interest.

Railton may object to this line of criticism by suggesting that all the objectionable traits of a person will not work during the process of decision making under idealized circumstances. First, one might believe that the effects of different traits tend to wash out under conditions of full information and rationality. But I doubt that as a matter of fact all fully informed persons would have the same reactions. Second, the process of becoming fully informed and rational will require certain changes in a person's motivational system, which are the changes that will eliminate what we intuitively regard as objectionable traits in favor of ones we all would be inclined to regard as desirable in an ideal person. But different persons react differently to increased information and the experiences. Some people become more open, concerned, and empathic while others become increasingly distrustful, cold, and indifferent. So, I believe that, in order to get rid of the influence of the objectionable traits in one's reduction basis, Railton must either give up his commitment to internalism or to specify which personal traits are allowed in one's motivational system and can be further regarded as a constituent of one's reduction basis. However, it seems to me that the real option for him is to do the former, because, by doing the latter, he will return to the original Ideal Observer Theory by giving up the possibility that everybody can be his own ideal observer under idealized circumstances regardless of his personal traits.

Chapter 2: Moral Values

1. Railton's Account

(1) Moral Facts

Railton claims that there are moral facts and properties to which we should be able to gain epistemic and semantic access. There is no special mystery about what sort of things they are or how we come to have knowledge of them, refer to them, and so on, because they are natural properties rather than *sui generis* moral properties. Such a naturalistic account could have the advantage not only of permitting an explanation of epistemic and semantic access to moral facts, but also of affording a straightforward explanation of the cognitivism of moral discourse.

(2) Social Rationality

The strategy Railton uses to identify a society's moral good is much the same as the one he uses to identify a person's intrinsic non-moral good. The analogy is between an individual's reduction basis and a society's reduction basis. The latter consists of the actual realization of the interests of each social member. The notion of social rationality is introduced "by considering what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information" (Railton 1986b, 150). The idealization of social rationality is equivalent to what is considered as rational from a social point of view with regard to the realization of

intrinsic non-moral good of each social member. We can realize if a society's social rationality is obtained by looking at its reduction basis. "Just as an individual who significantly discounts some of his interests will be liable to certain sorts of dissatisfaction," Railton remarks, "so will a social arrangement – for example, a form of production, a social or political hierarchy – that departs from social rationality by significantly discounting the interests of a particular group have a potential for dissatisfaction and unrest" (*ibid.*). An individual whose wants do not reflect his interests or who fails to be instrumentally rational may experience feedback of a kind that promotes learning about his good and development of more rational strategies. Similarly, the discontent produced by departures from social rationality may produce feedback that, at a social level, promotes the development of norms that better approximate social rationality.

On Railton's account, the conception of social rationality permits the notion of moral good to be explanatory, to play an important role in the understanding of social history. For societies in which the interests of some are favored, in contrast to institutions and policies that are rational from the social point of view, have a potential for dissatisfaction and unrest, alienation, loss of morale, and decline in the effectiveness of authority. Such a potential produces feedback that "promotes the development of norms that better approximate social rationality" (*ibid.*, 151). The same pressures have led to the broadening of the concept of 'society' to include all other peoples including very different people in different circumstances. The same applies to movement in the direction of attaching important moral requirements to the conditions for well-being, for the satisfaction of human interests. These pressures occur because persons whose

interests are denied are led to form common values and make common cause along lines of shared interests, thereby placing pressure on social practices to approximate more closely to social rationality. Social history can give us information about which social structures best satisfy this ideal of social rationality.

(3) The Inescapability of Morality

Railton claims that all the three features of morality – its scope, well-foundedness, and impact – have shown that moral evaluations are inescapable. He says, “Moral evaluation would be formally escapable were its scope hypothetical; it would be substantively escapable were its standards ill-founded (criticism based upon a mistake can be avoided by pointing out the mistake) and it would be social-psychologically escapable were it without widespread evaluative impact” (Railton 1992, 47).

First, our notion of moral evaluation has a non-hypothetical scope. To show its non-hypotheticalness we do not have to appeal to Kant’s categorical imperative, because there are various other classes of critical standards with non-hypothetical scope in the relevant sense, such as esthetics, Logic, and the theory of instrumental rationality. For example, we believe that we can use logical standards to criticize one’s public speech. No matter whether the speaker cares about the logic of his speech, we can easily dismiss his arguments by pointing out that what he says is not consistent. Similarly, we use aesthetic criteria to evaluate a painting no matter what the painter himself thinks about it. It is in the same way that we deem moral evaluations to be properly applied even to agents who lack any significant intrinsic desire or extrinsic motive to give these evaluations weight in their deliberations. Moral evaluations are non-hypothetical because the contribution of

one's action to the well- or ill-being of others is independent of one's interest in it. Railton remarks, "After all, the basis of logical evaluation is not to be found in reasons but in reasoning, and bad reasoning is not made good by having compelling reasons for engaging in it. Similarly, the basis of morality is not to be found in the agent's reason's alone, but in the reasons or well-being of others he might affect, and bad actions are not made good when the agent has compelling reasons of his own for engaging in them" (*ibid.*, 49). In other words, the right application of moral standards to a certain action is the same as that of logical standards to a certain speech, which is independent of the particular ends and desires that anyone happens to have, but rests upon the correctness of the reasoning itself or the well-being of other people it will affect.

Secondly, some, or perhaps even most, of our moral evaluations are well-founded. Morality does not rest on error as the gods of the past did, because pains and pleasures matter to us. The facts of pain and pleasure may constitute an important part of a theory of well-being which is the core of the justificatory basis of morality. Railton argues, "If pains and pleasures are genuine states about which we can have evidence, as it seems to me they are, then to that extent we have a notion of well-being that is epistemically respectable as well" (*ibid.*, 44). For him, morality can survive ideology and aim to promote the well-being of all human beings.

Finally, Railton explains the evaluative impact of moral judgments, which is not categorical, because, on his view, the person who assigns moral standards no weight in deliberation need not be *irrational* at all. The penalty of this kind of person is not the threat of being irrational or unhappy, rather, "it is the penalty of being unable without deception to see oneself and one's action in a certain way, of being unable without

deception to defend oneself and one's action in light of a certain standard, or to claim a kind of objectivity or autonomy in one's deliberative practice or agency" (*ibid.*, 43). But it is, as Railton has realized, that "[t]his is a penalty only to the extent that this way of seeing and justifying ourselves has evaluative impact for us" (*ibid.*).

2. Critiques of Railton's Account

(1) The Problem of Railton's Justification of Morality

Railton tries to use the same strategy to justify both an individual's objective non-moral good and a society's moral good. For the former, a person's objective good is connected to her reduction basis which includes her capability, emotional system, environment and so on. For the latter, a society's moral good is related to a society's social rationality which is concerned with the satisfaction of each social member's non-moral interest. But the similarity between the two arguments seems to me artificial, because the moral psychology that the two approaches appeal to conflict with each other in the sense that one is internalist and the other is externalist. For example, the reason for an individual to pursue her self-interest is an internal reason for humans are motivated to pursue their own goods while the reason for an individual to act morally is an external reason for, in tending other's interests, she may have to give up some of her self-interests.

Let us look at an example where one's moral good is incompatible with one's non-moral good. Beth quit her accountant job and seriously started a writing career. But years went by, her writings were one by one refused by publishers. She felt so frustrated that she would like to return to her old job. But she was too ashamed to admit her failure.

One day, to her surprise, she found some novel manuscripts in an old box left by her late mother. She immediately sent one of them under her own name to a publisher. This time, it was accepted. She was so happy with the news although she knew that, from a moral point of view, what she did was not good. Though she did feel a bit guilty, she also felt the contribution of the 'success' to her psychological well-being offset the negative feelings brought by the belief that she did something morally wrong. Suppose that Beth would want more to do what she did when she was fully informed because the full information would make her be sure that nobody would ever know the truth. Railton will have to accept what Beth did as something intrinsically good for her. In this way, however, her intrinsic good is in conflict with her moral good. If Beth is naturally motivated to have concerned her psychological well-being only, Railton will not be able to convince her to do the morally right thing, for the moral reason is an external reason and he does not offer a resource to explain whether it can override the internal non-moral reason.

