

Kant's School of Morals: The Challenge of Radical Evil and the Need for Moral Education  
in *Religion within the Limits of Reason Alone*

Joseph Goski

Thesis submitted to the  
Faculty of Graduate and Postdoctoral Studies  
in partial fulfillment of the requirements  
for the Master's in Philosophy degree in Philosophy.

Department of Philosophy

Faculty of Arts

University of Ottawa

© Joseph Wyllie Goski, Ottawa, Canada, 2014

**Abstract**

My aim in this thesis is to demonstrate that common interpretations of Kant's theory of respect do not account for the motivation a subject feels to follow the moral law. A large number of interpreters focus on Kant's early ethical works—such as the *Grounding* and the *Critique of Practical Reason*—to justify how the moral law alone motivates a subject to act rightly.

However, by the time he published *Religion within the Limits of Reason Alone*, Kant had discovered the problem of radical evil—the fact that people tend to feel more motivated by the inclinations than by the moral law. Kant's solution to this challenge comes in the form of moral education: the contingent practices of historical institutions (factors that are extraneous to the moral law) are required to learn respect for the law. By the end of the *Religion*, it will be asked whether duty for duty's sake is ever achieved.

## Table of Contents

Abstract.....	ii
Table of Contents.....	iii
List of Abbreviations .....	v
Introduction.....	1
Chapter 1: The Motive behind Justice and the Problem of the Sensible Knave .....	8
Two strains of interpretation on the question of moral motivation .....	8
Evaluating Moral Motives.....	8
Interpretations emphasizing Self-Interest as the Motive behind Justice and Promising.....	13
Problems with Self-Interest: The Sensible Knave and Interpretations favouring Motivation by Justice Itself .....	16
Second Stage in the Development of the Virtue of Justice: Learning how to be Motivated by Justice Itself. ....	19
Stage 2: Learning how to be Motivated by Justice Itself .....	19
Chapter 2: Moral Law and Practical Reason.....	32
The Distinction between Morality and Legality through Respect.....	32
Reason and the Moral Law .....	36
Reason as Consonant with the Moral Law .....	36
Second Interpretation: distance between Reason and the Moral Law.....	42
Attempts to Avoid an Interpretive Impasse .....	53
Chapter 3: Real Opposition and the Discovery of Radical Evil.....	57
Why Religion?.....	57
What is Evil?.....	63
A priori Deduction of Evil.....	63
Consequences of Defining Evil as Subordination .....	75
Chapter 4: Moral Training and the Need for Non-Moral Incentives .....	78

The Social Character of Evil: a Crisis of Autonomy .....	79
The Ethical Commonwealth.....	82
Historical Churches as Means to the Ethical Commonwealth.....	88
Solution to the Corruption of the Churches: A Final Attempt.....	95
Consequences of this Analysis of Radical Evil.....	105
Conclusion .....	108
The Ideal of Autonomy .....	108
Limits of Autonomy .....	110
Kant's Ethics as a Solution to Empiricism? .....	112
Bibliography .....	116
Primary Sources .....	116
German editions of Kant's works .....	116
Secondary Sources on Kant .....	117
Secondary Sources on Hume .....	119
General Works .....	120

## List of Abbreviations

**Works by Immanuel Kant** (all citations of Kant’s works refer to the pagination of the *Akademie Ausgabe* edition found in the following English translations. I decided to use the Greene and Hudson translation of the *Religion*—since it is the edition most often cited in the secondary literature—which does not include the *Akademie* pagination. To fix this problem, I included both the *Akademie* and the Greene and Hudson pagination in my citations from the *Religion*.

- A.* *Anthropology from a Pragmatic Point of View*. Trans. Robert B. Louden. Cambridge: Cambridge University Press, 2006.
- Auf.* “An Answer to the Question: What is Enlightenment?” in “Toward Perpetual Peace” and *Other Writings on Politics, Peace, and History*. Ed. Pauline Kleingeld. New Haven: Yale University Press, 2006. 17-23.
- CF.* *The Conflict of the Faculties*. Trans. Mary Gregor. New York: Albaris Books, Inc. 1979.
- G.* *Grounding for the Metaphysics of Morals*. Trans. James W. Ellington. Indianapolis: Hackett Publishing Company, 1993.
- KpV.* *Critique of Practical Reason*. Trans. Werner S. Pluhar. Indianapolis: Hackett Publishing Company, 2002.
- KrV.* *Critique of Pure Reason*. Trans. Werner S. Pluhar. Indianapolis: Hackett Publishing Company. 1996.
- MM.* *The Metaphysics of Morals*. Trans. Mary Gregor. Cambridge: Cambridge University Press, 1991.
- NG.* “Attempt to Introduce the Concept of Negative Magnitudes into Philosophy” in *The Cambridge Edition of the Works of Immanuel Kant: Theoretical Philosophy, 1755-1770*. Trans. David Walford. Cambridge: Cambridge University Press, 1992. 203-241.
- Rel.* *Religion Within the Limits of Reason Alone*. Trans. Theodore M. Greene and Hoyt H. Hudson. New York: Harper & Row Publishers, 1960.

## Introduction

It is easily noticed that David Hume's critique of causality has a devastating effect on both the common and the philosophical notion of cause and effect. If, as Hume argues, the mind does not perceive a necessary connection between two objects<sup>1</sup> (except for their resemblance and contiguity), then it follows that the concept of cause and effect must be a chimera of the mind, formed through the habit of regularly observing one object follow from the movement of a similar object preceding it (*Treatise* 63-66; 73-74). The argument depends on assumptions common to empiricism: all simple ideas start and can be resolved into impressions of the senses (*Treatise* 2-4); all knowledge about natural objects is obtained through the senses (*Treatise* 1-4); the senses do not receive an impression of the necessary connection between two objects (*Treatise* 54); therefore, the necessary connection commonly referred to as cause and effect cannot concern natural objects outside the mind.

If metaphysical concepts, such as cause and effect, are not perceived through the senses, then these concepts must remain mirages of the mind rather than descriptions of the world. The result, therefore, is that philosophical, theological, and ethical ideas relying on universal concepts for their conclusions can no longer be considered valid once these concepts are discovered to be fanciful ideas of the mind born through habit, rather than statements of fact. Shortly after Hume's publication of his argument, another philosopher was quick to notice the implications of these findings: Immanuel Kant.<sup>2</sup>

---

<sup>1</sup> *Treatise* 54, 56

<sup>2</sup> It should be noted that Hume alone is not responsible for Kant's realizing that the concept of causality could not be attributed to the world outside the mind. Lewis White Beck points out that, during Kant's early career, scholars in Germany were already subjecting the metaphysical concept of causality to critique. Kant's reading of Hume probably brought to fruition a critique of traditional metaphysics that he had already been reflecting

In the *Critique of Pure Reason*, Kant notes the effects that empiricism has on traditional notions in metaphysics, religion, and ethics:

On the side of *empiricism* in determining the cosmological ideas of reason...we find the following.

*First*, we find no such practical interest issuing from pure principles of reason as morality and religion carry with them. Mere empiricism seems, rather, to deprive both of these of all force and influence. For if there is no original being distinct from the world; if the world is without a beginning and hence also without an author; if our will is not free and the soul has the same divisibility and corruptibility as does matter; then *moral* ideas and principles also lose all validity, and fall along with the *transcendental* ideas that amounted to their theoretical support. (*KrV* 2: 496).

Since, according to the empiricist, “we are acquainted with nothing more than nature”, we are not allowed to “seek the cause for anything outside of nature (i.e., to seek an original being)” (*KrV* 2: 498). Such consequences of empiricism might be less troublesome to Kant regarding religion, since he is often critical of traditional religious practices or of the common notion of God (at least in 18<sup>th</sup> century European culture). However, Kant also observes that attributing the “same divisibility and corruptibility” to the soul as is the case for matter means that “the moral ideas and principles lose all validity, and fall along with the *transcendental* ideas that amounted to their support” (*KrV* 2: 496).

Kant’s strategy in responding to the critique of causality is to “remove the empiricism” (*KpV* 5:53). Kant’s solution lies in claiming that universal categories, such as cause and effect, are not features of the world, but are part of the structure of the understanding. The mistake made by Hume and other empiricist philosophers is to assume that the metaphysical categories of cause and effect apply to objects in themselves; they fail to notice, however, that these categories belong to the understanding in ordering sense data,

---

on for several years. For Beck’s historical analysis of the influence of Hume on Kant, see Beck’s essay “A Prussian Hume and a Scottish Kant” in *Essays on Kant and Hume*.

or *phenomena*. In other words, “objects can...be *thought* through [the categories] although not determined a priori” (*KpV* 5:54).

For Kant, the moral law is the only universal principle that differentiates between the good and evil object of the will<sup>3</sup> and, furthermore, subjectively determines the proper motive for acting.<sup>4</sup> The moral law acts as a moral determinant of the will precisely in virtue of the fact that it is *a priori*; otherwise, the law would remain contingent and the maxim of one’s action could not be probed for its conformity to the moral law through a test of universalizability. Furthermore, since the moral law requires a finite rational being to fulfill its demands, practical reason must postulate the existence of God, the freedom of the will, and the attainment of happiness.<sup>5</sup> The foundation of an *a priori* universal principle in the form of the moral law thus provides justification for concepts that had succumbed to the empiricist critique.

An important problem emerges, however, for an ethics established on an *a priori* law: why would a human being, who feels the influence of sensible inclinations, feel motivated by the moral law? Kant’s answer is found in the notion of respect: in removing all obstacles to its determining the will to right action, the moral law inspires respect—an *a priori*, yet sensible feeling. Commentators, however, often disagree as to how the law inspires respect: a common solution to Kant’s enigmatic argument concerning respect is to claim that respect consists of having esteem for one’s reasons for acting.<sup>6</sup> Even should this

---

<sup>3</sup> See the ‘Analytic’ of the *Critique of Practical Reason*, chapter II.

<sup>4</sup> See the ‘Analytic’ of the *Critique of Practical Reason*, chapter III.

<sup>5</sup> See the ‘Dialectic’ of the *Critique of Practical Reason*. I will explain in more detail the relation between the moral law and the traditional concepts of metaphysics at the beginning of chapter 3 of the thesis.

<sup>6</sup> I will cover various interpretations of respect in chapter 2 of the thesis.

interpretation be granted,<sup>7</sup> there remains evidence that respect does not provide sufficient motivation to heed the *a priori* law.

The disagreement and lack of clarity concerning respect and certain other key concepts of Kant's ethics are perhaps Kant's own fault. In a short article, Ralph Walker suggests that the misunderstandings surrounding Kant's philosophy are the result of his tendency to oversimplify or suppress key concepts "that are necessary to understand his position" (Walker 10). This tendency is worse when it comes to Kant's ethical writings, since the *Grounding* and the *Critique of Practical Reason* contain simplified ethical concepts that would find their full expression only later in the *Metaphysics of Morals*. Walker claims that these simplifications "have conveyed the idea that for Kant all that matters morally is that one should act out of a sense of duty, and that it is at best irrelevant whether one's feelings and sympathies are engaged; only the moral motive matter" (Walker 10). If Walker's claim is correct, then it is necessary to explore Kant's later works in order to overcome these caricatures and follow the evolution of Kant's ethical thought toward its more complete form. When we do so, however, a new set of problems emerge, one which led Kant himself to raise the question toward the end of his life as to whether or not human beings really do regularly feel motivated to heed the demands of the moral law.

