

Spring 2006

## **Immanuel Kant & Bernard Williams: Discovering the Foundations of Morality**

Jonathan Comish

Follow this and additional works at: [https://digitalcommons.lsu.edu/honors\\_etd](https://digitalcommons.lsu.edu/honors_etd)



Part of the [Social and Behavioral Sciences Commons](#)

---

*Immanuel Kant & Bernard Williams:  
Discovering the Foundations of Morality*

Honors Thesis  
Jonathan Comish  
Spring 2006  
Thesis Director:  
Husain Sarkar

# Table of Contents

<b>Introduction</b>	<b>1</b>
I. The Issue	1
II. The Laws of Freedom	2
III. A Zero Sum Game	5

## **Section I:** **The Ethics of Kant**

<b>Chapter 1. Why Must Freedom be Transcendental?</b>	<b>11</b>
I. The Tyranny of the Laws of Nature	13
II. Can there be Free Will?	15
<b>Chapter 2. <i>Groundwork of the Metaphysics of Morals</i></b>	<b>20</b>
I. Hypothetical Imperatives	21
II. The Categorical Imperative	26
III. The Four Examples	33

## **Section II:** **The Ethics of Bernard Williams**

<b>Chapter 3. "Internal and External Reasons"</b>	<b>42</b>
I. Internal Reasons	43
II. External Reasons	52
III. The Problem With Irrationality	57
<b>Chapter 4. <i>Ethics and the Limits of Philosophy</i></b>	<b>61</b>
I. Williams' Critiques of Reason	62
II. Can there be <i>Objective</i> Practical Deliberation?	73
III. Does Kant Destroy Integrity?	79
IV. The Lines Are Drawn	83
<b>Chapter 5. The Constructive Ethical Theory</b>	<b>85</b>
I. A Critique of Obligation	88
II. The Composition of Practical Deliberation	98
III. Practical Necessity	106
IV. Williams' Ethics	111

**Section III:**  
**Reponses to Williams**

<b>Chapter 6. The Critics</b>	<b>117</b>
I. Hollis: The <i>sub-Kantian</i> Approach	118
II. Scanlon: A Matter of Preference	128
III. Korsgaard: The Internalism Requirement	136
IV. Allison: The Existence of Universal Principles	145
 <b>Chapter 7. A Kantian Reply</b>	 <b>149</b>
I. Kant's Rational Agency	150
II. Pure Practical Deliberation	152
III. Kant and Integrity	158
IV. The Locus of Value	163
V. In Defense of Morality	166
 <b>Works Cited</b>	 <b>171</b>

## Introduction

Kant's *Groundwork for the Metaphysics of Morals*, written in 1785, is one of the rare philosophical texts which radically changes the state of discourse. Speaking solely in terms of influence, very few treatises are able to be compared to Kant's moral masterwork. This surprising amount of influence is due to the fact that Kant was able to develop a new moral system, which broke substantially from the theories that preceded it. When we examine the philosophy of Bernard Williams, we find another philosopher who seeks to break from the prevailing philosophical traditions of his time. In his book, *Ethics and the Limits of Philosophy*, published in 1985, Williams explicitly sought to take ethics in a radically new direction. Of course, in doing so, he will come into conflict with Kant's philosophy, and this conflict will be the central concern of my thesis.

### I. The Issue



From the time of its inception to the present day, Kant's moral philosophy has not lacked for critics. It is to Bernard Williams, one of the more influential of these critics that this thesis aims to respond. Williams, perhaps the most prominent of Kant's critics of the last thirty years or so, wishes to do away with not only Kant's moral philosophy, but also with the entire branch of ethics that he labels "morality." The most prominent reason that Williams cites in his quest to remove morality is that morality limits, perhaps even destroys, the possibility of our living a fulfilled life in which we are true to ourselves. His claim is that morality, in placing absolute demands upon us (in the case of utilitarianism, always having to strive for the greatest utility), requires us to

relinquish certain fundamental desires or projects that we have undertaken so that we might fulfill our never-ending moral obligations. Williams' contention is that, ultimately, morality does not allow me to be *me*. Of course, the Kantian would argue just the opposite, since only by acting for the sake of the moral law can I be autonomous and free from the tyranny of the laws of nature. There is a reason that Kant terms moral laws the "laws of freedom," perhaps best demonstrated by an analogy.

## II. The Laws of Freedom



When I first started to become a proficient musician, I felt that the beauty of music was entirely contained in the act of creation and the absolute spontaneity thereby expressed. I thought that if one were to place limits on this act of creation by, for example, playing a different chord because that is what the rules of musical theory demanded, then the beauty of music would somehow be threatened. As a result, I neglected to study musical theory. Instead, I contented myself with the basic chords and scales that I already knew, thinking that I could easily augment them to suit what I was playing at the moment. Much to my surprise, I found that I became a much less creative musician. I would hear other musicians improvise beautifully, playing much more complicated scales than I knew, and knowing exactly when to play the extra notes in these scales. When I would try to emulate them, I inevitably became lost and what I produced could only be called, very charitably, creative music.

After struggling with this frustration for a year or two, I finally broke down and enrolled in a music theory class. To say that my eyes were opened would not do the experience justice. I soon learned exactly how to use the diminished fifth to heighten the emotional content of my music. I could use sixth chords to provide smooth transitions between soothing major sevenths and staccato ninths. I could all of a sudden play modally, changing the scale that I was using to better compliment the chord over which I was playing. In short, instead of telling me what I could not play, as I feared they would, the rules of music theory told me what I *could* play, thereby dramatically increasing the range of my playing.

Before learning music theory, my playing was never as free or spontaneous as I had imagined it. Instead it was ruled by hesitancy, boredom and insecurity. I always wanted to play just the right note to express the aesthetic that I was imagining at the moment, but I was never sure of which notes I could play. I was worried that if I learned too much about music theory, then my playing would be determined; it would be so limited that I would sound just like all of the other musicians whom I was resolved not to resemble. Instead what I found was a dramatic expansion both of the options available to me and of the confidence with which I could play those notes and chords.

This analogy applies fairly well to the laws of freedom. Williams argues that Kant's moral law reduces agents to mere automatons, since those who wish to follow the Categorical Imperative must act based on pure practical reason; they must reason the same as anyone else would in the same circumstances. Given a certain reading of Kant, it is easy to see how one could come to such a conclusion. I am always under some obligation imparted by the Categorical Imperative, and it seems that, if I wish to follow Kant's theory, then I will have no time left to be myself; just as I used to think that if I

were to follow the rules of music theory, I would somehow not be able to create authentic music. However, just as in the case of music, this is not necessarily true.

While it is true that the Categorical Imperative commands categorically, it often allows us some leeway in how we act, such as when we are performing an imperfect duty – one that is concerned with furthering some agent's ends, such as her happiness. We are allowed some room in how we decide to further those ends. Moreover, how we pursue our own ends (for this is enjoined by the Categorical Imperative as well) depends upon what my ends are. Kant's moral law commands only that I relinquish those ends that are not in accord with its strictures – ends that are immoral anyway. Certainly I lose nothing by not pursuing immoral ends, and if I am thereby forced to reevaluate and change myself, so much the better.

The result of the process of deliberation enjoined by Kant is that I know that I am able to act morally when I do act. I am able to act with confidence after I consult the Categorical Imperative, there is little doubt, if I have deliberated properly, that I am doing the morally permissible thing.

In contrast, Williams leaves us with only the notion that I should act how *I* want to act. In his eyes, ethical considerations have only as much weight in my practical deliberations as I want them to have. This means that if an agent was deliberating about a certain course of action, any inclination to behave ethically would not necessarily have any decisive weight in his deliberations. The agent would have no point of reference to which he could refer for guidance except his own desires. In Kant's eyes such an agent would be trapped, in thrall to the laws of nature; he would have no autonomy.



### III. A Zero Sum Game



What explains the radical gulf between these two brilliant minds? There can be little doubt that their contest is a zero-sum game; each proposes a theory that is utter anathema to the other. As such, it is imperative to resolve the issues that have given rise to this standoff. To find the ultimate source of this chasm, we must explore within the philosophy of each, and examine the foundational claims at the very heart their philosophies. These claims, the metaphysics of Kant and the psychology of Williams, give rise to their respective ethical theories. It is only when the conflict between the roots of these two competing philosophies has been resolved that we will be able to accurately assess the resulting ethical theories.

Although some might have an immediate reaction of alarm at the apparently subjective doctrine of Williams or the seemingly impersonal theory of Kant, it is important to recognize that each of these theories was carefully constructed as a result of the considerations arising from the deeper philosophies of their respective creators. As such, it is important to reserve judgment until these issues have been resolved. If Kant is correct in his metaphysics, then his ethics could have turned out no other way. Likewise, if Williams is correct in his psychological account, then we have no choice but to reject Kant's theory in favor of that of Williams. Therefore, a significant part of this thesis will be devoted to examining the roots of these theories, and how the roots of the theories gave rise to the fully developed account. Since Williams is a much later philosopher than Kant, and his theory arose partly in reaction to Kant's philosophy, his criticisms of Kant will also be carefully analyzed in their turn. Once we have achieved a

clear understanding of how these theories were developed, we will be in a position to analyze them critically.

Considering all of this, it seems most appropriate to order this thesis chronologically. We will begin with an examination of some passages from *The Critique of Pure Reason* (1787) which raise issues critical to the development of Kant's moral theory. Next, an examination of *The Groundwork of the Metaphysics of Morals* will give us the essential elements of that moral theory. Once we have a full understanding of the workings of Kant's moral account, we will be in a position to analyze how this contrasts with Williams' philosophy. Accordingly, we will then begin with an examination of Williams' seminal paper, "Internal and External Reasons." I intend to argue that this paper serves the same role in relation to Williams' mature ethical theory that the positions advanced in *The Critique of Pure Reason* play in relation to Kant's fully developed ethical theory. The next step, once we have culled the central arguments of this paper is to examine Williams' fully developed theory, as presented in *Ethics and the Limits of Philosophy*. Finally, we will examine some of the more prominent critics who have responded to Williams, so that we might determine whether any of their arguments give us the resolution that we are seeking. These critics include: Martin Hollis, who argues that Williams' sub-Humean account of practical deliberation is inherently problematic, Timothy Scanlon, who argues that, while he does not believe in the sub-Humean account, he also does not believe that it excludes any moral theory, Kristine Korsgaard, who argues that while Williams was essentially correct with his internalism requirement, he was mistaken in concluding that this requirement excludes the Categorical Imperative, and finally Henry Allison, who takes Korsgaard's argument a step further and argues that Kant included mechanisms in his theory by which it can

account for the internalism requirement. This thesis will conclude, after the exegetical work has been completed, with a critical analysis of the central issues, in order to see if a resolution is possible.

Of course, if a resolution of the central problem is impossible, then we are no further on. I intend to argue that Williams was, in fact, mistaken in his reading of Kant, and to show in fair detail how his arguments in “Internal and External Reasons” relate to the Categorical Imperative. I intend to argue that I believe that, once we take this perspective, then the issues will become much more tractable and a solution will be within reach – namely that the Categorical Imperative not only could be, but is an internal reason. Enough has been said about the aims of this thesis in general. Without solid philosophy, these words will turn out to be just so much fluff. I am anxious that we should begin the serious work of the task that I have set for us. Therefore, let us begin.



# **Section I**

## **The Ethics of Kant**

# **CHAPTER 1**

## *Why Must Freedom Be Transcendental?*

If one wishes to understand Kant's moral philosophy, then the best place to begin would be a close examination of Kant's metaphysics, as laid out in his work, *The Critique of Pure Reason*. In this monumental book, Kant develops his metaphysical theory, one that has particular consequences for his ethics. In fact, these consequences are so immediate and direct, that it would be very difficult to understand why he shapes his moral theory the way that he does without a certain degree of familiarity with his metaphysics. Therefore, because this thesis is concerned with a defense of Kant's moral theory, it will begin with an examination of his metaphysics or, at least, that part of his metaphysics which has important consequences for his ethics.

## **I. The Tyranny of the Laws of Nature**



Kant wrote his philosophy during the time in which Newton's theory of physics was in its ascendancy. Considering this, it should come as no surprise that Kant's metaphysical theory is informed a great deal by Newtonian physics. According to Newtonian physics (at least as it is presented by Kant), the universe is governed by the Laws of Nature, and within the universe, the rule of these laws is absolute. (*Critique of Pure Reason*, A 444/ B 472.)<sup>1</sup> The classic example of this is that of the apple falling and striking Newton. Once the apple's connection with the tree was severed, it had no choice but to fall, for it had to obey the law of gravity, itself one of the Laws of Nature, and the path of its descent was determined by others of the Laws of Nature; to such an extent that its trajectory was entirely predictable. If we were to attempt to recreate the incident, as long as we kept the initial conditions and the Laws of Nature the same, the

---

<sup>1</sup> The *Critique of Pure Reason* will hereafter be referred to as *CPR*

experiment would have the same result, no matter how many times we repeated it. This is but an example of a microcosm within the universe, but it nevertheless serves as a suitable demonstration of the implacable force of these laws.