It seems to me that the theories which can account for the normativity of morality and are compatible with Railton's account are those of modern Kantians, such as John Rawls and Jürgen Habermas, because he defines social rationality as "what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information" (Railton 1986b, 150). If Kantian agents choose Rawls' two principles of justice or what all the members of a society under ideal conditions will agree on as a law, they cannot violate the law that they give themselves, because the autonomy of moral agent is the source of moral normativity. However, the Kantian reason for acting morally is an external one, for Beth can ask,

‘Why should I be moral if I can achieve more non-morally intrinsic good when violating moral standards?’ Therefore, it is clear that Railton’s justification of moral goodness is fundamentally different from the one he offers for non-moral goodness even though he takes them as the same.

One possible response from Railton is that an agent might not voluntarily take other people’s interests into account, but he will, as a matter of fact, be forced to. Railton believes that a social arrangement that departs from social rationality by significantly discounting the interests of a particular group has a potential for dissatisfaction and unrest. In other words, a set of institutions is morally justified if and only if it satisfies an ideal of social rationality. For example, slaves’ interests were significantly discounted in a society where slavery was permitted. This is why slaves fought for their freedom and the abolition of the slavery. The owners of the slaves might not want to give up the benefits they obtained from slavery, but they were forced to adopt a new social arrangement. Railton seems to suggest that a certain morality will be justified if nobody in the society claims his interests is discounted or nobody is capable of initiating any social change. But why should the fact that disparity from the ideal causes disfavored groups to initiate changes tell us what is justified? If a social ideology is so effective that all the members in a disfavored group are happy enough not to initiate any change or if the people whose interests are significantly ignored are too weak to fight, we still want to know what social arrangement can or cannot be justified anyway. Moreover, Railton seems to have ignored the details of how moral changes are brought about in historical cultures, for instance, by missionary activities or the changes in standards of sexual behaviour brought about by introduction of the pill. The explanatory power of Railton’s

ideal of social rationality comports best only with a broad, rather undetailed look at history.

(2) The Problem of Moral Convergence

Railton's proposal suggests that moral properties can be known to be, or be identical to, natural properties, just as the property of being water is identical to the property of being H₂O. If he is right, the truth of moral propositions can presumably be known essentially in the same way the truth of propositions in empirical science can be known. Moreover, we should expect there to be agreement among factually informed, intelligent people in relevantly similar situations just to the degree that there is about the affirmations of empirical science. As a matter of fact, however, there seems to be great diversity of valuations of moral judgments around the world. And it seems we must concede that moral standards have not converged historically, around the world, to the same extent as have scientific beliefs. What might Railton say in reply to this?

Railton has offered the following sort of explanation of the extent and persistence of observed moral disagreement. First, the category of moral evaluation as found in ordinary language is hardly a determinate one. Secondly, we find wide variation in the circumstances under which people live; these differences do much to explain variation in moral judgement. Thirdly, the existence within and across societies of significant conflicts of interest and differences in social role or group identity. Finally, moral judgements typically concern complex interpersonal phenomena where relevant particular and general facts are difficult to ascertain and where multiple dimensions of assessment may be at work, giving rise to problems of comparison. Appeal to similar

considerations could help to explain why in some areas of morality there is significant moral agreement, such as the keeping of promises, truth-telling, and strictures against theft and assault. (Railton 1993b, 281-283)

Although all the reasons given above for explaining the existence of moral disagreements seem to me plausible, I suspect Railton's account itself does not support the thesis he argues for even when the agreement on empirical facts has been achieved. He may argue that the principle of the supervenience of moral properties upon natural properties strongly supports moral convergence among people who hold same factual beliefs. For example: Two situations can differ in their moral qualities just in case they also differ in some natural qualities and if one judges two situations to be morally distinct, one is committed to judging them to be different as well in their natural properties. According to Railton, this principle implies the possibility of moral convergence.

By contrast, however, I believe the principle of supervenience says nothing about moral convergence among people but only about the consistent moral judgments within one individual. Suppose that there are two situations X and Y, an individual A makes a judgement that X is morally good while he judges Y to be morally bad. Thus, A commits himself to judging them to be different in their natural properties. In other words, if A denies that X's and Y's natural properties are different, he would have no reason to judge their moral qualities to be different. Moreover, if A judges X as moral at this time, he must make the same judgment next time when A considers X1 which has no relevant difference from X. Since the supervenience Railton is talking about is asymmetric, the non-moral must constitute or produce the moral. However, all of these arguments say

nothing about what another individual, say, B, should think about the situations X and Y. B may agree with A on all the facts concerning situations X and Y, but still makes an assessment that is contrary to A's by judging that X is morally bad while Y is morally good. When B judges X1 which is similar to X, she makes a judgement, which is consistent with her earlier one, that X1 is morally bad. In this way, A and B can be both consistent with themselves, but conflict with each other.

Moral cognitivists believe that, if agents A and B agree on all the facts concerning situation X and Y, they will not disagree on their moral characters. However, moral non-cognitivists do not think so. They believe that there are no moral facts or properties and to understand moral evaluations is to understand that moral judgments express a human response of a distinctive sort to the world, not the discovery of a special, additional set of worldly features. They think that human responses to a particular natural situation can conflict with each other while the cognitivists deny there is such a freedom. What I am trying to argue is that, although Railton is a moral cognitivist, his theory actually supports a non-cognitivist claim.

Railton advocates a relational notion of good and rejects value absolutism. According to the relational good thesis, any good thing only can be good for a certain subject because the subject's good is closely connected to his particular reduction basis. Something good for one individual may be no good at all for another one. Nothing can be identically good for two different individuals with different reduction bases. One's reduction basis consists in not only one's capacity, circumstances, but also one's emotional system. Only when all the elements in one reduction basis are the same as those in another reduction basis, when two reduction bases are compared, can the value

judgments given rise to by the two reduction bases be the same. More specifically, when two agents are different from each other regarding their motivational systems, they will not make the same moral judgment even when they agree on all the facts concerning the same object. This conclusion, if correct, renders Railton susceptible to a charge, his intentions to the contrary notwithstanding, of being committed to non-cognitivism which holds that genuine moral disagreement involves at bottom disagreement in attitude and therefore is not necessarily resolvable by empirical investigation.

Chapter 3: The Objectivity and Non-Hypotheticalness of Moral Judgements

1. Railton's Account

(1) Belief

In order to argue that moral judgements are objective and non-hypothetical, Railton starts from arguing for the objectivity and non-hypotheticalness of beliefs. He believes that there are facts about what an individual believes that can contribute to the explanation of his behaviour in ways not dependent upon his interpretive scheme. Since belief is necessarily aiming at truth, it is not hypothetical upon any contingent aim of the believer. He suggests that belief furnishes us a model of how there can be a genuine issue of realism and objectivity in the domain of morality.

Railton rejects Platonism in the moral domain because he believes the nature of realism about any domain depends upon the nature of that domain itself. Since there is a big difference between moral judgment and ordinary judgement about objects in the world around us, moral properties could hardly do their job if they were prior to or independent of any condition or experience of subjects whereas physical objects could hardly do their job if they were not. He holds that a realist attitude does not necessarily bring with it a commitment to a sort of metaphysical independence, namely, to think that

the nature of moral facts owe nothing to our natures, because 'subject-ive' does not mean subjective⁹.