The problem of moral motivation emerged most acutely only *after* Kant had published the *Grounding for the Metaphysics of Morals* and the *Critique of Practical Reason*: a brief glance at history reveals that human beings do not appear to be regularly motivated to follow the law at all. Once Kant begins to focus on more historical topics, the problem of leading sensuously inclined human beings to follow an *a priori* law becomes

---

<sup>7</sup> I will attempt to demonstrate in chapter 2 that this interpretation of respect should not be maintained.

more acute. The problem reaches a head in *Religion Within the Limits of Reason Alone*, in which Kant defines radical evil as the tendency of human beings to subordinate the influence of the moral law to the influence of the inclinations (*Rel.* 6:36). The task of the *Religion* will be to solve the problem of radical evil by slowly training corrupt human beings to feel respect for the moral law. Moral training, however, involves contingent, empirical institutions and motives leading human beings to respect the law. On the question of the subjective motive for following the law, Kant's solution to radical evil will involve the very empirical, contingent conditions that he had sought to eliminate in his response to the empiricist and Humean critique of causation.

My aim in this thesis is to establish that, as he confronts the problem of radical evil, Kant resorts to a form of education similar to that proposed by Hume in order to train the will into feeling motivation to follow the moral law. Kant's interest in later works with regard to human beings in their historical context leads him to confront the problem that people often feel greater motivation to follow the promptings of their inclinations than to follow the exigent demands of the moral law. This problem—defined by Kant as radical evil—recurs continuously in the *Religion*, especially once Kant notices that the simple presence of other people lessens motivation for the moral law. Kant's solution is to use the rites and statutes of the historical churches in order to gradually train people into feeling respect for the moral law. Gradually, as people come to feel respect for the law symbolized through contingent religious rites, they will be able to leave the influence of the churches behind. Whether this process is completed (or ever can be completed) in the *Religion* and what are the implications of Kant's solution on his notion of respect will be the focus of my analysis of the *Religion*.

The thesis will begin with an analysis of Hume's explanation in the *Treatise* of the artificial virtue of justice and the institution of promising. It will be argued that the movement of Hume's historical argument is to lead the will in order to gradually feel motivated by the virtue of justice itself. With the results of Hume's argument in the background, the thesis will shift to an analysis of Kant's notion of respect. Once an interpretation of the relation between reason and the moral law has been established in chapter two, I will then move on in chapters three and four to an analysis of the problem of and solution to radical evil in the *Religion*. In the conclusion, it will be seen that the solution to radical evil in the *Religion* is similar to Hume's argument concerning the gradual formation of the artificial virtue of justice.

Several avenues for further research will result from this analysis. First, the dichotomy between empiricism and Kantianism should perhaps be explored from a different angle: if Kant requires the use of the historical churches in order to educate people into feeling respect for the moral law, then it is not so certain that the moral law acts as a sufficient motive on the will. And if people do not respect the *a priori* law except through the use of historical institutions that slowly teach this respect, then is Kant proposing a moral theory that, on the question of motivation, is substantially different from that of the empiricist Hume? Second, it should be explored whether Kant, in the historical works, ever achieves the ideal of duty for duty's sake. In other words, if the moral training offered by historical institutions is not overcome, then can the autonomy of heeding the moral law *for the sake of the law* be realistically achieved? Finally, if Kant has nonetheless noticed an important problem in empiricism for metaphysics, ethics, and religion despite his inadequate solution, then it would be worth exploring the key empiricist assumptions regarding nature

and metaphysics that prompted the critique of causation and the Kantian reply in the first place. These larger questions could not be directly explored in these pages, which is why the thesis reads like a propaedeutic setting the stage for further research into foundational questions of modern philosophy and ethics.

## **Chapter 1: The Motive behind Justice and the Problem of the Sensible Knave**

The difference between Hume and Kant on the issue of moral motivation appears obvious. Kant defines the moral act as one motivated solely by respect for the moral law. Hume, on the other hand, maintains that virtue and vice are labels attached to qualities and characters of human beings that are felt as pleasing or painful in experience. Between the two philosophers, therefore, lies an immeasurable gulf: Kant sets his ethics on the wings of the *a priori*; Hume locates virtue and vice on the hard rock of experience.

This depiction of the difference between the moral theories of Kant and Hume is correct insofar as the objective criteria for evaluating a moral act differ from each other. It will be seen, however, that their depiction of the subjective motivation for moral action paints a similar picture of social formation, making the Scottish empiricist and the Prussian innovator of transcendental philosophy stand on similar ground. I will begin the story with Hume.

### **Two strains of interpretation on the question of moral motivation**

#### *Evaluating Moral Motives*

When evaluating the virtue or vice of an act, the object of investigation cannot be the act itself; for thinking an act virtuous because of the act's virtue is to assume what needs to be proven. It is necessary, therefore, to separate the motive from the act in order to determine whether or not the act was performed out of a laudable motive or intention. To argue otherwise would result in a logical circle:

To suppose, that the mere regard to the virtue of the action may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle. Before we can have such a regard, the action must really be virtuous; and this virtue must be deriv'd from some virtuous motive: And consequently the virtuous motive must

be different from the regard to the virtue of the action. (*Treatise* 339-341).

This passage makes it clear that settling the difference between a virtuous or vicious act depends on the motive that moves the will to act. Hume's claim that the motive exciting action cannot come from reason is also clear: since reason only concerns matter of fact and the relation of ideas—neither of which excite the passions—then the only other impression acting on the passions must be those that are felt, i.e. the impressions of pleasure and pain.

Hume's argument that the motive for moral action comes from feeling and not from reason is laid out in three short sentences: "morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason" (*Treatise* 325). In deciding whether moral concepts are created by impressions or the relation of ideas, it is important to start at the first premise of Hume's argument: moral rules and ideas of virtue and vice influence the passions and actions; if this were not the case, then no one would exert plenty of effort, through education, to communicate moral concepts in the hopes that students will direct their behaviour toward what they should regard as good and away from what they should regard as bad (*Treatise* 325). The act of reasoning, however, is only performed on matters of fact or the relations between ideas, none of which influence the passions (*Treatise* 326). Matters of fact concern the existence of an object or the existence of a quality in an object (*Treatise* 326-27). Such judgments, however, do not constitute morality in and of themselves. It is not a matter of morality for reason to judge that a stone is currently placed in front of my feet and that this, hard as it is, will hurt if I kick it. The existence and hardness of the stone is neither moral nor immoral—they are simply matters of fact.

Likewise, the relation between ideas is neither moral nor immoral. To use Hume's example: that a young tree should grow to overshadow and hence destroy its parent can hardly be considered an act of murder, even though a relation of kin exists between the older tree and the young tree that sprouted from its seed (*Treatise* 332). If matters of fact or the relations of ideas were the products of moral judgment, then any natural object or natural relation between objects would be evaluated morally; judgments of this sort, however, are simply not made by human beings. The most that reason can do in determining moral or immoral acts is in alerting the passions to the existence of a pleasant or painful object, or in pointing out the means for obtaining a pleasant object and for avoiding a painful one (*Treatise* 327). Since the argument has proven that the ideas of reason do not influence the passions, then the only perception left for influencing the passions are the impressions, which do not directly influence the mind through the activity of reason.

The impressions influencing the passions in moral or immoral motives, therefore, are impressions of pleasure or pain. Virtue and vice, in other words, are words used to describe our approval upon observing qualities or character<sup>1</sup> conducive to pleasure or pain. Experience shows that people judge the impression belonging to virtue as pleasant, while the impressions belonging to vice as disagreeable: "An action, or sentiment, or character is virtuous or vicious; why? because its view causes a pleasure or uneasiness of a particular kind. In giving a reason, therefore, for the pleasure or uneasiness, we sufficiently explain the vice or virtue" (*Treatise* 335).

---

<sup>1</sup>It is important to remember that the labels of virtue or vice are attached to characters and qualities in a person, since the motive—and not the act itself—is responsible for the moral quality of the act.

Since it has been established that evaluating the motive of an act depends on the pleasure or pain produced by the perception of the motive, it is necessary to focus on the task Hume sets himself at the end of the last passage cited above: *why* does the perception of certain actions, sentiments, and characters produce pleasure, while others produce pain? The answer to this question creates a set of interpretive difficulties that make Hume's claims regarding the relation between motive, on the one hand, and pleasure and pain, on the other, more difficult to take at face value. This problem becomes clearer when interpreting the virtue of justice.

Justice is widely considered a virtue, yet self-interested people do not seem *naturally* inclined to act justly. If a person is led by natural instinct to regard his own interests ahead of those of his fellow citizens, then what could possibly lead him to respect the conventions of justice, which involves refraining from taking another person's property, or binding his will through a promise? The rules of justice, which exist in every society, seem inexplicable through appeal to a natural motive, for there does not appear to be any such motive leading a person to observe the rules of equity.<sup>2</sup> Hume discovers the motive for observing justice by eliminating its natural component: the motive for justice results from artificial convention.

Human beings are naturally selfish, so are "not easily induc'd to perform any action for the interest of strangers, except with a view to some reciprocal advantage" (*Treatise* 370). The riddle of promising, however, can be solved through the "reciprocal advantage" that this "mere form of words" offers. The development of a motive for promising runs parallel to the development of the respect for property. Though human beings are naturally

---

<sup>2</sup> Hume makes the same observation with respect to promises, arguing that human beings do not manifest any natural motive for observing promises (*Treatise* 369).

selfish, they are physically weak compared to other animal species. Nature, furthermore, does not provide enough natural goods to fully satisfy human desire. The need to enter society results from what have been called the “circumstances of justice.”<sup>3</sup> The combined *limited* self-interest and the scarcity of natural goods makes society and the convention of justice *useful* to human beings:

[human conventions] are intended as a rem'dy to some inconveniences, which proceed from the concurrence of certain *qualities* of the human mind with the *situation* of external objects. The qualities of the mind are *selfishness* and *limited generosity*: And the situation of external objects is their *easy change*, join'd to their *scarcity* in comparison of the wants and desires of men. (*Treatise* 351)

Hume offers the same story with respect to promises. Hume writes that, were we “to follow the natural course of our passions and inclinations, we shou'd perform but few actions for the advantage of others, from disinterested views; because we are naturally very limited in our kindness and affection” (*Treatise* 370). Distrust of fellow human beings dissolves the society needed for our welfare; without the assurance that people will keep their word, the benefits of society through an exchange of goods and services would be “lost among mankind, and every one reduc'd to his own skill and industry for his well-being and subsistence” (*Treatise* 370).

As the invention of rules for the stability and transfer of property makes society possible, so rules for the transfer of property makes life in common beneficial. Rules dictating the transfer of property, however, only permit exchange of specific goods between close neighbours: the rights of occupation, prescription, accession, and succession all concern either the possession of goods worked upon by an individual, or the transfer of

---

<sup>3</sup> See Martha Nussbaum, *The Frontiers of Justice* (26-28) and John Rawls, *A Theory of Justice* (§22).

specific goods among close relatives and friends (see *Treatise* III, Part II, section III). The convention of promising remedies the limitations of the initial rules by permitting the transfer of general objects over long distances (*Treatise* 371). Without the ability to trust that a stranger living in a distant country will provide payment once I have shipped him a specific set of goods, I would never take the risk of investing money to supply him with those goods. Distrust foils the purpose of society, which lies in providing the goods and services that human beings could not obtain on their own in a natural condition of solitude and weakness. As in the case of justice, the solution to the impairment of the transfer of goods by the obstacle of self-interest lies in making promising appear useful for better obtaining the goods and services desired.