That the physical universe is governed by the Laws of Nature does not seem to have any direct relevance to an ethical theory. But keep in mind that we – humans - are also part of the physical universe. What this means is that we are also subject to the Laws of Nature, not only physically, but psychologically as well. Though this is not immediately apparent, since we do not normally regard our psychological particulars as things with physical states, and we usually hold that the Laws of Nature govern only those things of a physical nature, Kant explains the role of the Laws of Nature in the formation of our psyche through his theory of causation.

Kant gives a full explanation of his theory of causation in the section of *The Critique of Pure Reason* titled “The Antinomy of Pure Reason; Third Conflict of Transcendental Ideas.” (CPR, A 444/ B 472) The transcendental ideas which are in conflict are the ideas that there is transcendental freedom and that there is no transcendental freedom, but this does not concern us just yet. What is of immediate concern is the theory of causation that Kant describes in this section, and the implications that that holds for any ethical theory. Kant defines causality according to Natural Law, saying that, “everything **that happens** presupposes a previous state, upon which it follows without exception according to a rule.” (CPR, A 444/ B 472) When we are considering his theory of causation in relation to psychological states, we must pay particular attention to the phrase “everything **that happens**,” for this would necessarily include our psychological states in the set of things which are under the dominion of Natural Law.



The idea that Natural Law holds sway even over our psyches can easily be shown through examples. Suppose that an agent, call him A, is feeling joyful at one particular moment. It would be unlikely that the feeling of joy that A is experiencing happened of its own accord; our emotions do not work in that fashion.<sup>2</sup> Rather, to say that something caused his feeling of joy would make much more sense. This is why our emotions cause us humans so much consternation, for we have little, if any, control over them and we often find that the emotions that we are feeling at the moment are not the ones that we want to be feeling. To return to agent A, perhaps he just received a raise at work, or found out that his wife was pregnant, or a host of other occasions which normally cause a feeling of joy in people. It seems now that, instead of causing the joy himself, the joy was caused in A by something outside of himself, namely some event that occurred.

We have just given the proximate cause for A's joyfulness, but there are still some remaining questions which need to be answered in order to have an adequate understanding of how the Natural Laws work in determining the state of our minds. For instance, it has not yet been made clear why a certain event would cause A to be joyful. It is unlikely that A decided that getting a raise or finding out that he was about to be a father should be a cause for joyfulness; that is not how our emotions work. As will be argued presently, through our reason, which is outside of the influence of the Laws of Nature, we do have control over how we react to our emotions and how we allow them to influence our behavior, but we do not have control over which emotions we feel. Therefore, there must be something embedded deeper in A's psyche which determined

---

<sup>2</sup> One might object by saying that A just has a happy disposition, but this only reinforces my point – in this case, A is again happy because of something determined by the Laws of Nature.

what his reaction to the event would be. If we wished to pursue the enquiry further, we would have to investigate the other circumstances of his life, and determine what effect those circumstances have in shaping his psyche. However, the purpose of this example is not to investigate A's psychological makeup; instead, it is to demonstrate how our psyches are determined by the circumstances which surround us. What we should take away from this example is that the Laws of Nature, which govern *all* causal interactions, also determine the causal interactions that shape our psychological makeup, and so, by inference, determine us as well.

In theory, then, if we were all completely determined by the Laws of Nature, a psychologist would be able to predict my actions, if he was aware of all of the particular circumstances which determined my psychological makeup, just as surely as a physicist could predict the trajectory of the apple which fell to strike Newton. In fact, some psychologists purport to do just that. However, allowing the Laws of Nature sole dominion over this universe would not be without some serious consequences. The obvious consequence would be that the entire universe, and everything in it, would have been completely predetermined, from alpha to omega. Every human accomplishment, from the poetry of Ovid to the sculpture of Rodin would become meaningless as a feat of humanity, since it would have been determined from the dawn of time that Ovid would write what he wrote, and Rodin would sculpt what he sculpted.

Outside of the other considerations that a completely predetermined universe would have for how we think about ourselves and our history, a completely predetermined universe would be fatal to any ethical enterprise, and for many reasons. After all, we would have no control over our actions, and without choice or volition,

there could be no such thing as moral responsibility.<sup>3</sup> How can I be held accountable for an action over which I had no control, or possibility of influence? It is quite difficult, if not impossible, to formulate an ethical theory without some notion of responsibility. This consideration recalls the famous assertion that, "Ought implies can," meaning that I cannot be obligated to do something that I am not able to do. Since most ethical theories are centered around the idea of obligations, and since there could be no such thing as an obligation in a completely predetermined universe, there could be no such thing as ethics in a predetermined universe.<sup>4</sup>

## II. Can There Be Free Will?



The task that Kant has taken upon himself now seems to be a daunting one. His task is none other than to rescue human freedom from the tyranny of the Laws of Nature, a task which is as difficult as it is important. Where is Kant to begin his attempt? He cannot try to posit human freedom within the universe with which we are familiar, the universe of space-time; for he has already established that the Laws of Nature enjoy an uncontested dominance over that realm. Considering this, Kant makes a bold move and posits the location of human freedom as being *outside* the universe of space-time. (CPR, A 446/ B 474) By doing so, Kant has allowed room for "an **absolute** causal **spontaneity** beginning **from itself**." (CPR, A 446/ B 474) If Kant is successful

---

<sup>3</sup> As will be seen later, Bernard Williams will see no problem with there being no such thing as morality, if not ethics, and he denies that there is anything resembling what we would call moral obligation.

<sup>4</sup> Of course, one could argue that one could be obligated to do those things that one was fated to do in a predetermined universe, thus fulfilling the *ought implies can* formulation, but that consideration would not rescue the concept of obligation from incoherence in such a universe, because people in that universe would still have no choice but to act as they were fated to act.

in establishing that causal spontaneity is possible, and that we are able to be the sources of causal spontaneity, then, through this transcendental freedom, there is hope for ethics yet.

The method that Kant chooses in his attempt at proving the possibility of free will is a *reductio ad absurdum*. Thus he commences by assuming that there is nothing other than the Laws of Nature. This means that any given state of affairs implies a previous state of affairs which determined the current state of affairs. Of course the previous state of affairs must have been determined by some even earlier state of affairs, and so on *ad infinitum*. The assumption that there is no source of causality in the universe other than the Laws of Nature has led Kant to the consideration that, “at every time there is only a subordinate but never a first beginning, and thus no completeness of the series...” (CPR, A 446/ B 474) Therefore, if there was no other source of causation in the universe, other than the Laws of Nature, then it would have been impossible for the universe to have come into being in the first place, as there would always have been some antecedent cause determining every event. Kant claims that this leads into a contradiction, by saying that “now the law of nature consists just in this, that nothing happens without a cause sufficiently determined *a priori*. Thus the proposition that all causality is possible only in accordance with the laws of nature... contradicts itself” (CPR, A 446/ B 474) This somewhat cryptic passage is absolutely critical, for it is that contradiction which allows Kant to posit some source of causality other than the Laws of Nature.

This contradiction is among the most difficult passage in Kant’s philosophy, and for an adequate understanding of it, we would do well to consult Henry Allison. In his

book, *Kant's Theory of Freedom*<sup>5</sup>, Allison argues that this apparent contradiction rests in this: that if there were an infinite regress, then the universe would be incomprehensible, because it would break the principle of sufficient reason inherent in the Laws of Nature. To understand this, we must remember one of the conditions Kant sets out in the antinomy, that “nothing happens without a cause sufficiently determined *a priori*,” where Allison takes “sufficiently determined” to mean that all of the antecedents of a particular effect are in place. As such, it would be impossible for a cause to be “sufficiently determined” if there were an infinite regress, because there would be no ultimate cause, and therefore no point at which one could say that a cause had been sufficiently determined. Therefore, there must be some form of causality other than the laws of nature, or the laws of nature would lead to an inescapable contradiction, in that they require every state of affairs to have sufficiently determined causes, but no such thing would be possible.<sup>6</sup>

So, Kant has now established that there is room for transcendental freedom, but he has yet to tell us how it is possible for *humans* to partake of this freedom. It is one thing to say that the creator of the universe has transcendental freedom, and quite another to say that humans, whose physical bodies and minds are located in the universe of space time, are able to make use of transcendental freedom. This question has two dimensions, a theoretical and a practical one. The theoretical side of this question concerns our metaphysical ability to exist, in a manner of speaking, in both the noumenal and phenomenal realms, for Kant posits our reason (as I will discuss presently) within the noumenal realm and our bodies and minds in the phenomenal

---

<sup>5</sup> Hereafter *KTF*

<sup>6</sup> Allison, *KTF*, pp.16-19

realm. In addition to this consideration of a kind of metaphysical duality, there is a deeper question to be answered, namely a thing existing in the noumenal realm - our reason - is able to affect something in the phenomenal realm.

That Kant believes in the ability of things in the noumenal realm to affect the phenomenal realm is not in doubt, as is illustrated by the following passage:

But because the faculty of beginning a series in time entirely on its own is thereby proved (though no insight into it is achieved), now we are permitted also to allow that in the course of the world different series may begin on their own as far as their causality is concerned, and to ascribe to the substances in those series the faculty of acting from freedom. (*CPR*, A 450/ B 478)

The key part of this quote in understanding Kant's metaphysical position is actually the parenthetical remark, "though no insight into it is achieved." By making this remark, Kant is not saying that there is more work to be done in order to understand this metaphysical conundrum, he is saying that to prepare us for his ultimate conclusion: namely that no insight *can* be achieved.

For an explanation of this contention, we may turn to the third section of *The Groundwork of the Metaphysics of Morals*. Though this section is ostensibly about the incomprehensibility of the absolute necessity of the moral law, Kant defends his position through certain aspects of his metaphysics which are quite helpful in explaining his position regarding our inability to comprehend the connection of our noumenal selves to our phenomenal selves.

In attempting to understand Kant's position, the following is quite indispensable, "For we can explain nothing but what we can reduce to laws the object of which can be

given in some possible experience.” (G, 4:459) This means that, although we might be able to *prove* that there is some way for objects in the noumenal world to influence the phenomenal world, as Kant did through his *reductio*, there is no way for us to *understand* what Kant has proven. We can have no experience of the world of noumena, and therefore there is no way for us to abstract the laws governing the workings of the noumenal world, since we must derive these laws from sense experience and we can have sense experience only of the phenomenal world. Now, in terms of understanding how objects of the noumenal world influence the phenomenal world, it should be apparent that we are unable to understand these interactions, because they exist, at least partially, in the noumenal world, and we are unable to understand the workings of that world. (G 4:452)

Having disposed of the theoretical aspect of the question of how humans are able to partake of transcendental freedom, it is time to consider its practical aspect. Unfortunately, there is no brief way to answer that side of the question, as it consists of the entirety of Kant’s ethical theory, for Kant contends that only by acting for the sake of the moral law, and not merely in conformity with it, are we able to be free from the oppression of the laws of nature, and his arguments to that end make up *The Groundwork of the Metaphysics of Morals*. Therefore, since this thesis is concerned with a detailed defense of Kant’s ethical theory, it seems appropriate to conduct a detailed examination of the crucial aspects of that theory.

## **CHAPTER 2**

### *Groundwork of the Metaphysics of Morals*



In explaining Kant's theory, one of the toughest challenges is deciding where to start. This is a natural consequence of the theory being such a well-integrated whole, since every aspect of the theory supports and is supported by other aspects of the theory. It would be best to explain the whole *Groundwork* at one time, but that is obviously impossible. Fortunately, we have already discussed some of the metaphysical considerations that influence the groundwork, and there is a specific purpose behind this exegesis, namely to defend Kant from the critiques of Williams, so these considerations suggest to me that a suitable place to start would be an examination of the Categorical Imperative, with a discussion of Kant's moral psychology, and finally a presentation of the four examples. In this way, we are able to begin by focusing on that aspect of Kant's moral theory which so galled Williams, an aspect which we are able to cogently discuss due to the preceding section on transcendental freedom, an understanding of which is necessary in order to comprehend why Kant shaped the Categorical Imperative the way that he did.

## **I. Hypothetical Imperatives**



So that we leave no significant part of Kant's theory unexamined, let us begin by determining why morality needs to be based upon a categorical imperative rather than a hypothetical one, or a set thereof. This is a crucial distinction to understand, because, as

we shall see later, the constructive theory that Williams proposes is one based entirely on hypothetical imperatives. The best way to go about this would be to first draw a clear distinction between the two and closely examine the different qualities of each, so that we may be perfectly clear on the reasons why Kant thinks that hypothetical imperatives are unsuitable as a basis for morality, while a categorical imperative is the perfect fit.

In order to properly understand just what a hypothetical imperative is, and how it differs from a categorical imperative, we must first be sure to understand just what an imperative is. Kant defines an imperative as, "The representation of an objective principle, insofar as it is necessitating for the will, is called a command (of reason), and the formula for the command is called an **imperative**." (G, 4:413) So, an imperative is an expression, or formulation, of some obligation.

Now, we can more easily examine just what, exactly, a hypothetical imperative is. Simply put, it is an imperative with a specific end, meaning that a hypothetical imperative is only true, or obligating, in relation to the end for which it was formulated. This is illustrated by the following: "The former [hypothetical imperatives] represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will)." (G, 4:414) For example, right now, I am typing this chapter under the hypothetical imperative that, if I want to write an excellent thesis, then I must work diligently. However, if I did not care write an excellent thesis, then I would not be obliged by that specific imperative, since it was formulated based on a certain end, and that end no longer interested me.