Railton recognizes that radical nonsubject-ivity, where it can sensibly be sustained, does help to secure several seeming hallmarks of realist thought. Nonsubject-ivity belongs to the job description of physical objects. To believe that at least some physical object statements, literally understood, are literally true is to accept the existence of entities with this sort of explanatory potential as part of the furnishings of the actual world. If a domain of entities and properties D is independent of experience, for example, then it has at least three interesting kinds of explanatory potential. First, it can in principle support non-circular explanations of the patterns in our experience. Second, D can in principle support explanations unmediated by any sort of experience. This, taken together with the first, establishes an impressive 'width of cosmological role' for D. Third, these two kinds of explanatory potential enable us to see clearly how explanations invoking the properties and entities of D can in principle pass an 'Attitude Test' of a kind suggested by Gilbert Harman: if we were to replace reference to these properties and entities in our explanations with reference only to our attitudes about D combined with associated non-D facts, genuine explanatory information would be lost. This sort of radical nonsubject-ivity is said to require entities and properties that are 'out there' in the world, independent of the projections of our beliefs.

On the contrary to the case of physical objects, beliefs are the possessions of subjects. It is impossible to require beliefs to constitute *a priori* or independent order or to

⁹ Railton uses a hyphen between 'subject' and 'ive' to express, as he puts it, "the notion of that which is essentially connected with the existence or experiences of subjects" (Railton 1995a, 263). He believes that the domain of morality is one with subjects but is not subjective, which can be described as an objective 'subject-ive' domain. Similarly, 'nonsubject-ive' means without subjects.

possess a wide cosmological role. Commonsense holds that belief is an internal state of individuals that has a causal or quasi-causal role in shaping their conduct. This explanatory role is not always mediated by self-conception or interpretation. Therefore, belief discourse would pass the Attitude Test, which makes it possible for Railton to suggest that belief can be an objective feature of the subject-ive part of the world. If we are to make sense of the question of realism about belief, then there must be a way of being a realist about other subject-ive domains.

Will belief discourse also be non-hypothetical? Railton's answer is yes. He says, "To show that a norm or reason is non-hypothetical is not to show that it is utterly without condition. It is only to show that it would necessarily apply to any agent as such, regardless of her contingent personal ends" (Railton 1997, 58). From our ordinary observation, it is very likely that two individuals with the same evidence and same background beliefs would have the same reasons for believing any given proposition, regardless of possible differences in their personal goals. The philosophical argument that Railton offers is a constitutive one. In order for a propositional attitude to become an attitude of belief, it cannot represent itself as wholly unaccountable to truth or evidence. As a result, being in the belief business is a precondition of agency, as Railton puts it, "as an agent you must possess beliefs; as a believer you must represent certain of your propositional attitudes as accountable to truth and as disciplined by truth-orientated norms" (Railton 1997, 59). This constitutive argument concerns not an agent's actual conformity to epistemic norms but her self-representation as such.

(2) Action

Now let us consider whether there is anything in the realm of action which plays the role that truth plays in the realm of belief. This is important because the central purpose of Railton's project is to show that actions in the practical domain are also non-hypothetical and objective. We can think ourselves as looking for a feature F such that one must represent oneself as 'aiming at' F in action in approximately the same sense in which one must represent oneself as 'aiming at' truth in belief. There are two answers that suggest themselves. One comes from High Brow, someone such as Aristotle, who believes that an agent engaged in deliberate action necessarily aims at the good. Another comes from Low Brow, someone such as Hume, who holds that 'reason is the slave of the passions' and denies the plausibility of the thesis that one's action necessarily aims at the good.

Railton believes that both approaches succeed in showing the connection between norms and actions. The High road and the Low road thus both lead to non-hypothetical requirements for practical reasoning. The path in each case proceeds via a constitutive argument that has much in common with the constitutive argument made for theoretical reasoning. The linkage is made to the nature of agency, thereby avoiding dependence upon contingent personal goals.

The proponents of High Brow believe that deliberation and choice constitutively involve representing what one chooses as in some sense good. An agent might lack the nerve, or will, or energy to follow through on her judgements of choiceworthiness. But the extent that she is aware of this, she will recognize this as a lack, a gap between what she values and what she does. So they believe those among our fellow adults who have a

latent capacity to represent a course of conduct as good, but who fail to develop or exercise it can be properly considered as people who lack in the appropriate respects agency.

On the contrary, Humean individuals deny that their actions must aim at the good although they believe they possess full agency. For them, what is constitutively involved in deliberation is not choiceworthiness or the goodness of the end but the instrumental rationality of choosing the right means to realize an existing end. Railton uses two examples to illustrate the point of instrumental rationality. In one example¹⁰, when two children were ready to go to the beach, both of whom wanted very much to go and neither of whom liked the car ride, the first child refused to get into the car while the second one voluntarily did it. For the second child, the desire speaks on behalf of means toward its fulfillment, even unwanted means. In another example, two people both desired to take heroin. One was injecting himself in the right way while the other one swallowed it and choked to death. Although taking heroin is not a rational action, we do see one person's conduct as having a rationale in terms of what he believes and desires and the other's not. These two examples purport to show that the notion of an agent having and acting on ends of Low Brow is different from the notion of agent aiming at the good of High Brow. According to Low Brow, an agent's deliberations, decisions, and actions are in objective conformity with instrumental rationality and the connection between the action and the norm is non-hypothetical, because an agent as such must have ends and must also see herself as deliberating in a way that gives weight to the realization of what she takes to be her ends, independently of what these particular ends might be.

¹⁰ This example is originally given by Jean Hampton.

Are actions in the practical domain also objective? Railton believes that they do possess the sort of objectivity we found in the case of belief. For example, they can pass the Attitude Test and play an explanatory role in explaining an individual's behaviors or a society's social events. He argues, "The injustice of segregation helps explain why it produced widespread alienation and discontent and an eventual movement for change" and "[t]he deviousness of a parent helps explain the insecurity of his child later in life" (Railton 1995a, 272). Many moral explanations would pass the Attitude Test in the following sense: they could not be replaced without loss by an explanation adverting only to the moral beliefs of those involved. He remarks, "A social order's injustice may produce alienation and dysfunction well before any articulated sense of its injustice surfaces within the population. A parent's dishonesty may undermine the solidity of a child's sense of self long before any glimmer of a moral critique of the parent emerges in the child's mind" (ibid., 273). These explanatory accounts are in the first instance unmediated by moral concepts or beliefs.

As far as the practical domain is concerned, Railton is successful in arguing for the objectivity and non-hypotheticalness of actions and even the objectivity of moral actions, but he fails to show that moral actions are also non-hypothetical. I believe the failure is caused by his taking the Low Brow's approach. According to Low Brow, the only reason that we can use in our moral reasoning is instrumental reason. However, instrumental reason cannot tell us which end we should choose to have. If so, we will not be required by instrumental rationality alone to pursue a moral end over an immoral end. What we are required to do is merely to take a right means to realize an existing end

whatever this end happens to be. Therefore, a moral action is to be taken or not hypothetically depends upon an agent's existing desires or goals.

2. The Problem of the Argument of Non-Hypotheticalness

It seems to me that what the non-hypotheticalness of moral actions in the last section means is very different than the non-hypotheticalness of morality in the section "The Inescapability of Morality." Since High Brow's approach is not compatible with Railton's whole project, Railton can only take Low Brow's approach. However, Humean arguments are unable to carry us any further than showing the non-hypotheticalness of instrumental rationality. In other words, Humean agents cannot argue for the non-hypotheticalness of moral judgments from both an internalist and instrumentalist point of view. When Railton successfully argues that the scope of moral evaluations is very much like that of logical evaluations in the sense that they are both non-hypothetical, the moral psychology that he appeals to is externalist rather than internalist. That is to say, whether or not one cares about moral values, one's actions can be properly judged as morally right or wrong. But from an internalist point of view, a moral reason can only become a reason when he is motivated to obey moral norms. Therefore, it is hard for Humean agents to argue for the non-hypotheticalness of moral standards from an internalist point of view.