*Interpretations emphasizing Self-Interest as the Motive behind Justice and Promising*

The emphasis on mutual advantage as the spur for the creation of promising naturally leads some interpreters to claim that the real motive in moral matters remains the satisfaction of desire and happiness; justice and promising would not be motivating in and of themselves. Annette Baier, for instance, argues that each artificial virtue rests on “enlightened, cooperative, inventive, and ‘oblique’ self-interest” (Baier 41). Enlightened self-interest should not be considered a synonym for selfishness; the self-interested passion, rather, “acts cooperatively and in coordination with others” (Baier 44). The motive to justice, however, remains self-interest. Likewise, Baier argues that the motive for performing a promise is “enlightened egoism, since [Hume] requires that for any social convention to give rise to an

obligation each person must find himself (or herself?) ‘a gainer, on balancing the account’” (Baier 165).<sup>4</sup>

The interpretation provided by David Gauthier emphasizes the role of self-interest in the origin of justice and promising, as well as in their continuance. Gauthier rightly observes that, though justice requires a constraint on self-interest, the motive to adopt the constraint lies in better satisfying self-interest through life in society: “although adherence to these conventions requires each person to restrain her interest, this restraint is mandated by the interest itself, redirected by reflection on the greater satisfaction that it attains by maintaining the conventions” (Gauthier 414). It is important to note that, in Gauthier’s analysis, self-interest acts as the motive even once justice and fidelity to promises come to be regarded as virtuous in and of themselves. The desire to satisfy self-interest, however, does not fall victim to the objection of the sensible knave because a person will recognize that following the rules of justice and keeping one’s promises are necessary for remaining a trusted member of the community, which aids in better satisfying self-interest (Gauthier 424). Hume clearly argues that, in using the form of words necessary for making a promise, a person “subjects himself to the penalty of never being trusted again in case of failure” (*Treatise* 372). Though the penalty of ostracism will lead a person to avoid committing harmless acts of injustice beneficial to himself, the avoidance of penalty is nonetheless motivated by the interested desire to remain part of the community that best allows the satisfaction of self-interest: “he

---

<sup>4</sup> Annette Baier, for instance, observes that the threat of punishment constitutes another motive to the performance of promises: “in the story Hume gives us, threat is an essential element in promise, and it takes community cooperation in the relevant convention or practice to pose that threat of destruction of one’s power to ever have one’s promises accepted....Hume here quite clearly refers both to the community interest in giving each other security by ‘instituting’ these signs, with backup penalty of withdrawal of trust from those who take them lightly, and to the individuals promisor’s interest in performing what he promised, so as to remain in good standing, to keep his reputation as a trustworthy person” (Baier 182).

will expect these real costs to be outweighed by the benefits he receives in the form of opportunities that were he not disposed to the artificial virtues, he would not enjoy” (Gauthier 424).<sup>5</sup> The result is that, even in their civilized state, in which justice and fidelity to promises are regarded as virtues, human beings feel motivated to maintain these artificial conventions through the pull of self-interest.

Finally, Stephen Kalt observes the same self-interested motive as the basis for justice and promising. In analyzing the “self-hatred” passage,<sup>6</sup> in which Hume claims that a person who lacks a virtue may nevertheless perform it without the requisite motive, Kalt claims that the motive for performing the virtue nevertheless remains the desire for happiness. Pace Radcliffe, self-hatred, for Kalt, is not a species of the direct passions; it is, rather, a form of self-disapprobation, which belongs to indirect passions. Self-disapprobation, however, triggers a desire to avoid pain, which itself is a direct passion. The indirect passion of self-disapprobation, in other words, leads to the direct passion of aversion (to pain), which is simply “the flip side of, and thus identical to, the desire for happiness” (Kalt 153). On Kalt’s interpretation, the motive leading to the performance of duty turns out to be the external motive of happiness. Justice and promising, therefore, which depend upon our feeling an obligation to their performance, are really motivated by the desire for happiness, which is external to the perception of duty itself.

---

<sup>5</sup> A similar argument is presented by Éléonore Le Jallé: the fear of being ostracized from society and its useful system of conventions is enough to keep people from breaking their promises. However, Le Jallé emphasizes that it is the conventions themselves and their threat that create the self-interested motive to keep a promise; promising, in other words, is not maintained by a self-interested motive existing prior to social convention (Le Jallé 239).

<sup>6</sup> *Treatise*, p. 340

*Problems with Self-Interest: The Sensible Knave and Interpretations favouring Motivation by Justice Itself*

The rules of justice, however, are merely general; that is, they create greater benefit for people *most of the time* (*Treatise* 353). Equity opens the door to civilized and peaceful society, but individual judgments of justice may in fact be detrimental to the person involved. A clever financier may be able to avoid full payment to his creditor without the creditor's full knowledge; no one would be aware of the breach of justice and perhaps no one would be directly harmed. Fully repaying the loan would deprive the financier of what he could have received had he neglected the demands of justice.

Why, then, would a self-interested person feel motivated by the conventions of justice? Hume himself appears to have noticed the problem near the end of the *Enquiry Concerning the Principles of Morals* when he turns to the case of the “sensible knave”, who incarnates the tension between justice and self-interest. The knave is concerned above all with his own interest, but he notices that the rules of justice constrain him to act against his interest, since refraining from malevolently acquiring the property of others deprives him of goods he may enjoy. The reasoning of the knave should lead him to disregard the conventions of justice when the opportunity presents itself, leaving Hume with the trouble of finding an answer to the logical difficulty created by the philosophical foundations of his ethics:

And though it is allowed that, without regard to property, no society could subsist; yet according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That *honesty is the best policy*, may be a good general rule, but is liable to many exceptions; and he, it may perhaps be thought, conducts himself with most wisdom, who observes the general rule, and

takes advantage of all the exceptions (*EPM* 121-122).

Though Hume attempts an answer to the problem of the sensible knave in the couple paragraphs following this passage, he regarded the answer as unsatisfactory.<sup>7</sup>

The character of the sensible knave points to a tension between self-interest and the motivation to act justly, a tension that some interpreters exploit in order to claim that Hume could not have consistently maintained that pleasure and pain are the sole motives for acting justly. The distinction between natural and artificial virtue, in other words, leads to a distinction in the motives for moral action: since justice, at least at certain times, requires the sacrifice of a self-interested motive, a different motive must be at play in the case of acting justly. A certain number of interpreters, therefore, have claimed that the logic of the artificial virtues commits Hume to the position that it is the rightness of the act that motivates the agent to act justly. In reference to Hume's discussion of the sensible knave, Dorothy Coleman argues that, when it comes to the virtue of justice, self-interest cannot remain the sole motive for action, since the rules of justice and the pull of self-interest conflict with each other: "although the original motive for establishing a system of justice is self-interest, self-interest cannot explain the motivation to obey rules of justice in every instance, since in some circumstances they may work against self-interest" (Coleman 338). Likewise, Stephen Darwall, in observing that Hume does not directly answer the objection posed by the

---

<sup>7</sup> Stephen Darwall argues that Hume could not have considered his reply to the sensible knave as satisfactory because his reply does not defend the claim that respect for the rules of justice is the "best general policy" (Darwall, "Motive and Obligation in Hume's Ethics", p. 434). Darwall takes Hume's strategy, on the one hand, as the sign of the failure of an externalist motivation to justice via self-interest and, on the other hand, as an invitation to recognize that the problem of the sensible knave demands a commitment to the claim that, in civilized societies, the rules of justice must eventually become motivating in and of themselves.

sensible knave,<sup>8</sup> concludes that Hume must have eventually come to regard the rules of justice as inherently motivating. Though Hume may have considered in the *Treatise* that “self-interest invariably dictates justice, by the time of the *Enquiry* he has apparently abandoned this view” (Darwall, “Motive and Obligation in Hume’s Ethics”, p. 435). To solve this problem in the *Enquiry*, Darwall claims that Hume considers justice to consist of “regulation by rules”, which means that “just persons regard the rules *internally* as agents” (Darwall, “Motive and Obligation in Hume’s Ethics”, p. 436).

Interpreters<sup>9</sup> advocating motivation by the rules of justice themselves resist the passages separating motives from just acts by referring to a short passage in which Hume claims that an individual, lacking the proper motive for a moral act, may still fulfill the act despite the lack of a worthy motive:

When any virtuous motive or principle is common in human nature, a person, who feels his heart devoid of that motive, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it. (*Treatise* 340).

If an individual can still perform a just act without an anterior motive, then it appears as though the act could be motivating simply for its own sake. The consequences of this possibility would be important, since it would imply that motivation for performing artificial virtue does not always require self-interest. It is hence not surprising that Stephen Kalt’s interpretation of the motivation for justice begins with an analysis of the “self-hatred”

---

<sup>8</sup> See Darwall, “Motive and Obligation in Hume’s Ethics”, p. 434.

<sup>9</sup> Though I have only referred to Coleman and Darwall up to this point, I will refer to similar interpretation in the last section of this chapter.

passage in order to reject the interpretation that finds support in this passage for motivation by the rules of justice themselves.

Despite Kalt's rejection of this passage for claims that justice is inherently motivating, it will be worth exploring this clue a little further. It will be seen that the self-hatred passage stands as a clue that opens the door toward a second level of motivation; the first level indeed rests on self-interest—as many interpreters note—but the second level trains people to regard justice as motivating in and of itself. The remainder of this chapter will focus on analyzing the second stage of moral motivation. It will be seen that interpretations emphasizing self-interest as the motive for observing justice are not wrong; rather, they overlook the movement in Hume's text that surpasses self-interest by positing the rules of justice as motivating in and of themselves.

### **Second Stage in the Development of the Virtue of Justice: Learning how to be Motivated by Justice Itself.**

#### *Stage 2: Learning how to be Motivated by Justice Itself*

##### *a) The Limits of Sympathy*

As society grows larger, the encouragement provided by self-interest for maintaining the rules of justice becomes fainter. Accepting rules that constrain self-interest requires vision: people must perceive the advantages of property and promising in order to feel a motive to conform to these conventions. The harm committed by infringing on property rights for the sake of satisfying self-interest is apparent in smaller societies, such as hamlets or villages. The greater number of variables composing large societies, however, renders the effects of harm resulting from injustice more difficult to calculate. The result is that, in a

larger society, human beings do not “so readily perceive, that disorder and confusion follow upon every breach of these rules, as in a more narrow and contracted society” (*Treatise* 355).

The problem lies in the fact that self-interest is focused on a particular: the self. The growth of society, however, requires a more universal virtue to accord with the general interests of society. Though the mind cannot develop universal, *a priori* moral concepts—since its knowledge is built upon particular sense impressions—the feeling of sympathy extends the feelings of the subject beyond himself, toward most other members of society. Sympathy acts on the contiguity of ideas: the pleasure or pain felt by another person is close to my own because we both share similar characteristics (*Treatise* 355). The feeling of sympathy is beneficial in two ways. First, the feeling received from “contemplating the actions of others” is extended “even to our own actions” (*Treatise* 355). Second, our feeling of pleasure or pain induced by the pleasure or pain of others leads us to perceive justice as having a *moral* obligation:<sup>10</sup>

The *general rule* reaches beyond those instances, from which it arose; while at the same time we naturally *sympathize* with others in the sentiments they entertain of us. Thus self-interest is the original motive to the establishment of justice: but a sympathy with the public interest is the source of the moral approbation, which attends that virtue (*Treatise* 355).