Kant recognizes two types of hypothetical imperatives, those that are *possible*, and those, or, more specifically, the *one* that is *actual*. (G, 4:415) The possible hypothetical imperatives are those that possess a possibility of obtaining, and are, in

effect, infinite. Any and every conceivable end, therefore, has attached to it at least one possible hypothetical imperative as the way of obtaining that end. The preceding example, of me wanting to write an excellent thesis, is a possible hypothetical imperative. It is an end which someone is able to possess.

Against this must be contrasted the actual imperative. This is different from the possible hypothetical imperatives because this imperative is oriented to an end that all rational beings have specifically that of happiness. Kant goes so far as to say that the pursuit of happiness is an intrinsic part of man's nature. (G, 4:415) However, despite this imperative being present in every man, it is still unsuitable as the basis for morality, and Kant gives three reasons for this.

The first reason is that, despite its universality, the imperative of happiness is still a hypothetical imperative. Even though it is present in all men, it does not take the same form in all men. It is determined both by the psychological makeup of each person, for everyone has a different path to happiness, and the individual's skill in pursuing that path to happiness, what Kant calls the individual's "*prudence*." (G, 4:416) So, the imperative of prudence (as Kant names it) is universal only in its *presence* in all men and not in its *content*, as that is determined by the particular path that will lead to an agent's happiness, and in this respect cannot serve as the basis of an objective moral system.

The second reason that Kant has for rejecting the imperative of prudence as the basis for the moral law is that he does not wish to have a moral law that is determined by anything other than reason. He explicitly states that he wishes his moral law to be divorced from "everything that may be only empirical and that belongs to anthropology." (G, 4:389) He wants a moral system for all rational beings, not just

humans, and consequently must reject the influence of anything which might be limited to humans, such as the pursuit of happiness.

The final reason that Kant gives for not basing the moral law on the imperative of prudence is that “the concept of happiness is such an indeterminate concept that, although every human being wishes to attain this, he can still never say determinately and consistently what he really wishes and wills.” (G, 4:418) What use would there be of having a moral law that is, at best, indeterminate in its aims? If the agent who is pursuing his happiness cannot even say exactly what it is that he is pursuing, whether what he is pursuing is worthwhile, nor is he able to determinately say how he intends on attaining it, then it would scarcely be prudent to place that ephemeral object of his desire as the basis of something as fundamentally important as the moral law.

Therefore, hypothetical imperatives are present throughout our everyday lives and we use them constantly, albeit sometimes unknowingly. Why, then, can we not base our morality on them?<sup>7</sup> Kant gives a variety of reasons for asserting their unsuitability, but I will focus here on just two of them. The first reason is that they are formed by our desires and inclinations. They would form heteronomous principles of morality, something that Kant expressly forbids; saying that heteronomy of the will is “the source of all spurious principles of morality.” (G, 4:441) The second reason is that they would have no real normative force. This is because they are formulated for the purpose of attaining some end. Therefore, if the end sought by a particular imperative does not entice me, then I am able to say that that imperative holds no obligation for me. Kant feels that an imperative that holds no real normative force is an absurd candidate for being the moral law, and says so explicitly in the following passage:

---

<sup>7</sup> I have already mentioned this issue in passing, but it is quite critical and deserves further discussion.

Everyone must grant that a law, if it is to hold morally, that is, as a ground of an obligation, must carry with it absolute necessity; that, for example, the command “thou shalt not lie” does not hold only for human beings, as if other rational beings did not have to heed it, and so with all other moral laws properly so called; that, therefore, the ground of obligation here must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in the concepts of pure reason; and that any other precept, which is based on principles of mere experience - even if it is universal in a certain respect - insofar as it rests in the least part on empirical grounds, perhaps only in terms of a motive, can indeed be called a practical rule but never a moral law. (G, 4:389)

Thus, if something purports to be the moral law, then it must command absolutely, and without exception. Now we must determine why hypothetical imperatives are so unsuited for the task.

Let us first examine why all hypothetical imperatives must be heteronomously given. It has already been asserted that hypothetical imperatives are formed for the pursuit of some purpose. They all take the basic form of if you want/need/desire A, then do B. The problem lies in the fact that all of our wants, needs, and desires are based in the phenomenal realm and are formed by the laws of nature, not by our reason. The role that reason plays is in evaluating hypothetical imperatives and determining whether or not they are well suited for attaining the purpose for which they were formed, and whether or not their ends are in accord with the moral law. Therefore, hypothetical imperatives *cannot* be the basis for morality, since they are ultimately the products of the Laws of Nature, and if we were to base morality upon them, we would leave no room for human freedom or autonomy, since we would then be unable to escape the tyranny of the Laws of Nature.

The second reason why hypothetical imperatives are so unsuitable is that their normative force is easily denied. Kant illustrates this point when he says:

(hypothetical imperatives) can indeed be called *principles* of the will but not laws, since what it is necessary to do merely for a discretionary purpose can be regarded as in itself contingent and we can always be released from the precept if we give up the purpose... (G, 4:420)

Because they are predicated on an agent wishing to attain certain ends, if someone was not inclined to pursue those ends which are enjoined by the hypothetical imperative, then there would be no way for the imperative to assert any normative force upon him. What would be the purpose of having an ethical system, whose normative force could be easily escaped, simply by not having a particular end? Such an ethical system would surely be meaningless. Therefore, based on these two considerations, we are able to dismiss any system of ethics based upon hypothetical imperatives.

## II. The Categorical Imperative



The categorical imperative, however, is something which is altogether different. First of all, the categorical imperative, and this is what makes it categorical, has no end formed from our *desires and inclinations*. As Kant describes it, "The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end." (G, 4:414) That is why it forces itself upon us at all

times. Crucially, that is also why a categorical imperative would allow us to be autonomous agents, because it has no basis in our desires and inclinations and is not under the sway of the Laws of Nature, and hence must be without the universe of space-time, residing within the noumenal world. There is no method by which we could deny it as we can hypothetical imperatives, for the way to deny them is by denying that we have a desire to attain their ends, but the categorical imperative does not command in relation to some desired for end.

Now, the difficult part is, if this has no end, then how is it an imperative? Kant's answer is this that "It has to do not with the matter of the action and what is to result from it, but with the form and the principle from which the action itself follows..." (G, 4:416) Therefore, this imperative concerns itself with how we formulate our actions, and hence is a *moral imperative*. One might make the point that because this imperative has as its end morality, it really is only a hypothetical end, since someone can simply deny that he wants to live a moral life. Kant's answer would be that it is not possible to deny this, since the highest purpose of reason is in the production of a good will, and that is analogous to being moral, then if someone were to deny morality, then that would be to live in discord with reason, which could not be a possible end of a rational agent. (G, 4:396)

The importance of the contrast between the categorical imperative and hypothetical imperatives should now be clear. Hypothetical imperatives have their function and their purpose, but their purpose is at all times always hypothetical, and therefore they are not suited to be the basis for morality, since, because of the fact that they, the hypothetical imperatives, can always be denied or ignored, they could therefore never be respected nor give rise to duty that is a necessity for a moral law. In contrast,

the categorical imperative could never be denied, and is a suitable legislator of our morals. Due to this consideration, Kant asserts “that the categorical imperative alone has the tenor of a practical **law**.” (G, 4:420)

Now that we have established that the imperative on which morality must be based is a categorical imperative, we must examine why Kant claims that the Categorical Imperative takes the form (or forms) that it does. In the argumentation culminating with the Categorical Imperative, Kant introduces three considerations which will shape the ultimate form of the Categorical Imperative. Of these, an explanation of his second and third propositions in Section One of the Groundwork would be the most productive way of explaining why Kant formulates the Categorical Imperative the way that he does. It is in these two propositions that he explains how duty attains moral worth and just what duty is, two vastly important concepts in his moral philosophy.

His second proposition can be more easily analyzed in two parts, the first being, “an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon.” (G, 4:399) There are two words here which deserve special consideration, “duty” and “maxim.” Since duty is the focus of the third proposition, any extended discussion of it would be more appropriately located within the exegesis of the third proposition. Maxim, however, is a term that needs to be explained here, in this context. Kant’s definition of maxim is that it is “the subjective principle of volition,” meaning that it is the principle that we use to guide a certain action. (G, 4:401, footnote) For example, when a diabetic takes his insulin injections, he is not doing so arbitrarily. Rather he has reasons that guide his actions, and these reasons can be expressed in the form of a maxim. In this specific case, the maxim would be something like, “I ought to perform those actions that are



conducive to my good health.” So there are many maxims that we use in contemplating our actions, hence the subjective part of their nature.

How does this lead to the rest of the first half of the second proposition? How does the maxim lend an action its moral worth? To understand this concept, we must not view maxims as the ultimate sources of moral worth, but rather as the conduits through which moral worth passes from the moral law to our actions. That moral worth comes from the moral law should be clear, once we see that maxims, or at least those of moral worth, are shaped to be in accordance with the moral law. Considering this, it becomes easy to recognize how the moral worth of our actions comes to them through the maxim in accordance with which they were performed.

The preceding two paragraphs explain the function of maxims, but they did not explain how the purpose to be attained by an action has no relevance to its moral worth. The explanation of this is rather straightforward, for morality is a thing that must be pure. It must not be subject to anything else, specifically anything of this world, for if it were, it would not have its requisite normative strength. What reason would we have for being moral if our moral status was outside of our control? Therefore, moral worth cannot be attached to the outcomes of our actions, but must rest in our *intent*, something that is completely within our control, therefore leaving us in control of our own moral worth, and not subjecting morality to any outside influences. (G, 4:400)

The second half of the second proposition is, “but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire.” (G, 4:400) Now that the first half has been examined, it is apparent that the second half adds little that is critical, but is more of an emphatic restatement of the first half. However, that is not to say that there is not anything of importance in it.

It seems that the most important part of it is where Kant says that moral worth comes divorced from objects of desire. This is a very pertinent point, meaning that, for an action to have moral worth, it must be done *solely* with regards to the moral law. If I were to feed a starving child out of sympathy for her plight rather than from my moral duty, that would *not* be a moral action, for it would be guided by influences that have nothing to do with my will or reason, and hence to call that action moral would be, once again, placing morality in space-time, again placing our moral worth on things outside of our control, an action that would destroy morality.

Kant's third proposition is, "*duty is the necessity of an action from respect for the law*" (G, 4:400) To understand this statement, two questions need to be answered: first, what does Kant mean by respect, and second how does this give rise to duty. Since an answer to the first question is necessary in the answering of the second, it seems a natural place to start.

What, then, is respect? Fortunately, Kant gives us an explicit answer to this question, namely that respect is, "a feeling *self-wrought* by means of a rational concept."<sup>8</sup> (G, 4:401, footnote) The part of this statement that should be most worrying to us is Kant's use of the word "feeling." Was not morality supposed to be divorced from such things as feelings and emotions? In answering this question, the most important part of this statement for us to pay attention to is "*self-wrought*." This means that respect, though it is a feeling, is something qualitatively different from other inclinations. In fact, it is a feeling that is informed, and formed, by reason. Kant goes further to explain

---

<sup>8</sup> This footnote will play an absolutely critical role later in this thesis, when I attempt to respond to Williams on behalf of Kant.

that respect is “the *subordination* of my will to a law without the mediation of other influences of my sense.” (G, 4:401, footnote)

There are two things that should be drawn from this passage, the first of which being that the only things that I can respect are laws, or by extension, other people who represent these laws. Now, why should respect be reserved solely for laws? The answer is simply that there is nothing else to which we should subject our wills, nothing else to which *reason* would see fit to subject our wills, therefore only laws are deserving of this special feeling. The second is that to respect a law is to be in *subordination* to it. That means that respect is the recognition of a law as something that should be a governor of my actions, and it is only through this recognition and subordination that I can properly align myself with the law.

Now that we can see just what respect is, it should be readily apparent how respect gives rise to duty, for, in respecting the law, I have already subordinated my will to it, and therefore have taken upon myself the duty of acting in conformity with that law. So now duty, properly speaking, can only rise from respect for the law, for there is no other way to have an obligation that arises from reason; therefore, duty is the necessity of an action from respect for law.

Having dispensed with Kant’s antecedent arguments, we are prepared to receive the moral law. In its first iteration, Kant presents it thus:

*“I ought never to act except in such a way that I could also will that my maxim should become a universal law.”* (G, 4:402)

Kant reaches this formulation based not only upon the two preceding considerations, but also upon all of his arguments concerning what a moral law must be: the law must

be formulated such that it commands with an absolute necessity, based in pure reason, the law must be of such a character that an agent's reason would determine that his will should be subject to the law, and the law must be completely divorced from the agent's desires and inclinations. In Kant's own words, this means that "Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law." (G, 4:402)

However, Kant was not yet finished refining his presentation of the moral law at the point in which he wrote the first iteration. Eventually we are presented with the three formulations, which are, respectively, the Formula of Universal Law (FUL), the Formula of Humanity as an End in Itself (FHE), and the Formula of Autonomy (FA). They are as follows:

- 1) *Act only in accordance with that maxim through which you can at the same time will that it become a universal law. (G, 4:421)*
- 2) *So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means. (G, 4:429)*
- 3) *Act only so that the will could regard itself as at the same time giving universal law through its maxim. (G, 4:434)*

Kant's contention is that these formulations are but three facets of the same law; they are all derived from reason and all work towards exactly the same end. (G, 4:436) Together they are the driving force of Kant's moral philosophy. If we are able to act according to the precepts of these laws, then we are moral beings. Now our task is to determine how these laws may be applied to our actions, so that we can be clear on exactly how these laws function in terms of everyday morality.