Another problem given rise to by the argument of non-hypotheticalness is one that confronts all constitutive arguments. Railton has discussed these problems, but he believes they can be met by appropriate revisions of constitutive arguments. Constitutive arguments have the strength that comes from purportedly necessary connections. But this strength can also be a weakness. If the necessity turns out to be linguistic, the argument

may lack the power to sustain a substantive conclusion. But if the necessity is of a more substantive kind, it would be impossible to criticize an agent who refuses to give deliberative weight to his own acknowledged end, because, if the constitutive argument is right, an agent cannot fail to assign weight to his end. A very problematic idea resulting from this kind of constitutive argument is that whatever an agent is doing, she is pursuing something that is her real end no matter what end she believes she has.

Railton believes this problem can be explained away by appeal to the charge of irrationality. He says, "Failures of rationality come in many shapes and sizes, and do not form a unified type. But it may be useful to think of some forms of theoretical or practical irrationality as instances of incomplete-yet-nearly-complete approximation of believing or having an end" (Railton 1997, 73). A person can fail to do what she takes as her end because of many reasons, such as weakness of the will, lack of courage, shyness, etc. For example, if a person sets herself a long-time goal to write a novel, she might often be taken away by some small attractions from her desk. She knows that she should concentrate more on her work than she does because she considers writing as the most important goal in her life. She may none the less find herself giving it no weight in certain deliberations. Railton suggests that we should remove from the constitutive arguments an artificial rigidity, for there are a whole bunch of possibilities between the two extremes of having complete agency and having no agency at all.

But it seems to me that the original line of criticism of the constitutive argument is still on the mark. One could argue that it is true that a rational person can be irrational in certain deliberations, but it is not clear we should consider something as one's end when one cannot, in normal circumstances, actually pursue it most of the time. Suppose,

for instance, Lucy following Beth quit her accountant job and set herself a goal to write a long fiction. Sometimes she wasted a whole day to do shopping or to watch TV when she knew that she should have worked. Because of the existence of these facts, she realized that she was not a fully rational person because she could fail to do what her important end required her to do. However, after five years, her book was published and she became successful in her writing career. Beth was also trying to do what Lucy had done during those five years, but she wasted all the time on taking vacations, talking, and doing housework. She just could not sit down to get her job done. Was writing really Beth's goal when she could not bring herself to pursue it most of the time? According to Railton's theory of reduction basis, we should come to conclude that writing career is not good for Beth in the sense that she does not possess the right capability and personality that writing career requires in her reduction basis. So, what she was doing during those five years was actually something oriented to her real end which was different from the end of becoming a full-time writer that she took herself to have had. Railton could reply that her real end was exactly what she took herself to have had rather than what her actual actions implied. But one may question why one's real end should be characterized as such. Beth might be being ignorant during all those five years in not knowing what her real end was, but she was not being irrational in assigning no weight to her end of becoming a writer because writing was not her end at all though she thought it so. If so, a critic of the constitutive argument can insist that the constitutive argument cannot tell us how to distinguish an agent's actual actions from the actions prescribed by her end.

Chapter 4: Objections to Non-Cognitivism

The main difficulties that non-cognitivism confronts are as follows: non-cognitivism is unable to explain why moral statements are made in fact-stating form; non-cognitivists cannot admit any knowledge such as that something is morally wrong; non-cognitivist views cannot explain the authority of moral judgments. Railton discussed all these problems in his articles. Here I shall focus on another one that is related to all of them. Railton believes that the non-cognitivist conception of normativity can rightly capture the motivational component of moral evaluation, but needs supplementation by some account of the *ground* of it. The naturalist's understanding would provide a better theory which explains both the ground of moral evaluation and the connection between moral assessment and motivation, as a matter of social-psychological fact rather than conceptual necessity.

1. What the Non-Cognitivist Helps Us to See

Non-cognitivists have pointed out that the categorical reasons that the Kantian approach requires simply do not exist because practical rationality does not provide us with ends that are wholly independent of desire, manifest or dispositional. They also offer an explanation of our moral practice. According to non-cognitivism, to call something morally good is to express one's acceptance of a system of norms that permits it. For example, I say, if I am a non-cognitivist, 'X is morally good' means 'I accept a system of norms that permits X.' More precisely, it says not that I state that I accept a system of norms that permits X, but that I express my acceptance of a system of norms that permits

X, because to express a state of mind is not to say that one is already in it. Another part of why my attitude can be termed a moral attitude is that my attitude that X is morally good need not be conditioned by my sense of the particular desires or inclinations of mine. So my judgment is non-hypothetical. Moreover, If this attitude is fully internalized in me, it will function in an autonomous way psychologically. As a result, “We can have, act on, express, and be influenced by internalized, non-hypothetical attitudes without genuine categoricalness, that is, we can have the psychological equivalent of categoricalness in a world devoid of Kantian categorical reasons” (Railton 1993b, 286). If anyone worries about whether the detaching of categoricalness from moral reasons will cause the failure for them to be universal and necessary, non-cognitivists could reply that the force of categoricalness is not that of what we observe in our ordinary moral practice.

2. Objections to the Attitudinal Account of Normativity

Non-cognitivism can be rejected as an account unable to capture the peculiar importance or the ‘inescapability’ that we typically attach to the moral judgements. The normativity it accommodates is disappointingly weak. Railton remarks that even if a non-cognitivist account can be non-hypothetical at a certain level in the sense that a specific moral evaluation has nothing to do with any one’s particular ends, it must be hypothetical in the sense that an agent’s being able to make certain moral judgement depends on the fact that he happens to have this particular attitude. Consequently, a person on a non-cognitivist account can easily escape moral assessment. His actions can be criticized as morally wrong only if he happens to have this particular attitude but nonetheless does not act accordingly. If he happens to lack this attitude, then no matter what he does, he

should not be criticized or evaluated from a moral point of view. Railton rightly remarks that the normativity of moral evaluations is weaker than categoricalness but nevertheless is stronger than non-cognitivist non-hypotheticalness.

Another objection to non-cognitivism concerns the plurality and diversity of non-hypothetical attitudes that we find within individuals, which can be questions of taste or merely good manners. For instance, a person may have a very negative attitude toward the noise of eating and this attitude does not depend upon any particular desires or goals that he happens to have or anyone else happens to have. He does not like this kind of noise no matter who makes it, even himself. But this attitude is not a moral one. What the non-cognitivist needs is some way of characterizing a class of attitudes and expressions that captures the distinctiveness of moral evaluation.

3. The Advantage of Naturalistic Cognitivism

According to Railton, the specific normative nature of moral assessment is accommodated by looking beyond the non-hypothetical character of attitudes distinguished as moral attitudes to their characteristic ground or function. We can ask a further question, 'Why do moral attitudes have a special standing or importance in comparison to other non-hypothetical attitudes?' The non-cognitivist may point to the conspicuous importance to social life of this characteristic ground or function. This is why Railton favors an account of naturalistic cognitivism, which complements the foundational part of non-cognitivist account. On the naturalistic cognitivist view, an act or practice is better to the extent that it contributes to aggregate well-being impartially considered. First, it permits assessment that is non-hypothetical with regard to the agent's

own desires. Second, although the contribution to well-being does not encompass everything that people have found significant in moral assessment, a standard based upon such a contribution does capture something that has been widely held to be a significant part of morality. Third, the naturalistic standard provides a ground for moral assessment that would enable us to understand why moral considerations have a certain importance in social life. Railton thinks that the importance of well-being is not circularly dependent upon adopting a moral point of view of the sort the standard itself involves, since well-being and freedom from suffering are non-moral goods of the first order. Therefore, he claims that the naturalistic cognitivism not only completes a non-cognitivist account but also has its advantages.