Hume calls this movement of the sentiments from self-interest to sympathy “natural, and even necessary” (*Treatise* 355). Sympathy encourages motivation for the *general rule* of justice through our perception of the negative effects of injustice. A breach of justice

---

<sup>10</sup> Knud Haakonssen claims that, when we come to see the benefits of promise-keeping and sympathize with its good effects, then “the natural tendency to see behaviour as motive and motive as the result of good qualities of character leads them to imagine the existence of a *natural* motive, i.e. a natural obligation behind promises” (Haakonssen 13). Though Haakonssen’s argument appears to be the opposite of mine, it is important to note that people *imagine* the existence of a natural obligation to keep promises. Conventions make promise-keeping an artificial virtue; it is only on appearance that promise-keeping seems to be natural.

displeases in and of itself, since we observe and sympathize with the negative effects of injustice.

Even though sympathy acts as a bridge between the interests of a particular individual and the state of others in society, its power to bridge common feelings dissipates as two objects stand in less resemblance and at a greater distance. The sensible knave may still rationally conclude that he would be happier at acting unjustly in cases where he (mistakenly) believes there would be no negative effect. The virtue of justice must come to be motivating in and of itself once society has become large enough to mask the effects of injustice, a need well articulated by Sharon Krause.

In analyzing the relation between norms and motives in the *Treatise*, Krause notices that sympathy cannot remedy the particularity of self-interest because the feeling of sympathy dissipates once the resemblance and contiguity between ideas becomes greater: “But the communication loses its force the wider it travels. Hume is quite clear in the *Treatise* about the limits of sympathy in this regard. The sentiments of others, he says, ‘have little influence, when far remov’d from us, and require the relation of contiguity, to make them communicate themselves entirely’” (cited in Krause 641). Since the limited range of sympathy reduces the motivation to act justly in large societies, Krause concludes that Hume’s claim that sympathy makes of justice an obligation must remain doubtful: “the limits of sympathy compound the motivational gap associated with the limits of utility, and they further exacerbate the uncertain status of justice as an obligation” (Krause 642).<sup>11</sup>

---

<sup>11</sup> See Le Jallé for a different interpretation: sympathy is *always* present throughout both stages of Hume’s ethics. Sympathy leads us to feel the harm caused to another person as a result of injustice and sympathy is also active in producing the conventions of morality, since we feel the beneficial result of good actions (Le Jallé

Though Krause casts doubt on the status of justice as obligatory in itself, it is not so certain that the rules of justice cannot be motivating in and of themselves. An argument by Stephen Darwall can represent the starting point for examining how, in the second stage of the development of justice, the rules of justice come to be motivating in and of themselves.

Darwall observes that Hume's reply to the sensible knave is unsatisfactory. Since the knave is not motivated by justice, he judges that he can avoid the demands of justice when the act will not cause great harm to others and will benefit his self-interest. If Hume thought that the observance of the rules of justice always resulted in beneficial consequences, then he would simply have pointed out to the knave that committing injustice is actually harmful to the knave himself (Darwall 434). Hume, however, does not take this route; he despairs, in fact, of being able to adequately provide an answer for the knave: "if his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy and baseness, he has indeed lost a considerable motive to virtue" (*EPM* 283; cited from Darwall 434). The only way to answer the knave and to maintain the convention of justice in a larger and larger society is to regard justice itself as motivating. Since self-interest cannot provide a sure-fire motive for acting justly, Darwall claims that, by the time he wrote the *Enquiry*, Hume had begun to embrace the view that the rules of justice must be inherently motivating:

Since [Hume] cannot now think that rule-regulation is reducible to the pursuit of interests in maintaining rule-structured practices viewed *externally*, Hume's continuing talk of acceptance of...and regulation by rules must now be taken seriously. And this requires interpreting Hume as holding that just persons regard the rules internally as agents....Just persons acquire the "habit of justice" (E.203).

---

239). I do not entirely agree with Le Jallé's argument because, though sympathy remains integral to the extension of justice in society, the feeling of sympathy remains restricted by greater distances separating individuals in larger societies. Eventually, the rules of justice must come to be motivating in and of themselves, and not primarily through the motive provided by sympathy.

They are determined to follow the rules of justice, and although their original motive for being so determined is self-interest, they continue “without recalling, on every occasion, the reflections, which determined [them].” (E. 203) (Darwall 436).

Self-interest can only represent a first-stage, a transition from the foundation of justice to acting on justice for its own sake. Though human beings always feel the pull of self-interest, the genius of civil society lies in making justice more “interesting” than self-interest. The satisfaction of self-interest may have provided the motive for the creation of justice; the success and continuation of this artificial creation, however, demands seeing justice itself as inviolable.<sup>12</sup>

b) *Justice as Motive*

The way in which Hume introduces the problem of justice at the outset of Book III, Part II, indicates that the goal of the section is to explain how justice comes to be regarded as a motive for moving the will. Right from the start of his analysis, Hume develops a contrast between the virtue of justice for humans in their civilized state and the same virtue for humans in their natural state. The contrast comes out clearly in the passage that introduces the problem that Hume deals with in the rest of part II:

I suppose a person to have lent me a sum of money, on condition that it be restor'd in a few days; and also suppose, that after the expiration of the term agreed on, he demands the sum: I ask, *What reason or motive have I to restore the money?* It will, perhaps, be said, that my regard to justice, and abhorrence of villainy and knavery, are sufficient reasons for me, if I have the least grain of honesty, or sense of duty and obligation. And this answer, no doubt, is just and satisfactory to man in his

---

<sup>12</sup> J.L. Mackie argues that, in saying that the extension of justice to others relies on sympathy, Hume implicitly admits that justice is, in part, a natural virtue (Mackie 85). If Mackie's argument were correct, then my claim that a resort to moral education is necessary would turn out to be untrue, since sympathy would guarantee the success of artificial convention. However, even if sympathy is still active within a web of artificial convention, the distance separating people in large societies mitigates the influence of sympathy, as Krause noted in her article. As I will argue, as people develop a sense of obligation through moral training, they come to be motivated by the rules of justice themselves. At this point, sympathy is no longer the main element driving moral action.

civiliz'd state, and when train'd up according to a certain discipline and education. But in his rude and more *natural* condition, if you are pleas'd to call such a condition natural, this answer wou'd be rejected as perfectly unintelligible and sophistical. (*Treatise* 341).

The answer to the initial question in this passage assumes that justice is inherently motivating, since he expects the return of borrowed money out of a sense of duty itself. Hume's correction of the erroneous reply lies not in pointing out that the motivation for justice lies elsewhere, but that justice is only motivating *as a result of* a process of socialization: "and this answer, no doubt, is just and satisfactory to man *in his civiliz'd state*" (*Treatise* 341). The first step in the process of making a virtue out of justice is self-interest. Sympathy, however, begins the transition to the second stage—regarding justice as a virtue and hence motivating in and of itself. That process, however, comes to completion only once the rules of justice become inherently motivating.

In his search for a natural motive to justice, Hume rejects "public interest" because it is "only connected with [justice], after an artificial convention for the establishment of these rules" has been created (*Treatise* 341). Self-interest and sympathy led to the creation of these rules, first in small societies and later in larger ones. The work of educators and politicians, however, builds up the natural transition from self-interest to sympathy: "any artifice of politicians may assist nature in the producing of those sentiments, which she suggests to us, and may even on some occasions, produce alone an approbation or esteem for any particular action" (*Treatise* 355). The educational work of politicians, however, does not replace nature, for a distinction between vice and virtue must already be in place to which politicians can appeal for the content of their discourse to be considered salutary. The "publick praise and blame" of justice and injustice provided by politicians serves to "increase our esteem for

justice” (*Treatise* 356). Private education, provided by parents for their children, furthers the same purpose. Parents observe the utility a sense of honour offers to members of their society, and so they teach their children “to regard the observance” of the rules of justice “as worthy and honourable, and their violation as base and infamous” (*Treatise* 356). Hume argues that we are trained to regard the rules of justice as motivating out of an “interest of our reputation”: “there is nothing, which touches us more nearly than our reputation, and nothing on which our reputation more depends than our conduct, with relation to the property of others” (*Treatise* 356).

It may perhaps be argued that the ‘sentiment of morals’ established through education really depends on a desire to increase our reputation, which is only a particular manifestation of self-interest. Though it is certainly true that reputation is related to a self-interested motive, the increase of reputation depends upon regarding the rules of justice as unbreakable in and of themselves: “every one, who has any regard to his character, or who intends to live on good terms with mankind, must fix an inviolable law to himself, never, by any temptation, to be induc’d to violate those principles, which are essential to a man of probity and honour” (*Treatise* 356). Hume notices that parents regard justice and honour as *useful* for living in society. The utility of justice, however, makes the rules of justice pleasing in and of themselves, since we naturally regard any quality or character as pleasant because of its utility, as Hume had argued at the outset of Book III. It is thus not necessary, in a civilized state, to regard justice as pleasing *because of* the satisfaction of self-interest. Hume’s story is meant to show how such pleasure has become attached to justice by showing the utility that justice holds for society. The genius of society in creating artificial virtue is that a person becomes trained to regard justice as a virtue and hence as motivating

without the need to consciously seek out the utility of justice and certainly not at each moment. Moral philosophers seek the foundations of justice and find it in utility; the artifice of politicians and the work of educators ensure that not everyone has to become a philosopher in order to find a motive for acting justly.

c) *Interpretations of Justice as Motive*

The significance of sympathy and education for an overall interpretation of the motivating factors for acts of justice has not gone unnoticed by commentators. Arguing against commentators who locate the motive for justice strictly in an enlightened form of self-interest, Rachel Cohon claims that self-interest acts as motivation only in the “early days”, in “small, face-to-face communities that offer few temptations or opportunities to be a free-rider” (Cohon 204). A second stage in the formation of a motive for justice is needed, however, in large societies where “the effects of any one violation of trust on the social fabric are more diluted and distant” (Cohon 204). The second stage, according to Cohon, coincides with the “sentiment of morals” that arises after the initial institution of justice and of promising:

Sympathy with the (actual or prospective) beneficiaries of fulfilled promises and victims of broken promises, which may amount to everyone in society, generates approval for rules of promise and for general compliance with them, and disapproval for violations. Parents, politicians, and others with influence over people’s feelings then make use of this approval and disapproval. In a “second artifice”, they use psychological manipulation (for good purposes) to produce a strong feeling of moral approval for every act of promise-keeping, and a strong feeling of moral disapproval for every act of promise-breaking, especially the agent’s own, by associating the moral approval and disapproval with feelings of pride and shame. (Cohon 205)

The result of this process of social formation is for children and adults to feel motives for performing just acts and keeping promises “regardless of whether the agents see advantage

in it or not” (Cohon 205). Justice and promising, therefore, no longer depend on a direct perception of the satisfaction of self-interest for people to regard them as virtuous and hence feel a motive to act according to their demands.