### III. The Four Examples



So far, this explication has been mostly concerned with the theoretical aspects of the development of the Categorical Imperative. It is difficult to envision how Kant's moral theory works in practice without some concrete examples to demonstrate his theory in action. In the *Groundwork*, Kant recognizes this need as well, so he chooses to provide us with his famous four examples. A brief discussion here of the four examples would be appropriate for the purpose of demonstrating the practical functionality of the Categorical Imperative.

Kant's first example is of a man who is sick of life and is contemplating suicide. However, this man is not completely despondent and is still in possession of his reason. He seeks to find whether his maxim, which is, "from self-love I make it my principle to shorten my life when its longer duration threatens more trouble than it promises agreeableness," is in accord with the moral law. Kant's argument is that this man's maxim, far from being in accord with the moral law is in direct conflict with it. The man contemplating suicide forms his maxim out of self-love, which, by its nature, has the purpose of furthering life, not shortening it. If the man were to will that his maxim enjoining suicide out of self-love should become a universal law, then that would lead to a situation in which self love "would contradict itself and would therefore not subsist as nature." (G, 4:422) In other words, willing that this maxim should become universal law would lead to an intractable contradiction in which an inclination whose purpose is the furtherance of life has been subverted and now allows for the shortening of life.

Such a universal law could not possibly be willed without contradiction; so it can be judged that the man cannot morally act upon his maxim.

In the second example, Kant presents us with a man in financial trouble. He finds himself in need of a loan, but without the means of repaying that loan. His problem is that no one will lend him the money if he cannot hope to repay the loan. Therefore, his only option, if he still wishes to pursue the loan, is to promise that he will repay the loan within a certain period of time. Of course, he could only make such a promise knowing that it would be a lying promise, so still having a little bit of conscience, he seeks to determine whether or not his maxim, "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen," is in conformity with the moral law. Kant argues that it is not, because such a maxim, if it was willed to become a universal law, would lead to a contradiction, in which it would lead to its own nullification. That is because if it was a universal law that anyone could lie in order to secure a loan without any intention of ever repaying the loan, then no lender would believe anyone who made such false promises. Therefore, this maxim could not be a universal law, because the form of the law itself would make the law invalid.

Another perspective from which we may approach this example is from the FHE. Instead of seeking to know whether this maxim could be willed to become a universal law, we could instead determine if this law treats humanity as an end in itself of merely as a means to an end. If we approach the problem from this perspective, then it becomes immediately apparent that this maxim is not in conformity with the moral law, because it treats the money lender merely as a means to an end and shows no respect for

the money lender's own ends. The man who makes an empty promise to repay a loan is using his lender solely as a means to the borrower's end of getting some free money.

The third and fourth examples approach how we are supposed to use the Categorical Imperative from a different perspective than the two previous ones. Instead of dealing with an agent who has committed a specific action of questionable moral worth, such as suicide or lying, the next two examples present us with agents whose overall dispositions might be construed as being in conflict with the Categorical Imperative.

In the third example, Kant presents us with an agent who has been blessed with a certain talent. This talent, if cultivated, could be used for the great benefit of many people, but the agent in possession of the talent has no desire to pursue his talent or to develop it further. He has no need of money, for he is already in possession of enough money to ensure that he will live a comfortable life. Therefore, he has given himself entirely over to the pursuit of pleasure and does not concern himself with anything else. The maxim which he employs would look something like this: even though I am in possession of a talent that could be of a great benefit to a great many people, because I have no need of this talent, I may ignore it and pursue a life of pleasure. Now we must determine if this maxim could be taken up as a universal law. Kant denies that it could be, saying, "For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes." (G, 4:423) So, this maxim, if it allowed his talents to deteriorate, would be in conflict with his nature as a rational being, and that law which would command that we be in conflict with our rational natures cannot be willed.

Another way in which we could explain the example is through the FHE. In approaching it from this perspective, Kant argues that though the maxim that he has adopted does not force the agent to conflict with his ends, it does not harmonize with all of his ends. This is because “there are in humanity predispositions to greater perfection, which belong to the end of nature with respect to humanity in our subject.” (G, 4:430) This means that by allowing him to ignore the development of his natural talents so that he might pursue a life of pleasure, this maxim would actually enjoin that he neglect some of his ends, and would therefore, inasmuch as it does not seek to further all of his ends, be in conflict with the FHE.

The fourth example concerns a person who is not concerned with anyone else; let us name him agent A. The only ends that concern him are his own, and he has no desire to interfere with the ends of anyone else. His non-concern is rooted so deep in him that he even prescribes that no one is obligated to help him when he is in need. Thus he would not be inclined to act even if he were in a position to vastly improve the quality of someone else’s life, when it would cost him only a modicum of effort, and, of course, he allows that the reverse is true as well, not expecting any other agents to aid him if he were in need.

The question as to how we are supposed to morally evaluate this man now arises. In order for us to properly evaluate him, we must examine his maxims. In the case of agent A, instead of examining every maxim upon which he acts, perhaps we can, instead, determine a more general, overarching maxim, a fundamental maxim, upon which his other maxims are based. For A, it seems that this fundamental maxim would state: “I ought not to concern myself with anyone’s interest but my own, nor should anyone else concern themselves with my ends, and I shall not interfere with any of their ends.”



Now that we have a maxim upon which A bases his actions, and indeed all of his other maxims, we may evaluate that maxim in order to determine its moral worth. First, let us subject it to the FUL, so that we may determine whether that maxim passes the test of universalizability. In order to so determine, we should now examine the consequences of universalizing that maxim. The maxim, now reformulated as a universal law, would prescribe, "One ought to concern oneself solely with ones own ends and not the ends of any other and one ought not to interfere with the ends of others." Kant admits that this maxim passes the first test of universalizability, namely that it does not conflict with itself. So, the maxim of agent A could be formulated as a universal law, but that does not mean that it passes the whole test, for not only must it be formulated as a universal law, but one must be able to *will* that it be thus. According to Kant, this is where A's maxim fails the test, for A would, in effect, be willing that no one ever help him, even in situations of direst need, and, at least in Kant's view, a rational being could not will that such a state of affairs exist. How could a rational agent will that there should be a state of affairs in which he could very well be at a great disadvantage, especially when there is no apparent advantage for him if that state of affairs were to obtain?

Now that A's maxim has failed one formulation of the Categorical Imperative, let us see how well it fares against another, namely the FHE. Kant states that this maxim obeys this version of the Categorical Imperative only negatively, arguing that A gives accord to the ends of others only in the sense that A will not interfere with their ends, but that is not all that this formulation enjoins us to do. In that we are treating all people as ends in themselves, it is not enough that we merely cease hindering those ends, but we are also compelled to further those ends as well, as if they were our own

ends. As Kant contends, “For, the ends of a subject who is an end in itself must as far as possible be by *my* ends, if that representation (of the moral law) is to have its *full* effects in me.” (G, 4:430) This is where this maxim fails the test, for it does not obey the whole of the FHE. Inasmuch as the maxim has utterly failed two different formulations of the Categorical Imperative, it is not a maxim that should be acted upon.<sup>9</sup>

The four examples serve to show how the Categorical Imperative works in practice, and, additionally, Kant uses them to demonstrate how a maxim can violate the Categorical Imperative, as well as the types of duties that are entailed by the Categorical Imperative. A maxim may violate the strictures of the Categorical Imperative by being either a *contradiction-in-conception* or a *contradiction-in-will*. A maxim would be a contradiction-in-conception if it contained an internal contradiction. This is aptly demonstrated in the first example, in which the agent is attempting to determine whether suicide on the grounds of self-love is acceptable. In this instance, the maxim itself contains a contradiction, in that self-love, which should be directed towards furthering one’s own good, is leading one to commit suicide, the very opposite of a good. (G 4:422) A contradiction-in-will occurs when a maxim contradicts itself when it is willed, as in the fourth example. There is no contradiction in the formulation of that agent’s maxim, but one quickly becomes apparent when he attempts to will that maxim into a universal law, as it would then contradict his rational nature.

The types of duty that Kant wishes to demonstrate are perfect and imperfect duties. A perfect duty has to do with the rights of others, and allows for no variation. The second example demonstrates a perfect duty, because the agent’s maxim would entail a violation of the rights of the money lender. These are distinguished from

---

<sup>9</sup> Of course, it is sufficient for the maxim to have failed one formulation of the Categorical Imperative.

imperfect duties because imperfect duties concern the furtherance of our ends and the ends of others, and thus allow for a little flexibility, owing to the variation in those ends. (*Metaphysics of Morals*, 6:240)<sup>10</sup> The agent's maxim in the third example violates his imperfect duty, because it would prevent him from furthering those ends imparted to him by his rational nature, though the substantive form of that duty would have to be determined based on those particular ends.

I have tried to portray, admittedly schematically, that Kant's ethical theory is a consequence of his metaphysical position. If there was no need to look beyond the phenomenal realm for human autonomy, then Kant would not have been able to formulate his moral philosophy the way that he did. However, it was clear, at least to Kant, that if there was not something beyond this universe of space-time, then all hopes of human freedom would be doomed. Therefore, he had to posit the noumenal world as the location of human freedom, and the Categorical Imperative as the means by which we are able to be free. Thus he sometimes refers to the different formulations of the Categorical Imperative as the Laws of Freedom. (*G*, 4:387) Now that we have progressed this far in our understanding of Kant, we are at a position in which we are able to accurately examine Williams' critiques of Kant and to evaluate their force.

---

<sup>10</sup> The *Metaphysics of Morals* will hereafter be referred to as *MM*.



## **Section II.**

### **The Ethics of Bernard Williams**

## **CHAPTER 3**

### *“Internal and External Reasons”*

Now that we have examined Kant's moral philosophy, it is time to turn to that of Williams. Just as we did with Kant, we will begin by exploring the very basis of Williams' philosophy – his account of psychology – then we will look at the effect this had on his criticisms of Kant and his mature ethical theory. I will argue that Williams' paper "Internal and External Reasons" plays the same role in relation to his fully developed philosophy as the *Critique of Pure Reason* does to Kant's moral philosophy.<sup>11</sup> Therefore, if we can gain a detailed understanding of this seminal text, then we will be able to give Williams' later philosophy a much closer reading.

## I. Internal Reasons



In "Internal and External Reasons," Williams concludes that the only reasons that can provide an impetus for action are what he terms "internal reasons." Williams characterizes these reasons as reasons stemming from our *subjective motivational set*, or *S*. He characterizes *S* as containing:

dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying the commitments of the agent. (IER, 105)

In short, our *S* is what defines and identifies us as individuals. An external reason would then be a reason which bears no relation to *S*, and would therefore be given to the agent from outside of himself. Williams contends that there are no such things as

---

<sup>11</sup> Hereafter IER

external reasons, and the only reason that some would posit external reasons would be due to misunderstood internal reasons. (IER, 111)

Since Williams' eventual conclusion is that internal reasons are the only reasons which are able to motivate people to action, he spends the first half of this paper developing his theory of internal reason statements, especially a *sub-Humean* model. (IER, 102) In order to accomplish this goal, Williams develops four general rules governing internal reasons. His first rule is the most basic and concerns the very root of what an internal reason must be, saying:

- (i) An internal reason statement is falsified by the absence of some appropriate element from *S*. (IER, 102)

So, if I gave a reason for some action,  $\phi$ , that I performed, and I could not ground that reason in my *S*, then the reason that I gave would not be an internal reason; alternatively, the reason that I give for performing  $\phi$  must bear some relation to my *S*.

Williams identifies a problem with the preceding statement, asserting that it does not cover all of the instances in which an internal reason statement could be falsified. He does not wish to claim that false beliefs are able to give rise to reasons for action, since that would be, to a certain extent, to ignore the agent's rationality, something that Williams wishes to avoid.<sup>12</sup> An example of one of these instances would be an agent whose car has run out of gas and who believes that filling the tank with water will have the same effect as filling it with gasoline. Of course, if he filled the tank with water, he would be in a worse situation than he was previously, and he would have in no way approached his goal of getting his car to run again. Therefore, it would be absurd to say

---

<sup>12</sup> This issue will be dealt with presently



that he had a reason to put water in his gas tank, since his desire is based on a false belief.<sup>13</sup>

Considering this, Williams gives us his second rule governing internal reason statements:

(ii) A member of *S*, *D*, will not give *A* a reason for  $\phi$ -ing if either the existence of *D* is dependent on false belief, or *A*'s belief in the relevance of  $\phi$ -ing to the satisfaction of *D* is false. (IER, 103)

The example which Williams uses to demonstrate this rule is of an agent who wishes to drink a glass of gin and tonic. He believes that the liquid in front of him is gin, but it is really petrol. If we allow that elements in *S* which are based on false belief do give rise to internal reason statements, then the agent really does have a reason to drink the petrol, but the fact remains that the agent does not wish to drink petrol, he only wishes to drink the petrol that he mistakenly believes to be gin, which is a different case entirely. Then it would seem that if he knew that what he thought was gin was actually petrol, then there would be nothing in his *S* which would give him a reason to drink the petrol, meaning that he cannot have an internal reason to drink the petrol.