Chapter 5: Ethical Naturalism

Suppose that moral realism is true and that there are moral facts, then what are these moral facts like? Are they natural properties detected by natural power, supernatural properties detected by supernatural power, or *sui generis* detected by mysterious intuitions? As an ethical naturalist, Railton holds that moral facts and properties are natural facts and properties which can be detected or studied by empirical sciences. According to him, there are two kinds of ethical naturalism which are *methodological naturalism* and *substantive naturalism*. A philosopher can be a substantive naturalist without being a methodological naturalist by arguing on *a priori* grounds that the meaning of moral terms is naturalistic. Similarly, a methodological naturalist, such as Hume, does not have to make a substantive naturalist conclusion. Railton is a naturalist in both senses. Substantive naturalism can be further divided into two forms: reductionism and non-reductionism. Railton took a reductionist approach to argue for his substantive naturalism in his earlier papers. Then he became a non-reductionist since about 1995.

1. Reductionism

A reductionist holds that to reduce one kind of property A to another kind B is to show that A is nothing but B. By showing the identification of B with A, he is trying to give an account of A in terms of B, which is given a privileged status. For example, water is H₂O; genes are nothing but DNA molecules; heat is mean kinetic energy of molecular movements.

Railton's conception of 'reduction basis' will make many people ask: Is he a reductionist? For him, both moral and non-moral values have their reduction bases. In "Facts and Values," he identifies one's good with one's wants and desires when one is fully informed. He thinks that "an individual's intrinsic good consists in attainment of what he would in idealized circumstances want to want for its own sake – or, more accurately, to pursue for its own sake (for wanting is only one way of pursuing) – were he to assume the place of his actual self" (Railton 1986a, 17). The examples in the earlier chapters have shown that what a person would actually want under idealized circumstances is objectively determined by his reduction basis. For example, Lonnie's physical condition of dehydration is the reduction basis for determining the goodness of his drinking clear water; Beth's personality and capability is the reduction basis for determining the badness of her being a writer. It can be seen that values, on such a conception, are determined by their reduction bases. However, in his first two papers of 1986, he is hesitant to make a decisive claim that values can be reduced to or identified with natural properties. When he talks about the features of the specific kind of moral realism that he has in mind, he describes reductionism as something to study "Are moral properties reducible to, or do they in some weaker sense supervene upon, non-moral properties?" (Railton 1986b, 138). In the same paper, when he discusses moral realism, he expresses a conjecture that "moral properties supervene upon natural properties, and may be reducible to them" (*ibid.*).

Despite these hesitations, he maintains a reductionist position in his later essays. For instance, in "Naturalism and Prescriptivity," he claims that his naturalistic approach attempts to "locate value properties among features of the world that are accessible to us

through ordinary experience and that play a role in empirical explanations” and thus he treats “value properties as natural properties” (Railton 1989, 154). Although an ethical naturalist does not have to be reductionist about value, for he may hold that “moral properties might simultaneously be natural and *sui generis*,” Railton finds such a position unsatisfactory and is determined to entertain “a reductionist hypothesis, a synthetic identification of the property of moral value with a complex non-moral property”(Railton 1993c, 317). The most important reason for Railton to be a reductionist is that he believes “the identification can contribute to our understanding of morality and its place in our world – including such matters as semantic and epistemic access to moral properties – while preserving important features of the normative role of moral value”(ibid.). Another reason is that he believes “the naturalist who would vindicate the cognitive status of value judgments is not required to deny the possibility of reduction, for some reductions are vindications – successful reduction of water to H₂O reinforces, rather than impugning, our sense that there really is water” (Railton 1989, 161). Moreover, he suggests that “[w]hether a reduction is vindicative or eliminative will depend upon the specific character of what is being reduced and what the reduction basis looks like” (ibid.).

To achieve a vindicating reduction of good, he must identify it with a set of complex natural properties that, to a significant extent, permits one to account for the correlations and truisms associated with ‘good’ – i.e., is at most tolerably revisionist – and that at the same time can plausibly serve as the basis of the normative function of this term. For Railton, the reductionist account of the good for a person in terms of his being in a distinctive experiential state, say, happiness, can be understood as an analogy to the reductionist account of a vessel’s seaworthiness in terms of its certain dispositional

features. His sort of naturalist “reduces moral properties to complex social-psychological phenomena, not to extra-human Nature” (Railton 1993c, 325). Because values can be identified with complex descriptive properties, Railton believes that both a value and its reduction basis can play a causal role in the explanation of our beliefs. Therefore it makes no sense to ask “whether the causal work is ‘really’ being done by value or by its reduction basis” (Railton 1989, 161). If the identification holds, then each does.

“A successful, vindicative, identificatory reduction of moral value,” Railton remarks, “could help the case for moral realism if it could provide a foundation for moral value that removes some of the mystery surrounding this notion while at the same time providing support for many of its central characteristics” (Railton 1993c, 317). For example, he continues, “a reductive identification might indicate how moral value could be, on the one hand, a property to which we have epistemic and semantic access, and, on the other, a property which, through identifiable psychological processes, could engage people motivationally in the ways characteristic of moral properties” (*ibid.*). Another advantage of reductionism is that “a reductive identification might help us to understand how certain patterns of moral disagreement could persist, despite the cognitive character of moral discourse, and to understand as well the supervenience of the moral upon the non-moral to which Moore drew notice” (*ibid.*). It seems to me that Railton, at one time, firmly believed that reductionism presents so many advantages that it is the best form of substantive naturalism about moral values.

2. The Problem of Reductionism

First of all, someone who believes that Moore's open question argument has defeated ethical naturalism would think that Railton has committed what Moore called a naturalistic fallacy, because Railton has reduced values to natural properties. The second charge of reductionism comes from David Wiggins who appears sympathetic to a view he attributes to Moore, that reductionist ethical naturalism is 'crude' and 'scientific.' The third charge is from Railton himself when he argues against reductionism in his later writings.

"If these are open questions, then why is my naturalist so concerned to advance a particular reduction?" Railton asks (Railton 1993c, 320). "One reason for pursuing the definiteness involved in hypothesizing a given reduction," he continues to answer the question, "is that it is difficult to assess the adequacy of a cognitivist account of moral discourse which, like some contemporary anti-reductionist accounts, says little that is definite about the substantive content of that discourse" (*ibid.*). But if he wants to say more about the substantive content of moral discourse by adopting reductionism, will he be able to respond to Moore's charge?

He can claim that Moore's open question argument rests on a false assumption that synonymy is a test of property identity. According to Moore, 'Is an unmarried male a bachelor?' is a closed question because 'unmarried male' and 'bachelor' are synonymous by definition and nobody who understands the two words can sensibly ask 'Is an unmarried male a bachelor?' By contrast, 'Is water H₂O?' is an open question because 'water' is not synonymous with 'H₂O' and anybody can meaningfully question if water is H₂O. Moore's open question argument is designed to show not only the failure of

synonymy between moral and natural terms but also the failure of property identity between moral and natural properties. What Railton wants to argue is that synonymy is neither a necessary nor a sufficient condition for property identity. For instance, the claim that 'Water is H₂O' can be understood as an expression that 'water' and 'H₂O' both refer to the same property although 'water' is not synonymous with 'H₂O'. The scientific identification of water with H₂O is a significant empirical discovery rather than a result of *a priori* or definitional analysis. Another example is the talk about 'the morning star' and 'the evening star.' Although the two terms are not synonymous, they refer to the same object, namely, the same star at different times. Railton concedes that the question 'Is water H₂O?' is not conceptually closed and "this reforming definition could itself be altered or dropped under the pressure of further empirical discoveries or theoretical developments," but, he adds, the definition "does not really purport to close the question with conceptual necessity" (Railton 1989, 158). Therefore, the charge of 'naturalistic fallacy' can be refuted by an argument that the synonymy test is not enough for property identity and that an open question should be left open conceptually without affecting an ethical naturalistic claim of a synthetic form.