A similar interpretation is advanced by Páll Árdal, who claims that if the motive for the performance of justice were to remain solely based on the utility of justice, then the strength of justice could not be accounted for—the sensible knave could ignore the demands of justice when injustice proves to be more useful to himself (Árdal 186). The sense of duty is itself motivating and is developed through education:

[The just man] is able to [adhere steadfastly to the rules of justice] because the motive of duty has come to be so highly developed.... Propaganda and education help to create a situation where the non-performance of just actions comes to be firmly associated with pain in our minds. This helps to strengthen the behavioural pattern we call just behaviour (Árdal 187).<sup>13</sup>

The opposing pull between the outward-looking virtue of justice and the inward-looking motive of self-interest cannot be reconciled—it must be overcome. Justice, in other words, must become the motive in order to avoid the problem of the free-rider as expressed in the character of the sensible knave.

Charlotte Brown objects that Hume’s position only *looks* as though justice is motivating in and of itself (Brown 85). Brown claims that the regard for the rightness of an action “becomes practical only by triggering pride and the desire to be happy” (Brown 85-86). Brown concludes that “in both the natural and artificial case, the regard for the moral worth of the action motivates not in virtue of being moral but in virtue of its relation to pride

---

<sup>13</sup> See also Margaret Watkins Tate, who argues that, in their civilized state, human beings are motivated by the sense of justice itself (“Obligation, Justice, and the Will in Hume’s Moral Philosophy”, p. 114).

and the desire to be happy. This is not an internalist view” (Brown 86).<sup>14</sup> In reply to Brown’s argument, the sensible knave, however, is precisely the person who shirks the demands of justice when those demands promise to pass over the satisfaction of his own happiness.

Brown does not sufficiently take into account the fact that happiness and the satisfaction of self-interest only constitute the grounds of justice in small societies. Laying adequate emphasis on the distinction between the first and second stage in the formation of justice is crucial for finding the motive for justice in a “civilized state”.

Dorothy Coleman properly and rightly points out the difference between the first and second stage of justice in her reply to Brown:

Although the original motive for establishing a system of justice is self-interest, self-interest cannot explain the motivation to obey rules of justice in every instance, since in some circumstances they may work against self-interest. Hume maintains that once such a system is established, a person nevertheless may, in “his civiliz’d state, and when train’d up according to a certain discipline and education,” repay a debt simply from a “regard to justice,” or “sense of duty and obligation.” (Coleman 339)

We have thus come full circle. The problem that Hume started out with—why does a person in his civilized state regard a sense of duty as a motive for repaying a loan?—has now been answered. The development of duty and justice is initiated through the desire of self-interest. The regard that a just man holds for duty and justice in and of themselves, however, can only be explained as the end result of a process that quells the lure of self-interest through the influence of education and political rhetoric. The figure of the sensible knave and of the

---

<sup>14</sup> Gauthier recognizes that the artifice of politicians and educators can produce a motive for integrity and duty distinct from the motive arising from self-interest. However, like Brown, he is quick to point out that a seemingly disinterested virtue such as duty and justice rests on a motive rooted in the affections: “Only the peculiar working of the affections, in which the want of an imaginary motive proves to be the ground of a real one, can provide the materials needed for indoctrination to motivate us to be just” (Gauthier 420).

consequent disintegration of the social fabric demands that justice receive a motivating lustre of its own.<sup>15</sup>

d) *A Final Detail: making Room for Motivation outside of Justice.*

Though I have argued that Hume's account of the formation of the artificial virtues leads to unison between justice and motive, it is not true that an agent will find justice motivating *at all times*. Internal motivation is not a one-way street; Hume does not claim that a civilized, just person acts at all times from the motive of justice. Hume explicitly argues, for instance, that the convention of promising is partially enforced through the threat of exclusion resulting from infidelity to one's promises: "when a man says *he promises any thing*, he in effect expresses a *resolution* of performing it; and along with that, by making use of this *form of words*, subjects himself to the penalty of never being trusted again in case of failure" (*Treatise* 372). Annette Baier has noticed that, even in his civilized state, "threat is an essential element in promise": "Hume here quite clearly refers both to the community interest in giving each other security by 'instituting' these signs, with backup penalty of withdrawal of trust from those who take them lightly, and to the individual promisor's interest in performing what he promised, so as to remain in good standing, to keep his reputation as a trustworthy person" (Baier 182).<sup>16</sup>

---

<sup>15</sup> Anthony Pitson maintains a similar interpretation, though he claims that Hume himself did not think it possible. When it comes to promising, Pitson argues that Hume faces a puzzle as to how promising "can involve the voluntary undertaking of an obligation" (Pitson 188). The solution to the puzzle, however, requires recognizing that "what motivates the man of honour is neither interest nor displeasure at the non-performance of his promises, but that it is, rather, a recognition of the obligation under which he has placed himself" (Pitson 188). Pitson, however, claims that Hume dismisses this argument "as impossible" (Pitson 188). However, my preceding analysis of the stages of justice in the *Treatise* and of the problem of the sensible knave in the *Enquiry* suggests that this position is indeed held by Hume, though perhaps not as explicitly as it should have been.

<sup>16</sup> See also Le Jallé 227 for a similar interpretation on the importance of fear in creating a motive for keeping promises.

Though Baier's commitment to self-interest as the underlying motive of morality is not entirely correct<sup>17</sup> (for the reasons indicated in my analysis above), she does point out an important element in the text of the *Treatise* that suggests that the motive of fear can successfully motivate a person to fulfill a promise *when a promise is not found sufficiently motivating*. The ethical ideal remains one of finding motivation in the rules of justice themselves. Since human beings are imperfect, however, the actual movement of motivation in practice, consists of being led to regard justice as motivating, while being impelled by fear toward acts of justice in moments those rules are not felt to be motivating enough. Several motives, then, continually work on an agent; the goal, however, is for the motive of justice to take the lead.

The creation by artificial convention of a subjective motive for the performance of just acts provides a helpful framework for the study of Kant's response to the challenge of evil in *Religion within the Limits of Reason Alone*. By the time he sat down to write the *Religion*, Kant had realized that people do not always respect the moral law, even though the law exerts a positive influence on the will. Kant raises and tries to solve the problem of radical evil because the existence of acts motivated by the inclinations reveals a tendency in human nature to subordinate the influence of the moral law to the influence of contingent empirical inclinations. The way in which Kant proposes to re-subordinate the inclinations to the moral law involves a process of moral education similar to Hume's proposal that the will needs to be led to regard the rules of justice as motivating in and of themselves.

---

<sup>17</sup> On the motive of self-interest as constantly running through Hume's ethics, see Baier 41.

Examining the problem of radical evil in the *Religion* will open the door to exploring an element of Kant's ethics often overlooked by interpreters: the moral education required to lead a person to be motivated by the moral law alone. Before diving head-first into an analysis of Kant's *Religion*, it will be necessary to first critically analyze an assumption regarding the relation between the moral law and reason. The result of the analysis in the next chapter will set the stage for the remaining two chapters on the *Religion*.

## Chapter 2: Moral Law and Practical Reason

### The Distinction between Morality and Legality through Respect

Respect is an important concept in Kant's ethical works because it illustrates the human being's particular relation to the moral law. The human being, in Kant's technical lexicon, is a finite rational being; the classification is further illustrated by Kant's repeated comparison of human beings with God, on the one hand, and animals, on the other. God,<sup>1</sup> for Kant, is a rational being bereft of sensible nature; nothing in the divine nature can oppose God's reason in its legislation of the moral law. Animals, on the other hand, are sensible beings devoid of rationality; they act on impulse according to the determined causal order of *phenomena*, without the capacity of a rational nature to autonomously (hence freely) legislate the moral law to the will (*KpV* 5: 61-62). The human being stands between divinity and animality: though rational and hence capable of autonomously determining the will to act, the human being also possesses a sensible nature that does not necessarily follow the command of the law of reason.<sup>2</sup> Unlike animals, the will of a human being is able to be *objectively* determined by the moral law through the formulations of the categorical imperative—making the maxims of one's actions accord with universal legislation. Unlike God, the will of a human being is able to be *subjectively* determined by motives *other* than the moral law itself—for reasons of happiness, fear, or other motives that promise to satisfy the inclinations.

---

<sup>1</sup>In Kant's critical and post-critical writings, God is not a being that can be known or proven. When referring to God, Kant refers to the idea of a divine being; the question of whether this being exists or not surpasses the capacity of theoretical reason.

<sup>2</sup>In an important passage in the *Grounding*, Kant claims that only a "rational being has the power to act according to his conception of laws" (*G* 4: 412). The human being, however, is subject to incentives, which means that its "will does not in itself completely accord with reason" (*G* 4: 413).

Respect is the feeling through which the moral law subjectively determines the will. The moral law itself must determine the will; for, otherwise, “if the determination of the will, although occurring *in conformity with* the moral law, does so only by means of a feeling—of whatever kind—that must be presupposed in order for that law to become a sufficient determining basis of the will... then the action will indeed contain *legality*, but not *morality*” (*KpV* 5: 71). Respect is a feeling capable of affecting a sensibly inclined being, yet it is *a priori* since it is cognized only through the idea of a law that furthers its own causality by impairing or even ‘striking down’ any inclination that opposes the command of the moral law (*KpV* 5: 73). Respect is thus a feeling that applies only to finite rational beings whose sensible nature can turn their subjective impulses into an objective determination of the will.

For an act to be considered moral—not merely *legal*—the will must be determined both objectively and subjectively by the moral law; in other words, the maxim of the will must objectively accord with the universality of the moral law, while the subjective motive incorporated into the maxim of the will must be based on the law itself (see *KpV* 5: 72, 81). The model of ethics presented in Kant’s critical and post-critical writings is one whereby the will is determined solely by the moral law. Respect for the law marks the difference between legality (conformity to law—*pflichtmässig*) and morality (acting *for the sake of* the law—*aus Pflicht*):

the concept of duty demands *objectively*—in the action—agreement with the law, and *subjectively*—in the maxim of the action—respect for the law, as the sole way of determining the will by the law. And thereon rests the distinction between the consciousness of having acted *in conformity with duty* [*pflichtmässig*] and *from duty* [*aus Pflicht*], i.e., from respect for the law. The first of these (legality) is possible even if only inclinations were to have been the determining basis of the will; but the

second (*morality*), moral worth, must be posited solely in this, that the action is done from duty, i.e. only on account of the law. (*KpV* 5: 81).<sup>3</sup>

Human beings are situated in life, in a historical context. Affected as they are by sensible desires and inclinations, human beings can make those inclinations (often unwittingly) the supreme condition of moral action.<sup>4</sup> Human beings frequently conform the maxim of their action to the universal command of the moral law through motives *other* than the moral law itself, such as happiness. Yet, the content of happiness—or any other emotion—is contingent, since experience shows that not all human beings agree on what constitutes happiness: “the concept of happiness is such an indeterminate one that even though everyone wishes to attain happiness, yet he can never say definitively and consistently what it is that he really wishes and wills” (*G* 4: 418). Since the morality of an action is determined by the will’s conformity to the moral law, then respect, through which the universal moral law is the sole motive of action, is the criterion through which legality and morality may be distinguished.

The causality of the moral law through the *a priori* feeling of respect nonetheless remains enigmatic. Kant immediately avoids the question of *how* the moral law can subjectively determine the will through the feeling of respect, since the problem of *how* is one that belongs to—and surpasses—the capacity of theoretical reason (*KpV* 5: 72). Kant is left with the idea *that* the law should determine the will subjectively through the feeling of respect, which he claims is the only feeling that can be attributed to the moral law *a priori*:

---

<sup>3</sup> Two other passages present similar discussions of respect as the subjective determining cause of the will: Kant’s note on respect in *G* 4: 402 and making the moral law the supreme *life-principle* in *KpV* 5: 86.