Now, this does not mean that false beliefs cannot be used to *explain* actions. There would be nothing wrong in our using the fact that the agent believed that the petrol was gin in attempting to figure out why he drank the petrol. What Williams is worried about is that we would use this explanation for his action to claim that he, in fact, had a *reason* to drink the petrol. Williams warns us away from such an interpretation by saying,

---

<sup>13</sup> One might assert here that the agent has an external reason not to put the water in his tank. We need not go so far, as his desire is to fill his car with gas, and not with water.

It (equating explanations with reasons) looks in the wrong direction, by implying in effect that the internal reason conception is only concerned with explanation, and not at all with the agent's rationality... But the internal reasons conception is concerned with the agent's rationality. (IER, 102)

It is doubtful that the agent would admit to having a reason to drink the petrol, if he knew that he was mistaken in believing it to be gin. If the agent persisted in drinking the beverage, even after he discovered that it was petrol, then our inclination would be to call him deficient in some way. If, however, internal reasons statements were only concerned with *explaining* actions, then we would be entirely unable to make such a judgment. Since internal reasons are concerned with an agent's rationality, there must be some further proviso by which we can distinguish between acting on true belief and acting on false belief, as the former gives rise to rational reasons for action, while the latter does not. That is the reason behind Williams' second rule.

Williams includes the second part of (ii), which discusses situations in which the agent possesses some false belief about the relevance of his  $\phi$ -ing to the satisfaction of some element of his *S*, to cover a different way in which a false belief could instigate an action. We may use the same example of the agent who believes that the petrol is gin in illustrating this new facet. Suppose that someone informed the agent that it was, in fact, petrol in front of him and not gin, but that our agent persisted in his desire to drink the petrol because he thought that drinking the petrol would have the same effect as drinking the gin. Though the particulars of this example are a little different from the preceding one, the effect is quite the same. The agent's drinking the petrol would not produce the same effect as drinking the gin, and thereby it would not satisfy the

particular element of his *S* that would be satisfied by drinking gin. Therefore, he would not have an internal reason for drinking the beverage because his real desire is for the pleasure of a gin and tonic, and drinking the petrol will not contribute to that in any way.

Having disposed of how false beliefs invalidate internal reason statements, Williams develops the third of his four rules governing internal reason statements:

(iii) (a) A may falsely believe an internal reason statement about himself, and (we can add)

(b) A may not know some true internal reason statement about himself (IER, 103)

This rule might seem self evident at this point, but it is necessary, given the previous rule, and, furthermore, it will have important consequences when Williams begins arguing against external reason statements in earnest later in his paper, which will be discussed shortly.

The ways in which one can falsely believe an internal reason statement about oneself are many. There is the more simple case of falsely believing something that is verifiable empirically, such as in the example of the agent who believes that the petrol is gin. The more complicated example would be a false desire or inclination, such as when an agent thinks that he wants to major in biological sciences while his true inclination is to become an English professor. Considerations such as these raise many other questions such as how the agent is supposed to identify a false desire, or how these false desires come about, but such questions need not concern us here; it is enough for our present purposes to have described this facet of internal reason statements.

That we may want something and not realize that we want it should be apparent to all of us. The tougher question is how we would incorporate this unknown desire into an internal reason statement. It seems somewhat implausible that an unknown desire can be strong enough to precipitate an action and still remain hidden to our conscious minds, yet all too often it is true; just ask a child why he has misbehaved and the invariable response will be, "I don't know."<sup>14</sup>

The lingering problem with this feature of such elements of *S* is that their relation to our actions is largely unknowable. (IER, 103) If an unknown element in *S* was to play some part in an internal reason statement, all of the rules governing internal reason statements would still apply, meaning that the unknown element would have to have some identifiable (although not yet identified) link to the statement. This is problematic in that, inasmuch as the element in *S* is unknown, it would be exceedingly difficult for the agent to discern the link. Though this is a definite problem, it is not necessarily a problem with Williams' theory of internal reasons. The problem may merely lie in the fact that we have imperfect self knowledge, and this has no bearing on Williams' arguments.

However, Williams' theory would seem doubtful if he did not include a way for us to discover these hidden elements of *S*, and he does so in his fourth rule:

(iv) internal reason statements can be discovered in deliberative reasoning. (IER, 104)

---

<sup>14</sup> I am not asserting here that this response is truthful all of the time; it only needs to have been true once to prove my point, and that is certainly plausible.

This rule is crucial in turning Williams' contentions about internal reasons into a viable theory. This consideration allows our *S* to be a dynamic thing, and allows for the many changes which people make to their personalities in the course of their lifetimes. After all, I am not born with a subjective motivational set that is broad enough to sustain me for the rest of my life.

Deliberative reason can play many different roles in determining and modifying my *S*. Williams, in describing the changes that deliberative reason can make on my *S*, says:

As a result of such deliberative processes an agent can come to see he had reason to do something which he did not see he had reason to do at all. In this way, the deliberative process can add new actions for which there are internal reasons...

(IER, 104)

It could be used to increase my *S* by evaluating two different elements and determining that, taken together, they give rise to a new element. An example would be of a person, who is a fan of both rock and roll and jazz, who realizes that he is likely to enjoy fusion, since it is a combination of themes from both rock and jazz. Similarly, deliberative reason can modify existing elements of *S*, by reducing a certain element's weight in deliberative matters, or even eliminating it entirely.

Williams does not just limit the changes which reason can make to *S* to the narrow confines listed above; those are just the obvious ways in which deliberation is able to alter my *S*. There are other possibilities. Examples which he gives of these "wider possibilities" include: "time ordering," "evaluating the relative importance of different elements of *S*," or "finding constitutive solutions." (IER, 104) An example of

time ordering would be an agent determining which element of *S* should be satisfied first; such as when I determine whether I should eat breakfast or take a shower first when I get out of bed in the morning. I will eventually do both, but the order in which I do them is not static, and is dependent on certain other elements in *S*; perhaps I had a late dinner and am not feeling particularly hungry when I wake up. After deliberating about this set of circumstances, it is likely that I will take a shower first.

Again, reason can be used when two or more elements of *S* conflict, in order to determine which element is subjectively more important to me. Suppose that I have a desire to buy both a car and a boat, but I only have the money to make one of these two purchases. Here my deliberative reason takes over and examines both of these desires in relation to one another and in relation to other elements of my *S*. I might come to realize that the car would be a much more practical purchase, since I will use it every day, and there is an element within my *S* which inclines me towards more practical desires, when my desires are in conflict. In this example, I have used my practical reason to evaluate different conflicting elements in my *S* so that I may determine which one is more important to me.

The final instance of the “wider possibilities” with which Williams provided us is the case in which deliberative reason can find “constitutive solutions.” This means that we are able, through our practical reason, to determine ways in which our desires can be satisfied. So, given that I am hungry and have a desire to eat, I am able, through rational deliberation, to determine the best way in which to satisfy this desire, taking other elements of my *S* into account, such as my desire to stay healthy, my predilection for Mediterranean food, my desire to remain financially solvent, etc.

What Williams is attempting to do here is to characterize his picture of *S* and its relationship with practical deliberation as being broad enough to cover the myriad range of human actions. In fact, he must do so, since he will assert that there is no such thing as a true external reason statement. If all of our actions are to be based upon internal reason statements, which in turn must be based in our *S*, then our *S* needs to be sufficiently able to provide for *all* of the actions that humans have been observed to undertake. This is why Williams takes great care to mention these “wider possibilities” for interaction between *S* and practical deliberation; if he did not grant those possibilities, then an *S* in itself would not have been able to support those actions, and vital territory would have been consigned to the external reasons theorist. However, it remains to be seen whether he has granted too much and has overstepped the boundaries of what can reasonably be expected to be solely within the powers of *S*.

At this point, then, we have a reasonably complete picture of Williams’ theory of internal reasons statements. They are the statements that we use in describing what impels us to action, and they are derived from our *S*. Since our *S* is subject to change and modification, so are our internal reason statements; a statement which was true a month ago may be false now, because of some practical deliberation that I have performed in the intervening time. Likewise an internal reason statement which was untrue a month ago may be true now, due to some modification of my *S* that came about through practical deliberation.

## II. External Reasons



Having completed his description of internal reasons, Williams proceeds to debunk the notion of external reasons. Williams gives as his example, the story of Owen Wingrave, a character from Henry James' story of the same name. Owen is a young man with a very martial heritage, in which every male ancestor from time immemorial has taken part in the military. Owen, however, has no desire to join the military; in fact, he despises every aspect of the military life. His father, of course, tries to convince Owen to join the military, giving him reasons such as that family honor requires it. So Owen's father's claim would be an external reason statement. It could be phrased "*that there was a reason for Owen to join the army.*" (IER, 106) The thing to notice about this statement is that it is phrased entirely in the third person. Owen has nothing in his *S* which would lead him to desire a military life, so any reason statement concerning why he should join the military has to be made without any reference to Owen's *S*.

The appropriate question to ask at this point, according to Williams, is how can an external reason motivate someone to act, if it refers to no internal motivations of the person? It seems doubtful that Owen's father claiming that family honor should impel Owen to join the military would convince Owen to do so, if Owen had no preexisting desires sympathetic to military life or to family honor. Williams claims here that:

nothing can explain an agent's (intentional) action except something that motivates him so to act. So something else is needed besides the truth of the external reason statement, some psychological link; and that psychological link would seem to be belief. (IER, 107)



What Williams means here is that when an agent is presented with what appears to be an external reason statement, he would have no motivation to act upon that statement, because there is nothing in his *S* which would impel him to so act, and, therefore, would not be able to act on that statement. If an agent were to act upon an external reason statement, there would have to be some hidden factor by which the agent could acquire the necessary motivation to act. Williams claims that belief in the truth of the external reasons statement could provide enough of a psychological link to the agent's *S* so that the agent would have sufficient motivation to perform the action enjoined by the external reason statement – which, by now, is no longer external.

Now that we have come this far, we must examine the way in which belief can give rise to an internal reason statement. First, we need what would appear to be an external reason statement, and an agent who does not initially believe that statement. Let us posit an agent, A, who is currently deciding a course of action. A is in the process of deciding whether or not to continue his higher education. His parents were assisting him with his financial obligations, but they have fallen on hard times and are no longer able to do so. If A wishes to complete his course of study, he will have to take out substantial loans, something he is loath to do. A is leaning towards dropping out of school when one of his closest confidantes tells him that he has an obligation to finish school, so that he might honor his parents sacrifice. A has not considered this line of reasoning before, because he generally does not think in terms of the sacrifices that others have made for him. Therefore, this would be an external reason statement, because A does not believe that he is in any way obligated by sacrifices made by others. However, A's persuasive confidant perseveres and causes A to believe in the external reason statement. At this point, a definite psychological change has occurred, in

relation to A and the external reason statement. Earlier, before he believed in its truth, it would be easy for A to disregard that statement, as it (at least in Williams' eyes) holds no claim on him, since it is outside of A's *S*, and there would be no way by which it could lay claim on him. However, now that A believes that the statement is true, that statement has, in a way, become part of his *S*. He now has, as a part of his subjective motivational set the disposition that he ought to honor the sacrifices that others have made for him, and this has happened because A now *believes* in the truth of that statement.

In light of this example, we should now return to that of Owen Wingrave, so that we might see what effect this new consideration might have on it. Would there be a valid reason for Owen joining the military, if he came to believe that it was his duty to uphold the family honor? Would this still constitute an external reason if Owen came to believe it? Let us grant that Owen comes to believe that he does have a duty to uphold the family honor. Whether Owen would be motivated to act on the external reason statement is still unclear, because a belief in a duty is not necessarily sufficient for one to be motivated to act on that duty. It would seem that there is still some piece missing from the equation, if the external reasons theorist is to satisfactorily explain how one could act on an external reason statement merely through belief in that statement. Williams claims here that it would now appear to be an internal reason, since one who would be motivated to act based on his beliefs, or based on what he believes his duty to be, must have "a certain disposition to action, and also dispositions of approval, sentiment, emotional reaction and so forth." (IER, 107) This would mean that the agent who acts on an external reason statement that he comes to believe would then have an internal reason on which to act, since it now refers to some element of his *S*.

If this is the only way in which someone could act on an external reason statement, then it seems that the external reasons theorist is in some trouble, because the only way for someone to act on an external reason statement is to turn it, through belief, into an internal reason statement. Williams asserts here that, "Now it does not follow from this that there is nothing in external reason statements." (IER, 107-108) The first thing to consider when evaluating this claim is that it is not enough for a person simply to have a disposition to act on his beliefs when trying to determine if external reason statements give rise to action. First of all, it is not the case that all beliefs require action. Just because Owen came to believe that family honor depended on his joining the military would not mean that *he* would be motivated to join the military, even if he was already disposed to act on his beliefs. For Owen to act upon his father's exhortations, for him to act upon an external reason, he must come to have a disposition to act on *that particular belief*. Since an external reason statement must be made independently of an agent's *S*, if the statement is to be a true external reason statement, the agent must be in no way inclined to act upon it at the time it is made.