Now let us consider how Railton could respond to Wiggins' criticism. Wiggins suggests that reductionist naturalism is 'crude' and 'scientific.' Railton replies by explaining what he believes scientism is first. He says, "Scientism, I suppose, is the view that all genuine phenomena are within the scope of the scientific method, and that no other means can issue in any legitimate form of knowing" (Railton 1993c, 318). Then he explains, "a naturalist interested in reduction in ethics need not believe that there is such a thing as 'the scientific method' (certainly the history and sociology of science give us

reason for doubting this) or that all truths in ethics are accessible to us through some mechanical, fact-gathering, 'value-neutral' procedure" (*ibid.*). He points out that Wiggins' criticism is meant to come not with the label 'scientism' as such, but with an adjective Wiggins uses in its company: 'crude'. If 'crude' means revisionism, Railton thinks it should not be feared as such, because any reduction would bring about revision and some revisions are limited in a tolerable degree. "Typically, even vindicative reductions are in some measure deflationary," he remarks, "if only of a pre-theoretical presumption of the *sui generis* character of the area of discourse in question" (*ibid.*, 319). For example, the reduction of water to H₂O denies water the special status it has been accorded as one of the few basic, irreducible elements of the universe. Similarly, any reduction of good will deny it the elemental status Moore claimed it to have. But he finds nothing crude in the reduction of water to H₂O, and fails to see why a reduction of good must be any worse in this regard. He says that he believes "an appropriate naturalistic reduction can furnish essential ingredients for this sort of vindicative account of morality" while "Wiggins appears to think that reduction stands in the way of vindication and that treating moral properties as *sui generis* somehow helps" (*ibid.*, 325). It can be seen that, between the two options: to reduce moral properties to natural ones and to treat moral properties as *sui generis*, Railton believes the reductionist option is more plausible than the other one. At this point, what he does not realize is that there should be the third option: to treat moral properties as natural and non-reducible.

Moreover, he believes his sort of reductionism can account for not only "how obligation is realized in social practices through the internalization of norms" but also "about the ways individuals draw meaning and significance from the social practices with

which they identify, in such a way that they may have grounds for action in accord with internalized norms but contrary to personal inclination” (*ibid.*), which means that his reductionist naturalism can have the advantages that are supposed to belong only to Wiggins’ *sui generis* account.

The third charge of Railton’s reductionist account is from himself, which is concerning the explanatory role of moral values. For example, the fact that a young adult suffers from being unable to trust anyone can be explained by his parents’ dishonesty. Were the moral features to be identical with or reducible to underlying good-making features, explanatory credit would then always be given to the latter and the moral features themselves will play no causal role at all. In this example, the parents’ dishonesty will be reduced to the fact, say, that they often cheat on their children and other people. On a reductionist view, it is this fact rather than dishonesty that actually plays an explanatory role in explaining the young person’s sufferings. Is this account plausible? Railton denies it so in his latter writings. “Why not always give all the explanatory credit to the underlying physical facts?” he argues, because “[p]sychologists, sociologists, and biologists have formulated worthwhile theories using the distinctive categories of their disciplines,” and “such theories contain generalizations [which] may not be strict or exceptionless, but that do illuminate functional connections, causal dependencies, and other relations at a particular (supervenient) level of description of the phenomena” (Railton 1998, 179). He believes that these relations support counterfactuals and afford explanatory insights that would not be evident at the subjacent level of a purely physical description of events. For example:

evolutionary biologists might explain patterns of male appearance or behavior partly in terms of sexual selection, in which females favor males able to visibly

manifest bodily health, so that we find showy (and often otherwise inefficient) coloration, physique, display, etc. in males across a wide range of species. Now down at the level of physics, some or other complex microphysical process does in a law-abiding way yield all the molecular motion that underlies these evolutionary histories. But that account, though itself explanatory, would not supply the additional explanatory insight evolutionary theory affords. Indeed, if complex life-forms have evolved on other planets, organisms might have a quite different physio-chemical base yet still show patterns of sexual dimorphism, sexual selection, and mating display. Neither physical nor biological explanation drives the other out; both contribute to the full explanatory story. (*ibid.*)

The example has shown that physical theories cannot always afford the explanatory insights that biological theories can, although physics is the underlying theory of biology. By the same token, non-moral properties cannot always afford the explanatory insights that moral properties can. From Railton's latter point of view, it is very problematic to think that moral properties are reducible to or identical with natural ones. Therefore, reductionism is problematic.

3. Non-Reductionism

Since about 1995, Railton started arguing for a non-reductionist form of substantive ethical naturalism. In "Made in the Shade: Moral Compatibilism and the Aims of Moral Theory" and "Moral Explanation and Moral Objectivity," he claims that the relation between moral properties and natural properties is supervenience rather than identity and that values asymmetrically supervene on non-values. 'Supervenience' is a relationship between two kinds of properties that things may have. According to Simon Blackburn's writing in *Routledge Encyclopedia of Philosophy*, "It refers to the way in

which one kind of property may only be present in virtue of the presence of some other kind: a thing can only possess a property of the first, supervening kind because it has properties of the underlying kind, but once the underlying kind is fixed, then the properties of the first kind are fixed as well. The supervening features exist only because of the underlying, or 'subjacent' properties, and these are sufficient to determine how the supervening features come out" (Blackburn 1998, 235). For example, a melody supervenes on a sequence of notes and a person's generosity supervenes on what he actually thinks and does. Railton remarks, "Bob's honesty, for example, supervenes on his psychology, and Transylvania's injustice supervenes on its institutional structures and social inequalities. It is these psychological, sociological, and political features, ...that are in fact doing the explaining" (Railton 1998, 178).

The idea of supervenience purports to prevent either a mysterious or an identical relationship between the properties of two different levels. On the one hand, if the relationship is too loose, it will become a mystery for it will not be able to account for how the lower level properties give rise to the ones of the upper level. On the other hand, if the relationship is too tight, the upper level properties become nothing but lower-level ones differently described. The conception of supervenience "offers the seductive promise of a path between full-scale reduction of upper-level to lower-level properties, and an uncomfortable dualism" and it "promises the relief from metaphysical anxiety, but without the costs of reduction" (Blackburn 1998, 237).

Railton suggests that, although moral judgments do not function semantically simply to report some natural state of affairs, they nonetheless have an intimate relation to such states of affairs, that is, moral qualities are constituted by or grounded in natural

qualities. This relation of constitution is asymmetric supervenience of the moral on the non-moral, as Railton recognizes that “the relation of the moral to the non-moral is not reciprocal, ... The non-moral must constitute or produce the moral” (Railton 1995b, 102). In arguing for asymmetric supervenience, Railton rejects the reductionism he once accepted by making the claim that “none of this entails a naturalistic reduction” (*ibid.*). When advocating non-reductionism, he also accepts the is/ought division. By contrast to the view he held in “Facts and Values,” he claims that there is a logical gap between ‘is’ and ‘ought.’ In this new position, he clearly points out that “[I]nnovations in empirical theory, for example, do not dictate innovations in ethics in the way that innovations in physics are typically thought to dictate innovations in chemistry” because “[t]here is a logical gap between any alleged fact and any conclusion expressed in moral terms” (*ibid.*, 104). Only when the is/ought gap is admitted, Railton suggests, can the autonomy of moral thought be secured, for the asymmetric supervenience of moral properties does not guarantee that fixing the natural properties of the actual world *a priori* guarantees the presence of any moral properties.

4. The Problem of Non-Reductionism

Relying on the failure of meaning implication between moral and non-moral terms, Moore’s open question argument can be used against the naturalist claim that moral properties, though non-identical with natural properties, are constituted by natural properties. The idea of meaning implication assumes that ‘A’ and ‘B’ are a pair of terms standing for properties A and B such that B constitutes A just in case that it is part of the meaning of term ‘B’ that B things are A. Accordingly, we can say that moral properties

supervene on natural properties or natural properties constitute moral properties only if it can be shown that it is part of the meaning of natural terms that natural properties are moral properties as well.