<sup>4</sup> Kant critiques traditional ethical philosophies on the claim that they have made the contingent inclinations the ground of morality. Whether it be happiness, the will of God, or any other contingent principle, traditional ethics makes the mistake of establishing the criterion of morality on contingent principles that cannot be accepted by everyone. For Kant’s critique of traditional ethics, see *KpV* 5: 58-65. For Kant’s argument that all of traditional ethics is established on a principle of heteronomy, see *G* 4: 441-444.

“For how a law can by itself and directly be a determining basis of the will (this is, after all, what is essential in all morality) is an insoluble problem for human reason” (*KpV* 5: 72).

Yet, even if the interpreter remains within the limits of Kant’s analysis, it remains difficult to see *that* law may inspire respect. When commenting on Kant’s argument concerning respect in the Second Critique, Lewis White Beck observes that:

in spite of what Kant says, the law itself is not the incentive. A law is just not the sort of thing that can be an incentive. At most, consciousness of a law can be an incentive. If the law itself were a determinant of conduct, without the intervention of consciousness (which means, for us men, also feeling), it would not be a practical law, and men would not be free agents (Beck 221-222).

Philip Stratton-Lake tries to find a solution by arguing that Beck should not conceive of the law as a kind of force and hence should not be concerned that our freedom would be removed by the moral law:

if the moral law were conceived as a motive in this sense [as a force], then it would not be a practical law at all, but would be something like a law of nature... But we also think of motives not as forces pushing us along, but simply as reasons we have for doing certain acts (where these reasons are not understood causally) (Stratton-Lake 39. My emphasis).

Beck’s bafflement at how the moral law can be an incentive leads him to emphasize that only “consciousness of law” can provide a motive to act in favour of the law. In similar fashion, Stratton-Lake resorts to an interpretation of respect as a feeling that arises as the result of one’s autonomously deciding on a cause of action and hence feeling respect for the reasons behind the decision. Yet, the strategy of Beck or Stratton-Lake only pushes the problem one step further back: why should a person feel motivated to make the moral law—or the “consciousness” of the moral law—one’s reason for acting?

Whether respect consists of esteem for one’s reasons for acting or of esteem for a law external to the will is of important consequence for Kant’s ethics. The first interpretation

leads to a picture of autonomy as self-determination and of the subject as being in constant possession of an autonomous ability to act. The second interpretation distinguishes more sharply the moral law from reason, suggesting that autonomy does not highlight the ‘auto’ of autonomy, but the ‘nomos’; autonomy would consist of obedience to a moral law that imposes itself on practical reason. This chapter will advance the thesis that it is the second interpretation of the moral law that should be privileged: the moral law is a positive influence imposing itself on the will. As will be argued, however, the logical structure supporting this second interpretation entails a principle resisting the influence of the moral law—radical evil. Thus, the interpretation of the moral law and of rationality will set the stage for the remainder of this thesis.

## **Reason and the Moral Law**

### *Reason as Consonant with the Moral Law*

#### a) *Outline of the First Interpretation*

Kant’s language in the chapter on respect, though slightly vague in key places, suggests that the causality of the moral law be understood as an *intellectual causality*, whereby the moral law is co-terminous with reason and respect results from the *rationality* of the finite rational being. The moral law, in other words, determines the will in virtue of reason: “thus the moral law, just as *through* practical pure reason it is a formal determining basis of action...” (*KpV* 5: 75. My emphasis). Further down the same paragraph, as Kant describes the moral law as a subjective cause of the will through respect, it is claimed that “the cause that determines this sensation [respect] lies [*liegt*] in pure practical reason” (*KpV* 5: 75) and, furthermore, “pure practical reason...imparts authority to the law” (*KpV* 5: 76). Though scattered throughout the paragraph, these passages serve to highlight the difference

between a subjective determination of the will through feeling, and one through reason. The motive (*Triebfeder*) of a moral character (*sittlichen Gesinnung*) must be free of any sensory condition (*KpV* 5: 75), since the sensory conditions consist of phenomena ordered by the deterministic categories of space and time. The cause of the particular feeling that determines the will—respect—must lie (*liegt*) in practical reason, as opposed to a ‘pathological’ cause (*KpV* 5: 75). The will is free in virtue of reason, since reason’s transcending the particular conditions of sensible nature allows it to legislate universal practical law. Practical reason, therefore, appears to be the autonomous cause of actions that qualify for moral universality. The text seems to affirm this interpretation several pages later when, in the context of explaining both the negative effect of the law—humiliation—and its positive effect—respect—Kant writes that “acknowledgment of the moral law is the consciousness of an activity of practical reason” (*KpV* 5: 79).

A couple of passages further affirm the activity of reason in legislating moral ends, suggesting that the direct causality of practical reason is alone responsible for both the objective and subjective determination of the will to produce a particular act. Respect is defined by Kant in the following sentence as resulting from the activity of reason: “the consciousness of a *free* submission of the will to the law, yet as linked with an unavoidable constraint inflicted—but only by one’s own reason—is respect for the law” (*KpV* 5: 80). Kant seems clear that the constraint exerted on the will is the effect of reason. Three claims, furthermore, point toward the conclusion that respect is produced by reason: 1) the moral law constrains the will (*KpV* 5: 72-73; 80); 2) practical reason legislates its own conception of law (see *G* 4: 412); 3) respect is produced by the awareness of one’s ability to overcome the influence of the inclinations through reason (*KpV* 5: 75-76). For practical reason to exert

a positive influence on the will, it must overcome all inclinations opposed to reason. It follows that reason produces a feeling of humiliation when “striking down” the inclinations that oppose the unconditioned command of the will by reason (*KpV* 5: 73). A positive influence is produced by the negative effect on the inclinations, since the act of pushing aside a contrary inclination on the will furthers the positive influence of reason on the will (*KpV* 5: 73). The structure of the argument indicates that the legislation of moral law results from reason: “since this constraint is exerted only by the legislation of one’s *own* reason, it also contains an *elevation* [of oneself], and the subjective effect on feeling, insofar as the sole cause thereof is pure practical reason, can only be called *self-approval* with regard to this reason” (*KpV* 5: 80-81).

The wording used by Kant suggests that the moral law is the product of reason. Since reason is distinct from the causal determination of *phenomena*, and since reason consists of the capacity to reach particular conclusions from the starting point of universal ideas (*KrV* 3: 357; 364), it follows that a moral determination of the will—which requires freedom and universality—can only result from practical reason alone. A definite picture of morality comes to light on this account, one that has been promoted by various interpreters of Kant’s ethics. Autonomy—the keyword used by Kant to designate acts motivated solely by the moral law—is thought to result from the individual’s own reason. It is thus easy to move toward an interpretation of autonomy as *self-determination* to act morally. Martin Heidegger, for instance, interprets respect according to this picture of autonomy, as a “self-projecting” [*Das unterwerfende Sich-entwerfen...*] (Heidegger 111; p. 146 in German edition). Respect is precisely respect *for* the moral law, but it is reason that grants this law to itself (Heidegger 111). Submission to the law reveals most fully the autonomy of the subject, since the will

submits to a law provided by the subject's own reason: "In submitting to the law, I submit to myself as pure reason. In this submitting to myself, I elevate myself to myself as the free creature which determines itself... Respect is the manner of the self's Being-responsible, face to face with itself; it is authentic Being-its-self" (Heidegger 111). Respect, in fact, is the feeling produced by the law precisely because the law is autonomously given to the subject *by reason*. Heidegger is thus able to conclude that "in respect before the law, I submit to myself. In this submitting-to-myself, I am as I myself" (Heidegger 111).

A similar interpretation of respect and autonomy is provided by Henry Allison, though without the evocations of authenticity characteristic of Heidegger. Respect for the law is understood by Allison, to a certain extent, as resulting from the fact that the law is equivalent to, or at least has a relation to, reason: "respect for the law consists simply in the recognition of its supremely authoritative character, which is to be taken to mean that it provides a reason for action that outweighs or overrides all other reasons" (Allison 123). And further: "if the moral law is self-legislated in the sense in which Kant maintains, then anyone who is conscious of being subject to it will also have a reason or incentive (*Triebfeder*) to obey it and therefore an interest in it" (Allison 237). Allison's interpretation rests on the assumption that the rules guiding moral action are provided by reason: "just as in the theoretical realm the proper, regulative function of reason is to guide enquiry by framing an ideal order involving the systematic connection of phenomena under laws, so too, in the practical realm, its proper function is to guide conduct by framing an order of ends or ought-to-bes" (Allison 40).<sup>5</sup> A consequence of Allison's interpretation of the practical function of

---

<sup>5</sup> Christine Korsgaard holds a similar interpretation. In an article that attempts to establish the feature of humanity that makes it "the appropriate material for a principle of practical reason" ("Kant's Formula of Humanity" 184), Korsgaard claims that the characteristic in question is the capacity of practical reason to adopt

reason is the “Incorporation Thesis”: since it is practical reason that frames the ends that determine the conduct of rational beings, it follows that a rational being cannot act unless practical reason approves or “incorporates” an object into the maxim of the will. Maxims (or reasons) are “incorporated” by practical reason as part of the law determining the action of the will.

*b) Problems with the First Interpretation*

In spite of the evidence provided in the preceding passages, Kant often describes the effect of the moral law in terms that evoke an external force exerted by reason on the inclinations. Kant claims that the moral law “strikes down” self-conceit and that respect arises from the removal of a “counterweight” (*Gegengewicht*) to reason.<sup>6</sup> Later on, in *Religion within the Limits of Reason Alone*, Kant argues that the moral law is a “motivating force”, one that opposes the contrary influence of the inclinations (*Rel.* 6: 23/18, note 43/1). If respect, however, results from the fact that a proposed action is based on one’s *own reasons* for acting, then it is difficult to imagine that reason could exert an actual force on the will.

---

ends in general (187). The difference between a moral and a non-moral end, however, is established by the nature of practical reason. Rational choice—as opposed to non-rational choice—has a “value-conferring status” (196). It is rationality itself, in other words, that determines the goodness of the end. John Rawls also holds a similar interpretation, arguing that the content of morality is defined by practical reason. Kant maintains a constructivist ethics, by which the categorical imperatives are defined through a process of construction, the “form and structure of which mirror both of our two powers of practical reason” (Rawls, *Lectures*, 237). The content of ethics results from a constructive activity of reason. The “constructivist procedure is framed to yield the principles and criteria that specify which facts about actions, institutions, persons, and the social world generally are relevant in moral deliberation” (Rawls 246). Finally, Leslie Mulholland argues that “a rational being is related to moral values of any sort...as the subject that produces those values by its relation to the law of its willing...the rational being is the ground of the categorical imperative, not through its being an end of moral (or nonmoral) value, but through its autonomously producing the moral law (Mulholland, *Kant’s System of Rights*, 108, 110).

<sup>6</sup> The moral law impairs natural self-love, restricting it “to the condition of agreement with” the moral law (*KpV* 5: 73). The moral law, however, *strikes down* self-conceit, which presumes the individual has self-worth prior to the will’s being determined by the moral law (*KpV* 5: 73). Overcoming self-love and self-conceit is compared by Kant to the removal of a counterweight that, as a result, furthers the influence of the moral law on the will. Respect arises from the furthering of the law’s influence (*KpV* 5: 75-76).