Therefore, coming to believe an external reason statement and acting on it would require "acquiring a new motivation." (IER, 108) The question now becomes how do we account for acquiring a new motivation? Williams claims that there must be a special connection between "acquiring the motivation and coming to believe the statement." (IER, 108) What this means is that the agent's acquiring of the motivation must come about as an effect of his coming to believe the external reason statement. Of course, now the external reasons theorist must explain how acquiring the belief would give rise to the motivation for performing the action, without turning the external reason statement into an internal reason statement.

One possible way, Williams suggests, for the external reasons theorist to explain how the necessary motivation arises in the agent purely from belief in the statement is that the external reasons theorist can claim that the agent “should deliberate correctly.” (IER, 109) What this, “deliberate correctly” means is that there is something in the nature of practical deliberation itself which would lead the agent to become motivated to perform the action. Given a certain set of circumstances, there is only one answer that proper practical deliberation could give the deliberating agent as to what action he should be motivated to perform.

Williams’ response to this effort of the external reasons theorist is that it is implausible because, “there is no motivation for the agent to deliberate *from*.” (IER, 109) Remember, if the statement upon which the agent is deliberating is to be an external reason statement, then it can bear no relation to his *S*; he can have no motivation to perform it to begin with. In fact, Williams goes so far as to say (and rightly so) that if there was some sort of earlier motivation upon which the agent was able to base the ultimate motivation for his action, then the external reason statement would have to be considered an internal reason statement because there was already some element of *S* to which it appealed. It is difficult to see how an agent who has no preexisting motivations for a certain action can become motivated to perform that action based solely upon his deliberative reasoning. What base would his practical deliberation have in creating this motivation where none previously existed?

### III. The Problem with Irrationality



Having disposed of one way in which an external reasons theorist could assert the validity of his claim, Williams anticipates another avenue of approach. This time instead of claiming that practical deliberation was sufficient in itself for a motivation to perform an action based on an external reason, Williams says that the external reasons theorist could argue that an agent could act on an external reason statement if the “rational agent would be motivated to act appropriately.” (IER, 109) This claim could be made based on the fact that an agent is rational who, “has a general disposition in his *S* to do what he believes he has reason to do.” (IER, 109) This would answer the question as to how an agent would come to be motivated to perform an action simply because he has come to believe the external reason statement that enjoins him to perform that action; it is in his nature, qua rational agent, to do so.

This avenue of approach seems a bit spurious, because, Williams argues, “this reply merely puts off the problem.” (IER, 109) It answers the question of how belief in an external reasons statement can give rise to a motivation to action, but it does so only by begging the question, asserting that there was some preexisting motivation which would impel the agent to perform the action enjoined by the external reason statement. This approach, then, is fatal to the whole idea of the existence of external reasons in that it claims that one could perform actions prescribed by these statements if one already had a motivation to do so. This sounds suspiciously like it would make all valid external reason statements into internal reason statements.

At this point, Williams feels that one of the only plausible defenses that an external reason theorist would be able to offer is to claim that one who is not able to be persuaded by a true external reason statement is "*irrational*." (IER, 110) If this line of defense is successful, and if it can be proven that there is something to the charge of being irrational, then there is hope for the external reasons theorist. Williams does not think that the claim that one who is not persuaded by an external reason statement is irrational holds any weight. He attempts to show that such claims are spurious through the Owen Wingrave example. Williams claims that the injunction laid upon Owen by his father could have taken the form of an external reason statement, i.e., that upholding the family honor was a reason for Owen to join the military. (IER, 111)

Now, the failure of Owen's father to convince Owen to join the military could be due to the fact that the external reason is itself false, but Williams claims that the real importance lies in the fact that, "What the (external reasons) theorist would have a harder time showing would be that the words *meant* something different when used by Wingrave from what they mean when they are, as he supposes, truly uttered." (IER, 111) That is, if there could be a correct formulation of that external reason statement, the external reasons theorist must be able to say what it is, and how the true formulation is substantially different from the one used by Owen's father. Of course Williams, being in doubt about the whole business of external reasons feels that, given their vague nature, no such distinction can be made. He goes on to claim that the point of Owen's father's speech was not that if Owen were to reflect rationally, he would become convinced of the correctness of his joining the military. In light of all of this, Williams leaves us with two main assertions: first, that what a true external reason statement entails is not at all

clear; and, second, not all external reason statements base their persuasive powers on the agent deliberating correctly.

Williams thinks that he has now soundly rejected the idea of external reasons statements, saying:

The sort of considerations offered here strongly suggest to me that external reason statements, when definitely isolated as such, are false, or incoherent, or really something else misleadingly expressed. (IER, 111)

Indeed, Williams seems to have done a rather thorough job of attacking the concept of external reason statements. It would seem that Williams would lump most external reason statements into the category of “something else misleadingly expressed” because, according to his arguments, all so-called external reason statements which impel an agent to perform an action actually do depend on there being some element in the agent’s *S* which is sympathetic to the injunction of the external reason statement; thus, they are essentially internal reason statements. As to those external reason statements which are unsuccessful at enjoining an agent to action, they may be conveniently dismissed as false or incoherent, since they did not contain enough of a reason for the agent to act to convince him that he should act.

Of course, this does not mean that the external reasons theorist is left with no recourse - far from it. One possible route that the external reasons theorist could take - especially if that external reasons theorist has an interest in defending Kant - is to develop further the role that the agent’s rationality plays in respect to potential external reason statements. Williams did not leave this aspect of the subject untouched, but his attempt to dismiss the charge of irrationality does not appear to be entirely satisfactory,

depending, as it does, entirely upon the implications of the Owen Wingrave example.

It is difficult to overstate the importance of the role that “Internal and External Reasons” has played in the forming of Williams’ philosophy. For our present purposes, it is enough to examine how this paper has contributed to Williams’ critique of Kant. Indeed, the echoes of Williams’ theory of internal reasons can be felt throughout his different criticisms of Kant in his book, *Ethics and the Limits of Philosophy*, most especially in Williams’ objection to objectivism. However, we do not need to search through Williams’ later works to discover the implications to Kant’s moral theory which are contained within this paper. In fact, although Kant and the Categorical Imperative are mentioned only once in “Internal and External Reasons,” this entire paper has a direct relevance to the success or failure of the Kantian enterprise.

It is reasonable to argue, in light of the conclusions found in Williams’ paper, that the contest between Williams and Kant has become a zero sum game; their conclusions are as close as philosophy approaches to finding polar opposites. After all, Kant asserts that we can act morally only by basing our actions, through *no* influence of our desires and inclinations, firmly on the Categorical Imperative, while Williams asserts that it is impossible, ultimately, to act from anything other than from one’s desires and inclinations. Therefore, a well-considered response to Williams’ claims in this classic paper is needed for those who wish to keep the philosophy of Kant alive.



## **CHAPTER 4**

### *Ethics and the Limits of Philosophy*

It has been argued in the previous chapter that the position that Williams takes in his paper, "Internal and External Reasons," precludes any possibility of the moral theories of Williams and Kant coexisting. What remains for us to do is to examine the finished version of the product that Williams began in "Internal and External Reasons," so that we may have a better understanding of Williams' exact arguments with regard to Kant and ethics in general. To do so, we must turn to examine some of Williams' later works, most notably *Ethics and the Limits of Philosophy*.<sup>15</sup> It is in this work that Williams explicitly lays out his objections to Kant's moral theory and introduces his own dispositional theory of value. However, there are relevant elements of his philosophy that are important to understanding some of his criticisms of Kant that are located in his work *Utilitarianism: For and Against*, written with J. J. C. Smart. With a close examination of these works of Williams completed, we will have a firmer grounding on Williams' moral position, as it relates to Kant, and thus we will be in a good position to begin our evaluation thereof.

## I. Williams' Critique of Reason



The appropriate place to begin our task would be with the first of Williams' several criticisms of Kant in *Ethics and the Limits of Philosophy*, in which he questions Kant's ability to derive ethics from reason alone. The boldest and most interesting aspect of this criticism is that Williams attempts to derive Kant's conclusions while using none of Kant's methods. (*ELP*, 55) From there he attacks Kant's claim to universalizability, saying that there is nothing in man's rational nature that would

---

<sup>15</sup> Hereafter *ELP*

necessitate such a claim. It seems a risky thing, to attack someone's conclusions while taking an entirely different approach to reaching them, but, according to Williams, his approach is, "simpler and more concrete than Kant's," so perhaps he is onto something, if his claim is accurate. (*ELP*, 55)

The method that Williams uses in reaching the Kantian theory without the Kantian method is to examine the rational nature of man in order to discover any universal desires or needs imparted to us by our rational nature. He begins his examination with the question, "Is there anything that rational agents necessarily want?" intending to derive some ethical considerations from the answer. (*ELP*, 55)

The first step that Williams takes in answering this question is to say that when we act, we want certain "outcomes" to obtain. So, when I perform an action, for instance, studying French, there is a certain outcome that I am pursuing, namely fluency in French. My studying is merely the means to this desired outcome.<sup>16</sup> The important claim that Williams is making is that I must, *qua* rational agent, desire certain outcomes, and this seems a reasonable claim, because there would be no reason for me to be rational if there were no ends that I desire *qua* rational agent.<sup>17</sup> Inasmuch as we use reason to give us suitable methods for pursuing our ends, the role of reason would be much circumscribed, if not entirely marginalized, if we were creatures without ends imparted to us by our desires and inclinations.

Williams sums his position up till this point thus: "on various occasions we want certain outcomes; we usually want to produce those outcomes; we usually want to

---

<sup>16</sup> Of course, there are instances in which the action is both an end in itself and the means to this end, such as learning to play music, but this does not change the main point that when I act, I do so in the pursuit of a certain outcome. It is just that in certain instances the means and the end are the same thing.

<sup>17</sup> It seems that when Williams mentions outcomes, he refers to them with regards to the fact that they are a rational agent's end. Considering this, I will sometimes use them interchangeably,

produce them in a way that expresses our desire to produce them.” (*ELP*, 56) The last two clauses need not concern us as much as the first; as it is the strongest of the three claims and is also the most directly relevant to his project of finding a rational basis for ethics without resorting to any of those metaphysical claims of which he is so dismissive. Williams’ next assertion is the obvious extension of the first claim, that of rational agents necessarily wanting certain outcomes. If I am a rational agent, who not only desires these outcomes but desires to produce these outcomes, then I must want not to be frustrated in the pursuit of these outcomes. (*ELP*, 56) Williams means this only as a general description of what we are as rational agents. We must note that Williams is speaking here only in general terms. We are quite able to imagine a situation in which an agent would desire some outcome, but this desire does not constitute an important part of his *S* (the subjective motivational set). This means that he does not pursue this object with any zeal, if he pursues it at all. In such a situation, it would hardly matter to him if someone or something acted to frustrate his obtaining of that end.

It is unlikely that Williams is unaware of the possibility of exceptions to his claims about how we behave as rational agents. Rather the point that concerns us here is that the exceptions to these rules are more properly regarded as the outliers to the general rules with which Williams provides us. We do attach a considerable importance to the majority of our desires, such that we also desire not to be frustrated in their pursuit. Therefore, though exceptions are possible, it seems as though this consideration, concerning what rational agents desire, is necessary if the very concept of desiring something is to be coherent.

Another ancillary consideration that Williams identifies is that we must also want to be able to produce that outcome. (*ELP*, 57) Once again, this seems a precondition to

being able to coherently want anything. It is not clear how I could want some outcome to obtain, especially if I had some desire to facilitate that outcome, without wanting at the same time to be able to produce that outcome. Here, we must not confuse producing an outcome with physically producing an outcome. Saying that I want to produce something sometimes connotes the idea that I want to physically produce it, but this is not always the case. Indeed, there are cases in which an agent possesses an express desire *not* to produce the outcome physically while at the same time desiring to produce the outcome. An example of this would be when there is an industrial accident which released toxic waste. The mayor of the city in which this accident occurred certainly has a desire to clean the mess as quickly as possible, but at the same time, we could hardly imagine the mayor donning the necessary equipment and physically cleaning the waste. What we can picture the mayor doing is calling in the emergency response professionals and coordinating the clean up effort. In this way, an agent may desire to be able to produce an outcome without desiring to affect that outcome physically himself.

Included in this desire to be able to produce that outcome is the desire to have the knowledge necessary for that ability. The necessity of this consideration should be fairly apparent at this point, since one could scarcely be able to produce an outcome without possessing the knowledge required to be able to produce that outcome. Of course, this reduction of necessary ancillary desires could go on for quite some time, if we were to take it to its full extreme, but this is sufficient for the purpose at hand.