However, this kind of meaning implication is hard to establish between any two terms when one term is thought to be constituted by the other. For example, chairs are constituted by certain arrangements of microphysical particles, and social revolutions are constituted by certain social, political and economic events. But nothing in the meaning of the notion of 'certain arrangements of microphysical particles' implies that it is a chair and nothing in the meaning of the notion of 'certain social, political and economic events' implies that it is a social revolution. If these claims concerning chairs and social revolutions can be made correctly by natural and social sciences, why cannot we claim that moral properties are constituted by natural properties? The proponents of the supervenience thesis need to reject the meaning implication test implied in Moore's open question argument, for meaning implication is neither a necessary nor a sufficient condition for either constitution or supervenience. The naturalist, such as Railton, can concede that there are neither synonymies nor meaning implications between moral terms and nonmoral terms, but still maintain that moral facts and properties are constituted by natural and social scientific facts and properties for the naturalist's constitutive claims can be construed as expressing synthetic moral necessities.

Ethical naturalists believe that moral facts and properties supervene on natural facts and properties. It should be noted that the supervenience of moral facts and properties on natural facts and properties can follow from, but cannot establish, ethical naturalism. So another problem arising with Railton's non-reductionist position is

whether he can remain an ethical naturalist if the moral is not identical with the non-moral, for instance, if moral facts and properties are not identical with natural facts and properties. An ethical non-naturalist could question the metaphysical status of the moral properties of Railton's sort. He can say that if moral properties are not to be reduced to natural properties, then they are not natural but *sui generis* properties.

To respond to this *sui generis* charge, Railton can explain that, even though moral properties cannot be reduced to or identified with natural properties, they *are* natural properties in the sense that they are constituted by natural properties and they may be nothing over and above organized combinations of natural and social scientific properties. In other words, they are neither supernatural nor *sui generis*. To understand how moral properties can be natural without being identical with natural properties, let us consider two examples. For instance, we do not think psychological properties are *sui generis* even though they supervene upon, instead of being identical with, physiological properties. By the same token, we do not think evolutionary claims *sui generis* while we think they are supervenient upon, instead of being identical with, biological claims. We can equally understand how moral facts and properties can be natural without being reduced to the natural.

Chapter 6: Reflective Equilibrium

1. Coherentism

If we are ethical naturalists of Railton's sort, then how could we come to know moral facts, and how could we justify moral claims? An answer offered by Railton is Rawls' method of reflective equilibrium. While accepting the logical gap between 'is' and 'ought,' Railton believes that "reflective equilibrium knows of connections that are more-than-logical" (*ibid.*, 104-5). The method of reflective equilibrium is a form of coherentism, which holds that no beliefs are noninferentially justified in the sense that one's belief p is justified insofar as p is part of a coherent system of beliefs and p 's coherence at least partially explains why one holds p . Ethical theory, presumably normative theory, is based on a 'reflective equilibrium' of moral beliefs as well as with certain non-moral beliefs and conceptions, which is essentially John Rawls's method of wide reflective equilibrium. In order to explain how to arrive at the point of 'reflective equilibrium' when we try to justify a certain moral belief, I will quote Rawls' description of the justifying process when we use this method to search for the most favored description of 'initial situation.' Rawls writes:

In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at

others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. (Rawls 1971, 20)

The central idea of this passage is that, in order to construct a theory of ethics based on background beliefs and to move from them to some ethical principles, we need to take the whole set of background beliefs of a given person together, along with other data from our general knowledge, and seek a general theory when working from both ends. The result is called one's 'reflective equilibrium' or 'wide reflective equilibrium.' Rawls explains, "it is an equilibrium because at last our principles and judgments coincide" and "it is reflective since we know to what principles our judgments conform and the premises of their derivation" (*ibid.*). According to this method, no norms are superior to other norms for they must, in good Quinean fashion, be considered as a whole and this equilibrium is not necessarily stable because it is liable to be upset by further examinations. But that does not show there is anything wrong with them for fallibilism is inescapable. Such *perfect* stability is not to be had.

Railton's arguments for reflective equilibrium are shown through two examples (Railton 1995b). One is a park rangers' killing action and the other is Kant's theory. In the former, a group of park rangers were shooting repeatedly and effectively from a moving jeep into a large herd of deer. We would not think they have a good reason to do so if we only consider the killing scene isolated from other facts. But when we know that the park rangers were trying to reduce a deer population that otherwise would face certain starvation in the coming weeks of winter, we will understand the killing action is morally

right contrary to the seemingly cruel appearance. Reflective equilibrium requires us to put the belief in question into all the beliefs we have and see if it is coherent with them. This example is a good one to show how we make moral judgment through reflective equilibrium. Our verdicts about particular cases emerge from this evolving equilibrium, so we seek a balance that tends to minimize tensions and maximize coherence within our moral world. The second example is not as plausible as the first one. Railton considers Kant's dualist theory of phenomenal and noumenal as a form of wide reflective equilibrium. He also thinks that modern Kantians disturb the reflective equilibrium that Kant constructs by finding the notion of noumenal agency implausible. This example seems to me a very confusing one. I believe Kant's theory can be an example of compatibilism of free will and determinism but not one of reflective equilibrium.

It seems to me that Railton places himself in a better position to justify moral beliefs after he adopts a coherentist approach and rejects his former reductionist claims, because reductionism is more compatible with foundationalism which is in conflict with coherentism. I shall discuss foundationalism briefly in the next section.

2. Foundationalism

Foundationalism holds that one's belief p is justified just in case p is either foundational (i.e., noninferentially justified or self-justifying) or based on the appropriate kind of inference from foundational beliefs. A moral belief about a particular action or situation can be justified in part by reference to another, often more general, moral belief. At some point this process of justification must come to a stop with moral beliefs whose justification does not rest on other beliefs. These are the foundational beliefs. Intuitionism

is a form of foundationalism. Intuitionists hold that a moral belief can be justified either by being a foundational belief directly detected by intuitions or through a set of appropriate inferences from foundational beliefs. Many philosophers believe that foundationalism fails because there are no beliefs that are self-justifying and therefore there are no foundational beliefs to stop the regress while avoiding circularity.

Railton's reductionist account attempts to show that moral facts and properties are nothing but natural facts and properties. Consequentially, moral beliefs cannot be justified without appeal to natural beliefs which are considered more foundational than moral ones. However, Rawls' 'wide reflective equilibrium' sets out a justification process in which all the beliefs must be considered as a whole and no one sort of belief is superior over the other sort. Rawls remarks, "A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view" (Rawls 1971, 21). Therefore, Railton's reductionist ethical naturalism is a form of foundationalism rather than coherentism. Only when he gives up reductionism, can he consistently adopt 'wide reflective equilibrium' as a method of justification of moral beliefs. This is also what actually happened after 1995 when he started to stress on 'supervenience' and to argue for 'reflective equilibrium.'

Conclusion

Peter Railton takes an ethical naturalist approach to construct his moral realism. He holds that there are moral properties that can be true or false; moral facts and properties are natural facts and properties; we can come to know or justify moral judgments through wide reflective equilibrium. In trying to argue that moral facts and properties are natural facts and properties, he first adopts a reductionist position in which he believes that moral facts and properties can be identified with or reduced to natural facts and properties. Then he rejects this position and starts to advocate non-reductionism, according to which, moral facts and properties are non-reducible to, but rather, asymmetrically supervenient upon natural facts and properties. On this view, moral properties are constituted by natural properties. Therefore, two situations cannot differ in their moral character without relevant difference in their factual character. I believe Railton's moral realism in its later form is more successful than that in its earlier one. In examining his account, I also criticized his ideas of instrumental rationality, ideal observer theory, and internalism, etc.