Kant's language, then, must be unclear, for reason determines the will through deliberation; the power of judgment subsumes a particular act under a universal rule of reason, but the rule belongs to reason and should not be considered a positive force exerted on the will (*KpV* 5: 67). Andrews Reath adopts this line of argument, claiming that respect, like any incentive, is felt as a result of one's "regarding it as providing a sufficient reason for acting" (Reath 296). Reath rejects the view that the moral law opposes the influence of the inclinations by setting up an "opposing psychological force", since the moral law "exerts no such force" (Reath 296). Respect, therefore, is not a force exerted on the will by the moral law or by practical reason; respect is, rather, "to *recognize* the Moral Law as a source of value, or reasons for action that are unconditionally valid and overriding relative to other kinds of reasons." (Reath 287. My italics.).<sup>7</sup>

The claim that reason 'confers value' or frames an order of "ends and ought-to-bes" raises a problem, however: is the spontaneity of reason in setting its own ends reliable for setting *universal* ends? In other words, if reason confers value on an end due to the fact that the end is rationally chosen, then how can diverse rational agents be assured of adopting the same end, without committing Kant to the seemingly empirical claim that, in actual fact, the reason of each individual manifests the same structure? Paul Guyer voices this concern in presenting the interpretations of Korsgaard and Mulholland on the value-setting status of reason. Guyer objects that such an interpretation

---

<sup>7</sup> A similar interpretation of respect that directly considers and rejects the « forces » model is provided by Stratton-Lake. For the details of that discussion, see Stratton-Lake 39-40, as well as my summary of Stratton-Lake's argument on page 3. Henry Allison also mentions and rejects the "forces" model of respect. His interpretation of respect as an effect of one's reasons for acting is similar to the interpretation provided by Andrews Reath (Allison 126-127)

does not appear to place any particular *constraint* on the creative value setting of the agent...unless it does so by begging the question, that is, precisely by using an already *moral* conception of rational agency and thus simply assuming that a rational agent is one who adheres to the moral law of universalizability and so will not create any nonuniversalizable assignment of values. (Guyer 151).

According to Guyer, the interpretation that equates morality with rationality risks rendering moral ends arbitrary, which is the result that Kant had hoped to avoid in grounding his ethics on the universal moral law. The necessity of the moral law, according to some scholars, is better assured by interpreting the moral law as established on foundations other than pure rationality.<sup>8</sup>

*Second Interpretation: distance between Reason and the Moral Law*

Despite the apparent equivalence of the moral law with rationality, Kant's language suggests another interpretation of the relation between reason and the moral law. Already in the *Grounding*, in the context of a footnote clarifying his notion of respect, Kant describes the moral law as "that very law which we impose on ourselves and yet recognize as necessary in itself" (*G* 4: 402). The previous interpretations of the moral law easily notice the self-imposition of the moral law as a function of rationality; on the other hand, they overlook—or perhaps underemphasize—Kant's claim that the law is "necessary *in itself*." The law appears to be both self-imposed *and* transcending any determination prior to its own determination of the will. The law appears to reside in a position (or lack of position<sup>9</sup>) that forbids considering the law as synonymous with reason:

---

<sup>8</sup> Guyer reinterprets the moral law as resting on the unconditionally good end set by rational agents. The rational agent, however, is an end in itself because the agent manifests his freedom in setting ends. It is thus freedom that has absolute value, which means that the moral law is unconditionally valuable in virtue of the freedom of the rational agent (see Guyer 151). Adherence to the moral law is thus not intrinsically valuable (Guyer 155). This last point, however, will be called into question by scholars who see adherence to the moral law as intrinsically valuable in itself.

<sup>9</sup> This phrase will become clearer once I discuss Rogozinski's interpretation of the moral law.

Here philosophy is seen in fact to be put in a precarious position, which should be firm even though there is neither in heaven nor on earth anything upon which it depends or is based. Here philosophy must show its purity as author of its laws, and not as herald of such laws as are whispered to it by an implanted sense or by who knows what tutelary nature. (*G* 4: 425)

Unless reason should be interpreted as residing “neither in heaven nor on earth”, then it appears that it is the law itself that cannot be situated in any exact position. The moral law appears to be distinct from reason, though reason itself might “dictate” the principles of the moral law, as the sentence immediately following the above-cited passage suggests: “such laws may be better than nothing at all, but they can never give us principles dictated by reason” (*G* 4: 425-426).

It is tempting to cite Kant’s arresting image of finite rational beings as “legislating members of a kingdom of morals” as proof that autonomous reason frames and dictates the ends and the obligation to which the will must conform. It is frequently overlooked, however, to cite the sentence immediately following this passage, in which Kant emphasizes the *submission* of finite rational beings to the moral law. It is worth citing the passage in full, beginning with Kant’s claim for the legislating capacity of rational beings:

We are indeed legislating members of a kingdom of morals possible through freedom and presented to us by practical reason for our respect; but we are at the same time subjects of this kingdom, not its sovereign, and a failure to recognize our low level as creatures—as well as self-conceit’s [attitude of] refusal toward the authority of the holy law—is already a defection from the law in spirit, even if its letter were fulfilled. (*KpV* 5: 82).

The first clause of the quotation already suggests a distance between reason and the law: practical reason presents a ‘kingdom of morals’ distinct from reason itself. If this interpretation were questionable, Kant’s claim immediately following this phrase should clarify the matter: we are subjects of the kingdom—in a position of submission to the

reigning “holy law”—and contrasted by Kant with the ‘sovereign’ of the kingdom. The sovereign, not the subject, dictates the content of the law. Practical reason may ‘impart’ authority to the law,<sup>10</sup> but the reason of a subject of the kingdom of morals does not “frame its own ends” (as Henry Allison claims), as if reason were the determinant of an objective good or evil end. The obligation and content of moral legislation belongs to the law, for it is the *a priori* law that is the “origin” of moral principles and “their authority to command” (*G* 4: 426). Autonomy does not consist of one’s ability to dictate an order of ends through reason (even though the ends set by practical reason are universal and objective), but to stand in obedience to the command of the moral law rather than to the whims of passing inclinations: “by virtue of the autonomy of his freedom [the human being] is the subject of the moral law, which is holy” (*KpV* 5: 87).

a) *Logic of Real Opposition*

If the distinction between the moral law and reason is taken into account, then Kant’s use of phrases such as “counterweight” or “diminution of obstacles” can be understood as an opposition of principles. The moral law exerts a motivating influence on the will (*Rel.* 6: 23/18, note 43/1), so the inclinations, whose particularity opposes the universality of the moral law, act as a counterweight to the law’s positive influence. Kant claims that the removal of the counterweight by the moral law (which results in humiliation) automatically furthers the influence of the law on the will, thereby producing respect through our awareness of a law capable of determining the will (see *KpV* 5: 73, 74). The language through which Kant claims that the mere removal of a “counterweight” furthers the causality

---

<sup>10</sup> *KpV* 5: 76

of the moral law can be better explained in reference to Kant's early essay "Attempt to Introduce the Concept of Negative Magnitudes into Philosophy".

Introducing into philosophy a new notion of causality taken from the domain of modern mathematics and the physics of Isaac Newton, Kant argues that opposition can be thought along two lines. The first is the traditional conception of logical opposition, by which one predicate opposes the other predicate, with the consequence that the object is "nothing at all" (*NG* 2:171).<sup>11</sup> A geometrical figure with a predicate of roundness and a predicate of squareness at the same time and in the same place simply fails to exist. Logical opposition formed the structure on which traditional ethical and theological ideas were thought: since existence and goodness were considered synonymous, evil could be said not to exist since it consisted of a privation of goodness as opposed to a positive resistance against goodness. The young Kant accounted for evil in this way in an essay on Leibnizian optimism.<sup>12</sup> Leibniz himself built his theodicy on these foundations.

In an important passage in the *Theodicy*, Leibniz analyzes the concept of the "natural inertia of bodies" because he considers the concept as a "sample of the original limitation of creatures" (Leibniz 140). The goal of the analysis is to exculpate God for the existence of evil by demonstrating that "privation constitutes the formal character of the imperfections and disadvantages that are in substance as well as in its actions" (Leibniz 140). Leibniz explains the action of God's grace using the image of several ships, each proceeding downstream at different velocities due to the differing amount of cargo loaded in each ship. The key assumption held by Leibniz is that matter naturally tends toward inertia:

---

<sup>11</sup> Translations from this essay are my own.

<sup>12</sup> See Kant's essay: "An Attempt at Some Reflections on Optimism."

it is matter itself which originally is inclined to slowness or privation of speed; not indeed of itself to lessen the speed, having once received it, since that would be action, but to moderate by its receptivity the effect of the impression when it is to receive it. Consequently, since more matter is moved by the same force of the current when the boat is more laden, it is necessary that it go more slowly. (Leibniz 140)

Matter does not actively lessen the speed of the ship through a force or formal cause, but slows the ship due to its natural tendency toward inertia; a greater quantity of matter, therefore, naturally slows the ship. God's grace can be compared to the river, granting being and relative perfection to the creature, producing and conserving "whatever is positive in creatures" (Leibniz 141). The slowness of the boats can be compared with the "defects to be found in the qualities and the action of the creature" (Leibniz 141). The absence of goodness is not the direct result of God's will, but is the consequence of the natural inertia of bodies (Leibniz 141). The assumption concerning the natural inertia of bodies allows Leibniz to avoid the common sense conclusion that the existence of evil is the consequence of God's creation of the world:

The current is the cause of the boat's movement, but not of its retardation; God is the cause of perfection in the nature and the actions of the creature, but the limitation of the receptivity of the creature is the cause of the defects there are in its action. Thus the Platonists, St. Augustine and the Schoolmen were right to say that God is the cause of the material element of evil which lies in the positive, and not of the formal element, which lies in privation. (Leibniz 141)

The discovery of negative magnitude led Kant to abandon Leibnizian optimism—with its concomitant justification of God and explanation for evil—only about four years after writing an essay defending the topic.<sup>13</sup> "–A" represents not a limitation, but a positive

---

<sup>13</sup> See Kant's essay "An Attempt at some Reflections on Optimism", published in 1759. Kant's discovery of negative magnitude led him to abandon Leibnizian optimism and traditional theological conceptions of God for the following reason: if all perfections are found in God and nothing 'negative' belongs to him, then the discovery of negative magnitude means that we must either predicate negatives of God (since a negative magnitude is a reality) or we must admit that not everything that is comes from God. Lebrun sums up the problem in the following way: "since, for us, two positive realities can enter into conflict, how can one unite all

resistance to “+A”: the two predicates oppose one another, but the consequence “is also something (*cogitabile*)” (NG 2: 171). The equal amount of force by two bodies opposing one another has as its consequence the absence of motion, but the body at rest is nonetheless “something (*repraesentabile*)” (NG 2: 171). The two predicates belong to “one and the self-same thing” and, as in logical opposition, “they belong to [the thing] simultaneously” (NG 2: 171-172). Unlike logical opposition, however, the consequence of real opposition is something—*repraesentabile*. The difference between Kant and Leibniz on the issue of causation can be summed up by Martial Guérout:

“The agreement is thus entirely established between a dynamic and a metaphysics which, everywhere suppressing real oppositions in order to leave nothing but logical oppositions, denies the possibility of a destruction of the positive by [something] positive—denies that negativity can possess its own reality. Supported by the tendency of Newtonian philosophy, Kant will radically oppose this dynamic and metaphysics.... Wherever there exists a positive principle and where the phenomenon=zero, it is necessary to admit a real, equal, positive, and inverted principle” (Guérout 168. My translation).<sup>14</sup>

The logic of privation is thus not the sole explanation for the absence of an object or quality; a real opposition is discovered by Kant to explain why an object or quality fails to occur due to the nullification produced by an opposing force.