The point at which all of this discussion is aimed is that I, as a rational agent, must want *freedom*. I must want the opportunity to do what is necessary in the pursuit of my ends, and that is what Williams has defined as freedom. Here we must realize that the freedom of which Williams is speaking is not an unlimited freedom. If there

was nothing preventing us from obtaining our ends; if we could obtain our ends merely by desiring them, then there would be no need for us to be rational agents, because the role of reason would once again have become marginalized, although not in exactly the same manner as was discussed previously. In this case, I would have outcomes that I want to obtain, but, since I could cause them to obtain simply by desiring them, there would be no need for me to have a faculty, such as reason, which would instruct me as to the best manner of causing these outcomes to obtain. This means, according to Williams, that there are certain things that I must *not* want as “a condition of being a rational agent.”<sup>18</sup> (*ELP*, 58) Williams leaves us with a short list of things that we do not want, primarily omniscience and omnipotence, for in desiring such things we would necessarily be abnegating our nature as rational agents.

What, then, is this limited freedom that I desire? The freedom than Williams is characterizing is qualitatively different from Kant’s transcendental freedom. For one thing, Williams is seeking to describe this freedom without mentioning metaphysics or free will. Williams’ freedom is something altogether more limited in its scope, and it could succinctly be described as freedom from coercion or interference. (*ELP*, 58)

The temptation at this point is to say that freedom is unqualifiedly good. Williams is not yet ready to concede this point, because making this point would entail assuming that “This (freedom as an absolute good) would be so only if it were a good, period, that I should be a rational agent.” (*ELP*, 59) However, this particular assumption goes beyond what Williams is trying to establish. Aside from it being unnecessary, claiming that my rational agency is a good in itself would require that others should assent to that, since saying that something is good “imports an idea,

---

<sup>18</sup> Though, perhaps we could speak of wanting those things in principle, without any hope of actually obtaining them.

however minimal or hazy, of a perspective in which it can be acknowledged by more than one agent as good.” (ELP, 58) It is not at all clear that my being a rational agent would be considered good from a universal, or at least another agent’s, standpoint. Therefore, while it might make his arguments a bit more persuasive if he was to assert that rational agency is a good in itself, Williams deems such an assertion too problematic.

If I should not view my freedom as unconditionally good, then how *should* I view my freedom? The fact that I should not view my freedom as an unconditional good should not deter me from desiring that I should have freedom and that it is good for me to have freedom. Rather, it is possible for me to view freedom as good for me, as long as that is the limit to which I attach goodness to freedom. Whether or not I *must* view freedom in this manner is not entirely clear, but it should be apparent that I am able to coherently desire freedom as a thing that is good for me, in that my freedom is a necessary condition in the pursuit of my ends. One thing that is certain is that I must desire freedom, either as an unconditional good or not, if I desire certain ends, because without freedom, there is no way for me to obtain those ends. (ELP, 58)

Once Williams has established that I must want freedom, he is able to introduce the idea of legislation. His argument is as follows:

Since I necessarily want my basic freedom, I must be opposed to any courses of action that would remove it. Hence I cannot agree to any arrangement of things by which others would have the right to remove my basic freedom. So when I reflect on what arrangement of things I basically need, I see that I must claim a *right* to my basic freedom. (ELP, 59)

His argument here is fairly straightforward, and appears to be the natural extension of his earlier claims. It would be irrational for a rational agent to accept any state of affairs in which it would be acceptable for anyone to interfere with the agent pursuing his ends.<sup>19</sup>

The corollary of my making such a legislation is that I must now concede that other rational agents have the same rights that I am claiming. If I am basing my claim to this right on the fact that rational agency in itself is sufficient to entail the right, then there is nothing preventing others from claiming the exact same right. In fact, others need not even claim the right, for it is already theirs. In this way, I am not unique in any relevant aspect and therefore have no ground to assert any right that would give me some greater privilege than the other agents. It would be just as wrong for me to interfere with their freedom as it would be for them to interfere with mine.

Once Williams establishes that a rational agent should stake a claim to his freedom, and that the claim is based solely on the fact that the agent is a rational agent, he begins to approach some Kantian conclusions. Since this is precisely what Williams does not want to do, he begins to determine where he made his mistake. It is not in the first part of his argument in which he determines exactly what rational agents want, and nor is it in the last step of the argument, in which he universalizes from my claim to a right to freedom. The point at which Williams feels that he was in error was when I asserted my right to freedom, as this passage illustrates:

---

<sup>19</sup> I mean this statement only in its most general sense. There are limits to freedom that most people would consider rational, such as prohibitions against murder. The issue of what limits to freedom are acceptable to rational agents is one for a different thesis entirely.



But is the argument correct? Its very last step - that if in my case rational agency alone is the ground of a right to noninterference, then it must be so in the case of other people - is certainly sound. It rests on the weakest and least contestable version of a 'principle of universalizability,' which is brought into play simply by *because or in virtue of*. If a particular consideration is really enough to establish a conclusion in my case, then it is enough to establish it in anyone's case. That must be so if enough is indeed enough. If the conclusion that brings in morality does not follow, it must be because of an earlier step. (ELP, 60)

Of course, this earlier step is the assertion of a right, with the expectation that this right would cause some sort of obligation in other agents to respect it. Just why this was a mistake, we will now discuss.

In examining the part of his argument in which I assert my right to freedom, Williams' ultimate goal is to discover if there is any way for my claim to have normative force. The way in which he does this is by examining two rational agents, call them A and B, who both stake their claim to a right to freedom. It is in seeing how these claims, and the agents who made them, interact that Williams demonstrates the unsuitability of rational agency as the sole basis for reaching my right to freedom, and so to morality.

Williams begins this examination by presenting the thought processes of agent A that led A into prescribing a law:

I have certain purposes.

I need freedom to pursue these or any other purposes.

So, I need freedom.

I prescribe: let others not interfere with my freedom. (ELP, 60)

When A regards his prescription, he must regard it as in some respect a reasonable prescription. When confronted with agent B, who has also made a prescription enjoining others not to interfere with his (B's) freedom, it would seem that agent A must allow that B's prescription is reasonable as well.

In conceding that B's prescription is reasonable, does it not mean that A has been obliged to abide by B's prescription? *Williams' contention is that this has taken A's concession a step too far.* A has not said that he has placed himself under that prescription, nor has he said that it would be rational for him to obey B's prescription. In fact, A might not even believe that B would be rational in obeying A's prescription. A's prescription did not contain any element which would have any obligatory force over any other rational agent. A has made a prescription stating what other agents should do, with regards to what would be good for A. But, is it a reason for B not to interfere with A, just because A does not want him to?<sup>20</sup> The situation is likewise for A. Remember, A has only admitted that B's prescription is reasonably related to B being a rational agent (just as his own prescription is reasonably related to A being a rational agent) but it gives A no reason not to interfere with B.

Another thing to consider here is the form of the rule that A enjoins. If A wishes to make a rule, why does that rule have to take the minimally beneficial form that it does? If we talk of an agent making a rule based upon the desires which stem from the agent's rational nature, then why must we stop with a rule that just prescribes that others not interfere with that agent's ability to pursue his ends? If it is rational for A to make a rule protecting his right to freedom, could he not also make a rule that enjoins

---

<sup>20</sup> Of course, our ethical impulses might tell us that there is reason for B to respect A's desires, but we must remember that we have not yet reached the stage in this situation at which either of our agents would have any ethical notions.

others to pursue A's ends? Certainly that would be more beneficial for A, if he was to make a rule that commanded other people to do A's bidding. In that way, not only would A be protecting his ability to pursue his ends, he would also be furthering his ends at the same time. But of course, such a rule would be beset by the same problems by which the first of A's prescriptions was troubled, namely, that the most that B's rational nature would force him to admit is that it is reasonable for A to make such a rule, not that it is reasonable for B to obey the rule which A has made.

Considering this, it is surprising that A would make such a limited prescription, until we consider that A has no way of coercing others into obeying his rules. This is not a case of Athens and Melos, but a case of equals, and if A made a rule that would command other agents to pursue not their own ends, but A's ends, there would be small likelihood that the other agents would obey. Instead, he would have to temper his rule in such a way that it would be appealing enough to all rational agents that they would *want* or would feel *rationally obliged* to accept it. It seems very unlikely that agent B would want or would feel obliged to follow a rule that commanded that he further A's interests, especially if this comes at the expense of his own.

Williams claims that this caveat would lead to such questions as "what law could I make that I could reasonably expect others to accept?" and finally, "what laws should there be?" (ELP, 63) In this way, Williams attempts to approach the Categorical Imperative, inasmuch as the Categorical Imperative is a rule that is made from a standpoint of impartiality and equality. (ELP, 66) But even this final formulation of a rule that an agent could make from considerations arising out of his rational nature still falls short in one respect: it fails to introduce any form of obligation.

Williams' conclusion, in the case of making rules solely from considerations introduced by our rational natures, is that "one could never get to the required result, the entry into the ethical world, just from the consideration of the *should* or *ought* of rational agency itself." (ELP, 61) In determining how Williams was able to reach such a strong conclusion, we must recall his arguments in "Internal and External Reasons," in which he asserted that I cannot act on a reason that is not based in my *S*. This would mean that, unless A had some desire or inclination to obey B's prescription in his *S*, he would have no reason to do so. Of course, if there was some such inclination in A's *S*, then we would be no further on, because either A would already have a disposition to behave ethically, or he would have a disposition to act according to B's prescription, in which case "rational agency itself" would be an idle wheel. Either way, there would be no need to search for a way for A to become an ethical agent based solely on his rational nature. Conversely, if A was still to make his prescription that no one ought to interfere with his freedom, how could A reasonably expect another rational agent to obey A's prescription unless that agent had some element in his *S* by which he would be inclined to obey A's prescription? Thus, according to considerations stemming from "Internal and External Reasons," there is no way that rational agency by itself could lead us to morality.

Williams has a strong objection to any philosophy that describes humans as "rational agent(s) *and no more*," as he purports the Kantian system does. (ELP, 63) This is because he feels that any philosophy that does so is ignoring a large part of who we are as human beings, namely, we are our *S*. If we were just beings with a rational nature and nothing more, perhaps an ethical theory could be wrought which focused solely on our rational natures, but, unfortunately, this is not the case. It is true that any

ethical theory must take into account the fact that we are beings with a rational nature, but Williams argues that any ethical theory that hopes to be successful must take into account that human beings possess much more than our rational natures. Williams feels this so strongly that the rest of his arguments against Kant will be based on the fact that we are unable to act without regards to our *S*, which, at least in Williams' view, is exactly what Kant wants us to do.

## II. Can there be *Objective* Practical Deliberation?



The rest of Williams' arguments against Kant are critical of his objectivity, and may be easily divided into two related parts: one from questions of the first personal nature of practical deliberations and the other from the issue of integrity. Once Williams' criticisms concerning practical deliberation have been explained, it will become easier to see the reasoning behind his criticisms from integrity, so we will begin by examining his criticisms from practical deliberation. The crux of his objection from practical deliberation rests in his distinction between factual deliberation and practical deliberation. Williams' contention is that Kant has conflated the two, making his ethical theory, in practice, an impossible task.

Once Williams finishes his arguments concerning the impossibility of reaching any ethical considerations from considerations of our rational nature alone, he decides to remold his attempt at reaching Kant's conclusions, with more of an emphasis on Kant's own methods this time, although Williams is still leery of using anything resembling Kant's metaphysics. So now, instead of giving an account of what an ethical

theory would look like when wrought solely from considerations arising from our rational natures, Williams seeks to make an ethical theory out of what he calls “rational freedom.” (ELP, 65) Williams is still allowing that a rational agent must be committed to having his freedom, and is attempting to see whether this commitment to rational freedom can take us to being “citizen legislator(s) of a notional republic.” (ELP, 65)

Williams begins this new line of argument by seeking to determine just what rational freedom is, so that we may be better able to determine its consequences. He says that a crucial part of understanding this rational freedom is realizing that I, as a rational agent “can act *on* reasons,” meaning both that I am able to, at least in some way, do more than just act in accordance with laws, and that my behavior is not reducible solely to my beliefs and desires. (ELP, 65) That, as a rational agent, my behavior is not reducible to my beliefs and desires should not be a surprising assertion, but the difference between acting in accordance with a law and acting *for the sake of* a law is a subtle, yet ultimately important one.

If I am acting in accordance with a law, then that means that I did not involve myself in rational deliberation upon the law. A simple example of this would be if I was obeying the speed limit just because I did not feel safe driving any faster. In this case, I am guided by my own inclination, which just happens to be *in accord* with the law, but was formed without any reference to the law. By contrast, if I felt like driving fast, and noticed that the speed limit was much lower than I was inclined to drive, but I lowered my speed anyway, *because* of the law, then I could be said to be acting for the sake of the law. In doing so, I would be forced to rationally evaluate both my reasons for wanting to drive fast and the reasons by which the speed limit was determined, eventually reaching

the conclusion that the reasons for making the speed limit what it was superseded my reasons for wanting to drive at a speed that exceeded that limit.