Railton's ethical naturalistic theory seeks to account for moral and value judgments in virtue of natural and empirical properties while seeking to avoid the charge of Moore's 'naturalistic fallacy.' He believes that the Moorean semantic test of synonymy or meaning implication is neither a sufficient nor a necessary condition of supervenience of the moral upon the non-moral. Ethical naturalism should be synthetic rather than analytic in the sense that a moral claim does not seek to close Moore's 'open

question' conceptually or logically. The connection between the moral and the non-moral known through wide reflective equilibrium is an empirical one rather than a logical one.

BIBLIOGRAPHY

- Blackburn, Simon. (1993) *Essays in Quasi-Realism* (New York: Oxford University Press).
- _____. (1998) *Ruling Passions* (Oxford: Clarendon Press).
- _____. (1998) "Supervenience," in *Routledge Encyclopedia of Philosophy*, edited by Edward Craig.
- Boyd, Richard N. (1988) "How to Be a Moral Realist," in *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord (Ithaca: Cornell University Press), 105-135.
- Brandt, Richard B. (1969-70) "Rational Desires," *Proceeding of the American Philosophical Association* 43, 43-64.
- _____. (1979) *A Theory of the Good and the Right* (Oxford: Clarendon Press).
- _____. (1992) *Morality, Utilitarianism, and Rights* (Cambridge: Cambridge University Press).
- _____. (1996) *Facts, Values, and Morality* (Cambridge: Cambridge University Press).
- Brink, David O. (1986) "Internalism and Moral Realism," in *Southern Journal of Philosophy*, supp. vol. 24, 23-42.
- _____. (1989) *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).
- _____. (1991) "Rawlsian Constructivism in Moral Theory," in *Equality and Liberty: Analyzing Rawls and Nozick*, edited by J. Angelo Corlett (MacMillan), 196-216.
- Darwall, Stephen, Allan Gibbard, and Peter Railton. (1992) "Toward *Fin de Siecle* Ethics: Some Trends," in *Philosophical Review* 101, 115-189.

- Firth, Roderick. (1952) "Ethical Absolutism and the Ideal Observer," in *Readings in Ethical Theory*, edited by Wilfrid Sellars and John Hospers (Englewood Cliffs: Prentice-Hall, 1970), 200-221.
- Foot, Philippa. (1978) *Virtues and Vices* (Oxford: Basil Blackwell).
- Frankena, William K. (1939) "The Naturalistic Fallacy," in *Readings in Ethical Theory*, edited by Wilfrid Sellars and John Hospers (Englewood Cliffs: Prentice-Hall, 1970), 54-62.
- Gibbard, Allan. (1984) "A Noncognitivist Analysis of Rationality in Action," in *Social Theory and Practice* 9, 199-221.
- _____. (1985a) "Moral Judgment and the Acceptance of Norms," in *Ethics* 96, 5-21.
- _____. (1985b) "Reply to Sturgeon," in *Ethics* 96, 34-41.
- _____. (1990) *Wise Choices, Apt Feeling: A Theory of Normative Judgment* (Cambridge: Harvard University Press).
- _____. (1993a) "Precis of Wise Choices, Apt Feelings," in *Philosophical Issues*, 4, 33-35.
- _____. (1993b) "Reply to Railton," in *Philosophical Issues*, 4, 52-59.
- _____. (1994) "Meaning and Normativity," in *Philosophical Issues*, 5, 95-115.
- Hampton, Jean. (1996) "On Instrumental Rationality," in *Reason, Ethics and Society*, edited by Kurt Baier and J. B. Schneewind, 84-116.
- Hare, R. M. (1952) *The Language of Morals* (Oxford: Oxford University Press).
- _____. (1963) *Freedom and Reason* (Oxford: Oxford University Press).
- Hobbes, Thomas. (1651) *Leviathan*, edited by C. B. MacPherson (Harmondsworth, Middlesex: Penguin, 1981), 120.

- Korsgaard, Christine M. (1997) "The Normativity of Instrumental Reason," in *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Nigel Gaut, 215-254.
- Lewy, C. (1964) "G. E. Moore on the Naturalistic Fallacy," in *Proceedings of the British Academy* 50, 251-262.
- Mill, John Stuart. (1861) *Utilitarianism* (Indianapolis: Hackett, 1978).
- Moore, G. E. (1903) *Principia Ethica* (Cambridge: Cambridge University Press).
- Nielsen, Kai. (1977) "Mill's Proof of Utility," in *New Dimensions in the Humanities and Social Sciences*, edited by Harry R. Garvin (London: Associated University Presses), 110-123.
- _____. (1979) "On Deriving an Ought from an Is: A Retrospective Look," in *Review of Metaphysics* 32, 487-514.
- _____. (1989) *Why Be Moral?* (Amherst, New York: Prometheus Books).
- _____. (1996) *Naturalism without Foundations* (Amherst, New York: Prometheus Books).
- Putnam, Hilary. (1981) *Reason, Truth, and History* (Harvard: Cambridge University Press).
- _____. (1990) *Realism With a Human Face*, edited by James Conant (Cambridge, Mass.: Harvard University Press).
- Railton, Peter. (1986a) "Facts and Values," in *Philosophical Topics* 14, no.2, 5-31.
- _____. (1986b) "Moral Realism," in *Moral Discourse and Practice*, edited by Stephen Darwall, Allan Gibbard, and Peter Railton (New York, Oxford University Press, 1997), 137-163.

- _____. (1989) "Naturalism and Prescriptivity," in *Social Philosophy and Policy* 7, 151-74.
- _____. (1992) "Some Questions about the Justification of Morality," in *Philosophical Perspectives* 6, 27-53.
- _____. (1993a) "Noncognitivism about Rationality: Benefits, Costs, and an Alternative," in *Philosophical Issues*, no.4, 37-51.
- _____. (1993b) "What the Non-Cognitivist Helps Us To See The Naturalist Must Help Us To Explain," in *Reality, Representation and Projection*, edited by John Haldane and Crispin Wright (New York: Oxford University Press), 279-300.
- _____. (1993c) "Reply to David Wiggins," in *ibid.*, 315-328.
- _____. (1995a) "Subject-ive and Objective," in *Ratio* 7, no.3, 259-76.
- _____. (1995b) "Made in the Shade: Moral Compatibilism and the Aims of Moral Theory," in *On the Relevance of Metaethics*, edited by Jocelyne Couture and Kai Nielsen (Calgary: University Calgary Press), 79-106.
- _____. (1996) "In Search of Nonsubjective Reasons," in *Reason, Ethics and Society*, edited by Kurt Baier and J. B. Schneewind, 117-143.
- _____. (1997) "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Nigel Gaut, 53-79.
- _____. (1998a) "Aesthetic Value, Moral Value, and the Ambitions of Naturalism," in *Aesthetics and Ethics: Essays at the Intersection*, edited by Jerrold Levinson, 59-105.

- _____. (1998b) "Moral Explanation and Moral Objectivity," in *Philosophy and Phenomenological Research* 58, 175-183.
- _____. (1999) "Normative Force and Normative Freedom: Hume and Kant, But not Hume Versus Kant," in *Ratio* 12, 320-353.
- Ramsey, Frank P. (1926) "Truth and Probability," in *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, edited by D. H. Mellor (London, Henley: Routledge and Kegan Paul, 1931), 58-100.
- Rawls, John. (1971) *A Theory of Justice* (Cambridge, Mass.: Harvard University Press).
- Rosati, Connie S. (1995) "Naturalism, Normativity, and the Open Question Argument," in *Nous* 29, 46-70.
- Strugeon, Nicholas. (1985) "Moral Explanations," in *Morality, Reason and Truth*, edited by David Copp and David Zimmerman (Totowa, N.J.: Rowman and Littlefield, 1985), 49-78.
- Wiggins, David. (1993) "Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton," in *Reality, Representation and Projection*, edited by John Haldane and Crispin Wright (New York: Oxford University Press), 301-313.
- _____. (1993) "A neglected Position?" in *ibid.*, 329-336.
- Williams, Bernard. (1985) *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press).