Williams contends that an essential part of this rational evaluation is the agent's ability to "reflect on himself as an agent." (*ELP*, 65) In my previous example, when I was evaluating my reasons, a crucial step in my doing so was my evaluation of myself as a rational agent, for without viewing myself as one among many agents, it would be impossible for me to grasp the reasons for setting the speed limit. To do so necessarily entails viewing myself from some objective standpoint. When I am able to view myself as one among many agents, as I must do when I am determining the reasons behind the speed limit, then I am approaching what Williams calls "the standpoint of impartiality." (*ELP*, 66)

Another example that explains how rational reflection entails taking an objective standpoint, and does so clearly and precisely, would be an example illustrating an agent's thought process when purchasing clothing. When one makes a purchase of clothing, one does not do so from a strictly first-personal standpoint. Instead, the agent, who is evaluating a piece of clothing, looks at herself in the mirror and imagines how other rational agents would perceive her as she was wearing that particular piece of clothing. It is from this standpoint, the standpoint of objectivity, that the agent makes her decision. It is not how she feels about how she looks, but how she feels that others would feel about how she looks which determines whether or not she will make the purchase.

It is from this standpoint of impartiality that Kant wishes us to approach our ethical decision making. Before acting upon a desire, I am supposed to evaluate that desire from an objective standpoint so that I may determine if acting on that desire is a

moral thing for me to do. Williams characterizes this as turning a desire into a consideration. Instead of acting *from* desire, in determining how I will act, my desire merely becomes a minor element in my practical deliberations. So, objectivity plays an important, if not the most important role in practical deliberation, deliberation about what I should do, in this iteration. Now, in order to bring out the full thrust of this branch of Williams' critique of Kant, let us examine how he characterizes factual deliberation, and how this version of practical deliberation compares to factual deliberation.

Factual deliberation is deliberation over some empirically verifiable fact, such as there are two cars in my driveway, or that it is raining right now. As for factual deliberation, Williams admits that "what it [Kant's account of rational deliberation] says about reflection does indeed apply to factual deliberation." (*ELP*, 66) He does this because in any factual deliberation, there is an element of objectivity, there is something *outside* of myself to which I am referring; my desires and inclinations do not have to come into play. Unlike practical deliberation, factual deliberation does not have to occur in direct relation to me. To verify my claim about the cars in the driveway, all that someone would need to do would be to step outside and count the cars in the driveway. His judgment would not have to depend on his emotional state at the time, since he is deliberating about something outside of himself.<sup>21</sup>

Of course, there are factual deliberations which bear some relation to the self, such as judgments about color and one's physiological or psychological state, but these bear a relation to the self that is quite distinct from the relation to the self entailed by

---

<sup>21</sup> One might argue that the agent's deliberation might be affected by his emotional state, such as if he were in an agitated state and therefore miscounted the cars, but this error could be easily corrected, since the cars themselves do not depend on his emotional state.



practical deliberation. In the instance of color, there would be no disagreement about the fact that there was an object and that the object had a color. It would simply be a disagreement of perception or, even more simply, of nomenclature, neither one of which would be based strictly on desires or inclinations, thus they still could be resolved through rational discussion.

In the case of the physiological and psychological states, these are what one could call self-referencing factual deliberations, but they are still factual in the sense that they refer to some empirical aspect of the self. Williams gives us the example of the statement "I am ill." (*ELP*, 67) Even though this statement refers to me directly, it deals with my empirical self, and it is still a statement which could be verified objectively, for example, by my being examined by a doctor. Psychological statements such as "I am happy" or "I am sad" are somewhat more difficult of an issue in that they are difficult to verify empirically, and the states of affairs that they describe are located entirely in an agent's mind. Their verification, however, would still involve factual deliberation in that they are still statements describing states of affairs; it is just that the ways of observing these states of affairs are not as definitive as the ways of observing other states of affairs. So, it can be seen that factual deliberation is deliberation that has an empirical, and not a first personal base, even when the statements which are being deliberated about refer to oneself.

We have already touched upon practical deliberation, but since this argument rests solely upon the distinction that Williams is able to draw between practical and factual deliberation, it is worth reiterating exactly what practical deliberation is, so that there will be no confusion. Simply put, practical deliberation is deliberation about action. Therefore, "what should I do?" is the essential question to be answered by

practical deliberation. Williams holds that Kant's theory of rational deliberation is entirely inapplicable to practical deliberation because practical deliberation is "first-personal, radically so, and involves an I that must be more intimately the I of my desires than that account allows." (ELP, 67) Once again, we see the effects that Williams' arguments in "Internal and External Reasons" have in forming his later philosophy. There can be no doubt, if his conclusion held therein is true, that practical deliberation must be first personal; any practical deliberation that I undertake must have its grounding in my subjective motivational set, my *S*. To emphasize the importance that he attaches to this claim, Williams says that practical deliberation "is in every case first-personal, and the first person is not derivative or naturally replaced by *anyone*." (ELP, 68) Therefore, there is no way to have what could be termed an objective practical deliberation.

Williams' critique rests on the contention that practical deliberation is in no way congruent with factual deliberation. For, in his theory, factual deliberation can be done, in principle, by *anybody*, while it is *I* who must essentially perform practical deliberation. In the case of counting cars in the driveway, any rational human being would come to the same conclusion. My personal tastes and preferences would not have any effect upon the result of that deliberation. Therefore, the distinction, and in fact this whole line of Williams' criticism of Kant, rests upon the contention that I am able to be objective in my factual deliberation, while I *cannot* be objective in my practical deliberation. Being objective in our practical deliberation is the very thing that Kant commands us to do, and he contends that it is the only way in which we are able to *express* our autonomy; yet, it is one thing Williams cannot allow. In fact, Williams

would claim that deliberating in such a manner as Kant commands would effectively *destroy* our autonomy.

### III. Does Kant Destroy Integrity?



There is one final approach that Williams takes in his critique of Kant, and that is his critique from integrity. This is an idea which he only mentions briefly in *Ethics and the Limits of Philosophy*, but he had developed it more fully in *Utilitarianism: For and Against*.<sup>22</sup> His contention is that, if a Kantian system of morality was even possible, something of which he is in doubt, then there would still be the problem that in following this system, I would be violating my integrity. By integrity, I take Williams to mean, roughly, the state of pursuing my own *commitments* without any outside interference. Williams does not mean commitment in the everyday sense of the word, but attaches to it a special importance, which will be discussed presently. For now, it is enough to say that a commitment is a special kind of *project*. Of course, a Kantian morality would violate Williams' idea of integrity, because my personal projects do not attract a high deliberative priority in such a moral theory. In fact, Kant would most likely go so far as to claim that the only way that an agent is able to preserve his integrity is by living according to the moral law, since the projects given to him by his desires and inclination are given heteronomously and were not produced as expressions of the agent's free will. This is yet another point on which these two great philosophers are diametrically opposed.

---

<sup>22</sup> Hereafter, *UFA*

Why is integrity such an important issue to Williams?<sup>23</sup> He argues that integrity, in the way that he describes it, is essential to our very identities, which he claims are formed to a large degree by our “commitments.” If there was a moral theory which might require us to abandon these commitments, then that moral theory would require us to relinquish parts of our identities, something to which Williams stringently objects. It is, therefore, understandable that he devotes an entire chapter to the issue of integrity in *Utilitarianism: For and Against*. It is there that he introduces the idea of commitments.

For Williams, a commitment is a special kind of project that I place as central to my own happiness or well-being. As a rational agent, I naturally undertake certain projects, such as making an A in a particular class, learning another language, or aiding victims of injustice. To a certain extent, the ethical theory to which I subscribe influences the kind of projects that I undertake. A utilitarian’s projects will to some degree be informed by his duty of increasing the overall welfare. Likewise, a Kantian’s projects will be influenced by his duty of acting for the sake of the Moral Law. However, projects need not be very ambitious. A project can be something as simple as cooking a special dinner for a potential mate. Any outcome that an agent identifies as being important in the pursuit of his ends can be called a project. Williams does not wish to artificially constrain what may be considered a project, so he is at pains to make sure he includes a myriad of different possibilities.

Now, certain of these projects are so intrinsic to who I am as an agent that their importance takes on an entirely new order of magnitude, such that they deserve a

---

<sup>23</sup> The answer to this question has its roots, just like most of the other elements in Williams’ philosophy, in “Internal and External Reasons.”

different name: commitments. A commitment can be defined negatively as a project, the absence of which will cause a debilitating harm to my well-being or happiness. We must be careful here so that we do not say that one commitment is the pursuit of happiness. That would be an error, because happiness can only be pursued indirectly, through the pursuit of another project. To say simply that someone is committed to the pursuit of happiness would be devoid of content, since there is no one thing that universally leads to happiness, and happiness can only be attained through the pursuit of other projects.<sup>24</sup> Instead, all of our *S*'s are different, and so the conditions that need to be fulfilled in order to attain happiness are different for each of us. Therefore, a commitment may have happiness as its end, but it cannot be a commitment to pursue happiness, as that would not, strictly speaking, be a commitment to anything. Rather, commitments must be to some substantive project of requisite importance to the agent. An example of a commitment could be my desire to pursue a PhD, to teach philosophy, or even my desire to raise my children well. The only criterion that a project has to fulfill in order to be a commitment is that it must approach a certain threshold of importance to the agent.<sup>25</sup> Williams even provides us with a list of potential commitments to demonstrate the wide range of projects that could count as commitments, saying, "One can be committed to such things as a person, cause, an institution, a career, one's own genius, or the pursuit of danger." (*UFA*, 112) Another way to characterize a commitment is to say that it is a project which is integral to the identity of the agent. (*UFA*, 112-113)

---

<sup>24</sup> Interestingly enough, Kant and Williams seem to be in agreement on this point.

<sup>25</sup> Of course, we are entirely unable to determine exactly where this threshold lies, as it is exceedingly difficult, if not impossible, to so quantify our psyches.

Given how he has described commitments, and how he places them as intrinsic to our integrity, it is now apparent why they are not trivial things to Williams. Their pursuit is necessary for us to live as fulfilled and happy agents. Therefore, Williams has strong objections against any ethical system which could compel people to abandon their commitments. When it comes to utilitarianism compelling an agent to abandon his commitments, Williams contends that:

It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. (*UFA*, 116)

Though he makes this statement in relation to the demands of utilitarianism, its point is equally applicable to any moral theory. We could easily imagine a case in which someone, call him Agent B, had a commitment to hating a particular ethnic group, call it group C.<sup>26</sup> Agent B's life project is to cause group C as much harm as possible, perhaps it is even to eliminate group C. Certainly this would be characterized as an immoral project by most ethical theories, including that of Kant, and they would command that Agent B relinquish his commitment. Of course, as it is a commitment to Agent B, abandoning it would cause him no small amount of mental anguish, as it made up a significant part of his identity. He would be left adrift, and would have to reinvent a large part of himself. I have made this example an extreme one on purpose; I did so in order to highlight how an ethical theory could demand that someone abandon their commitment, yet not all instances of this demand need be immoral.

---

<sup>26</sup> Williams never said that a commitment had to be benevolent, and it would appear to be arbitrarily limiting to characterize them as such.

On the other hand, there are certainly cases in which a person has a commitment, perhaps even a benevolent one, and it happens that that particular commitment is in conflict with some ethical duty. The most immediate one that comes to mind would be of an agent who has a commitment to his job at a non-profit corporation. However, his job is so demanding of his time that his children suffer as a result. His job is of a nature that he cannot simply reduce his hours, he must continue as is or he must leave the job entirely. Most ethical theories would require him to dispose of his ethical duties to his children before returning to his commitment or, in some cases, to relinquish his commitment entirely. It can be argued that they would command that he give up his job for the sake of his family. To Williams, this is anathema. He would characterize this as an ethical theory requiring that a person abandon his very identity.

#### IV. The Lines Are Drawn



Williams feels that Kant's moral theory does not respect the value of an agent's commitments, as is evident when he asks this question concerning integrity in a Kantian system: "How can an *I* that has taken on the perspective of impartiality be left with enough identity to live a life that respects its own interests?" (*ELP*, 69) Since a Kantian must act in accordance with the Categorical Imperative – namely – that the maxim by which his action is formed can be willed to be a universal law - then it is a requirement of Kant's moral theory that *every rational agent would be ethically compelled to act in the same manner in any given circumstance*. If Williams is correct in this, then this is a serious problem for Kant, since it seems highly counterintuitive that the only way that I

can act autonomously is acting just like every other rational agent when in identical circumstances. Did Kant lead us right back into determinism, although of a different form? It seems that, beyond merely requiring that we give up our commitments in certain circumstances, the Kantian system would remove all vestiges of individual identity, if Williams is correct.

In the final analysis, there are three broad areas in which Williams locates serious problems in Kant's moral theory: the problem of establishing ethics from considerations arising from rational agency itself, the problem of performing practical deliberation from an objective standpoint, and the problem of integrity. All three of these approaches can be seen to be direct consequences of the issues that arise in "Internal and External Reasons." Therefore, it seems that if one wishes to defend Kant from Williams' sallies, then a critical examination of the conclusions presented therein is in order. Of course, one might attack Williams' criticisms on other grounds, such as his glossing over of Kant's metaphysics, or problems with his argumentation in establishing these criticisms, but that would be attacking the weed and not the root, and thus shrinking back from the problems which rest at the heart of the matter. The whole of the Kantian enterprise rests upon the possibility of a rational agent acting upon, at least as Williams has described it, an external reason. Until that can be established, no stable defense of Kant from Williams' criticisms can be established.