Already in this pre-critical essay, Kant claims that practical philosophy (*praktischen Weltweisheit*) can be interpreted according to the logic of real opposition (NG 2: 182). Vice is not a negation of virtue, but is an actual force opposing the ‘interior law’, though Kant has not yet devised the notion of a moral law (NG 2: 182). The existence of the negative relation,

---

these realities in a being that, by definition, excludes all negatives?” (Lebrun 195. My translation). See also Lebrun, *La Patience du Concept*, 290-293, in which he explains how Hegel tried to overcome the division of negative magnitude by claiming that, since a negative is only negative in virtue of its link to a positive force, there cannot exist a negative or positive in and of itself and hence real opposition can be overcome. Rogozinski claims Hegel’s innovation on negative magnitude lays the foundation for the *Aufhebung* (Rogozinski 25).

<sup>14</sup> For an explanation of Leibniz’s interpretation of causality and the impact of Kant’s rejection of Leibnizian optimism on Kant’s later philosophy, see Rogozinski chapter 1, especially pgs. 21-24.

however, depends on the existence of its opposite, positive force; otherwise, the positive and negative quality of the relation could not be established in the first place.<sup>15</sup> The category of vice, therefore, cannot be thought without reference to its opposite, the interior law:

For vice can only occur in so far as a being has within him an inner law (either simply conscience or consciousness of a positive law as well), which is contravened by this action. This inner law is a positive reason for a good action, and the consequence can only be zero if the consequence which would result from consciousness of the law on its own is cancelled (*aufgehoben wird*). (NG 2: 182)

Due to his discovery of real opposition, Kant begins to adopt a picture of ethics that would find its culmination in the moral law and an opposing principle which, for Kant, would become fully apparent only once his conception of the moral law had become fully formulated—radical evil.

Melissa Zinkin notices the importance that the concept of negative magnitude has on interpretations of Kant's notion of respect. The moral law, in constraining or “striking down” the influence of the inclinations on the will, *a fortiori* furthers its own influence on the will, leading to respect:

Respect is the free and proper motion of the will that is made possible when an obstacle or counterweight preventing its movement is removed. In this way, the feeling of respect is both that *with which* the will follows the moral law and also the *effect* of the moral law on the will. Imagine two weights on opposite ends of a lever. The removal of one weight is the same act as causing the other to fall. The removal of the hindering weight is what makes possible the falling of the other. And, just as the falling of this weight can be said to be its proper, unhindered, motion, so is it the nature of the will to act out of respect for the moral law, when unhindered by self-conceit. (Zinkin 45).

Zinkin observes that the problem of respect is centered upon one's reasons for acting. If an individual is supposed to make the moral law the basis of action in order to respect the law, then the question arises: why should I judge my reasons for acting against the standards of

---

<sup>15</sup> See Michaël Foessel, *Le scandale de la raison*, p. 97.

the moral law?<sup>16</sup> The answer to this question is not sufficient on a model that regards respect as a form of esteem for an agent's reasons for acting. The standard model of respect overlooks (or has difficulty accounting for) those passages in which Kant uses language—such as “counterweight”—that suggests the image of forces. Zinkin thus rejects Henry Allison's interpretation of respect as a “phenomenological” account of moral experience, arguing instead that Kant “has an *ontology* of the will that makes it possible for him to claim that the moral law has a necessary effect on us” (Zinkin 41). Since the moral law always exerts a positive influence on the will, the causality of the moral law is automatically furthered once the “counterweight” of the inclinations is removed.

b) *Autonomy and the Typic: further support for the Logic of Real Opposition*

Concluding the second chapter in the second *Critique* on the “object of pure practical reason” and preceding the beginning of the third chapter on the “incentives of pure practical reason”, Kant devotes several paragraphs to the “typic of the pure practical power of judgment.” The placement of these paragraphs is significant, for it provides a bridge between the object of practical reason, which is completely *a priori*<sup>17</sup> and the incentive of respect in chapter 3, which has reference to subjective feelings. The typic solves a problem for the determination of the will: in the act of making a moral choice, judgment subsumes a particular case under a universal rule of reason. The particular case is a material event in the world, while the rule consists of a universal *a priori* standard of reason. Since the universal

---

<sup>16</sup> “Why should I evaluate my reasons with regard to the moral law? There must be something about me that has an interest in making the law of reason into *my own particular reason* for action. I must want to be a self that acts in accordance with the moral law. But where does this interest come from?” (Zinkin 39)

<sup>17</sup> Since the traditional content of good and evil is related to material conditions (such as pleasure or pain), Kant redefines the meaning of ‘good’ and ‘evil’ in order to make both their meanings *a priori* and hence determined by the moral law. Kant takes advantage of a fortuitous distinction in German, by which good and evil can be rendered either as “*das Gute*” and “*das Böse*” (understood in a material sense) or as “*das Wohl*” and “*das Übel*” (understood in a formal sense). (*KpV* 5: 59).

rule of reason is also a law of freedom in virtue of its distinction from material conditions, it seems difficult for Kant to maintain that judgment may apply a universal rule of reason to a particular case that takes place “in the world of sense” (*KpV* 5: 68.<sup>18</sup>).

The requirement of independence from material conditions demands from the understanding a type of the law: the law of reason can be presented more concretely (*in concreto*) through the image of a law of nature, yet a law thought of “only in terms of its form” (*KpV* 5: 69). The situation of finite rational beings placed under a universal law demarcating the good from the evil of *particular sensible acts* demands that a concrete *type* of the law serve as a middle term between the moral law of reason and the law of nature. The *type* of the moral law can be expressed according to a rule reminiscent of the second formulation of the categorical imperative: “Ask yourself whether, if the action you propose were to occur according to a law of nature of which you yourself were a part, you could indeed regard it as possible through your will” (*KpV* 5: 69). The similarity between the law of nature considered as a type and the law of nature considered as a formulation of the categorical imperative is not coincidental, for a middle term is required to mediate the particular act of the will and the universal determination of the law. The requirement for practical reason of a type acting as a middle term between the universal law of reason and the particular act presented by the understanding holds important consequences for the interpretation of respect and autonomy.

Recognizing that the power of judgment requires a *type* in order to subsume a particular sensible act under a universal rule changes the interpretation of autonomy, due to the fact that the will cannot directly conform to the universality of the moral law. Since the

---

<sup>18</sup> For Kant’s full discussion of the problem, see *KpV* 5: 67-68.

human will is never pure (Krüger 102), a *type*—as the ‘form’ of nature—is needed to mediate between the particular, material acts of a human being and the pure *a priori* moral law. The *type*, in other words, is the necessary consequence of the nature of a finite rational being. Practical reason does not directly determine the will because the sensibly conditioned will does not conform to the formal universality of the law. Practical reason, in other words, is not the author of the law; instead, it provides ‘authority’ to the law, a detail noticed by Gerhard Krüger: “The moral necessity of reason is only possible ‘insofar as it constitutes itself as an *executive* power of the law’” (Krüger 101). Krüger cites in support of his claim a passage used to defend the ‘incorporation thesis’: “the freedom of the power of choice [*Willkür*] has the quite peculiar characteristic that it cannot be determined to an action by any incentive *except insofar as the human being has admitted the incentive into his maxim*” (*Rel.* 6: 23-24/19; cited in Krüger 101).

The passage cited by Krüger is often used by scholars to defend the interpretation of autonomous ‘legislation’ as one of reason framing its own ends: since Kant argues that the will can be determined to an act only through the individual’s *choice* to make the motive a reason for action, it follows that an individual is considered autonomous precisely because he always *determines* his own motive and is never *determined by* a motive. Krüger, however, claims that “the most important error in interpretation (*wichtigste Fehlinterpretation*), which must be kept away from this concept [of autonomy], is the misrecognition of its character as *type*”, for practical reason is still *human* reason, a reason that is empirically conditioned (Krüger 101). The human will, in other words, is never disinterested (Krüger 102), since, as Kant demonstrates in the *Anthropology*, the human being is situated in a historical context,

subject to the particular influences of the passions and the temptation to deceit.<sup>19</sup> Since the will can never attain the pure universality of the moral law, and since reason itself is also subject to the contingencies of human life, it follows that reason can neither be the author of a pure moral law, nor can it dictate that law without the mediation of a type of *a priori* law. The categorical imperative thus acts as a type, mediating the sensibly affected will and the pure moral law by means of an *als ob* (as if): “reason proclaims the law from itself (*von sich aus*), *as if* it were the author (*Gesetzgeber*) of the law.... The categorical imperative thus formulates itself as the demand that man should look upon his reason, that is, his own person, *as if* he himself were the author (*Gesetzgeber*) of the moral law” (Krüger 101). The dignity of the human being does not result from his autonomy, according to which he stands as the author of the moral law; rather, the human being is autonomous due to his dignity to *obey* the universal moral law:

For Kant, the concept of autonomy expresses nothing but the unconditional unreservedness (*Vorbehaltlosigkeit*) of obedience. The “self” in the concept of self-lawgiving does not mean unconditioned, “creative” freedom obeying only to oneself—which wants to be faithful to oneself according to a law—but *unconditional responsibility before the law*, precisely before which freedom itself cannot withdraw from. Here lies a historical difference between Kant and modern interpretation. (Krüger 103. My translation.)

The nature of autonomy as submission to the moral law is overlooked by commentators who do not notice that the categorical imperative should be understood as a type and not as a schema of the moral law.

---

<sup>19</sup> See the preface to the *Anthropology*, in which Kant defines an anthropology from a pragmatic perspective and highlights the difficulties of arriving at a fair judgment of a person, given our tendency to dissimulation upon being watched (*A. 7*: 119-123).

### **Attempts to Avoid an Interpretive Impasse**

The interpretations of respect and autonomy have so far been dichotomous: either reason legislates its own law—and respect consists of esteem for one’s own rational reasons for acting—or reason provides authority to a law it receives—and respect is directed toward the law itself. Two commentators, however, suggest a way out of this interpretive impasse.

Bernard Carnois attempts to mediate between the interpretations of Heidegger, for whom the moral law is given by reason, and of Krüger, for whom the moral law is a given (Carnois 119). Carnois begins with the claim that the moral law “is in us...since it is part even of the structure of our being, insofar as we are rational beings” (Carnois 120).<sup>20</sup> Reason, however, “does not produce the law”; the law, rather, is a “co-naissance”, through which reason is revealed to itself by revealing the moral law (Carnois 120). The moral law is somehow distinct from reason, yet reason is revealed to itself through the legislation of the moral law because the universality of the law is “inherent” to the “nature of reason” (Carnois 120). As already suggested above, the “form of an intellectual causality”, by which the moral law produces respect, can either be understood with respect to the consciousness of reason or with respect to the ‘transcendence’ of the moral law. Carnois attempts to mediate between these two interpretations: the moral law is distinct from reason, yet is only revealed through rationality. Carnois, however, does not clearly establish the nature of the ‘co-naissance’ that is characteristic of the knowledge (*connaissance*) of reason and the moral law. In explaining the ‘co-naissance’, Carnois seems to revert to Heidegger’s interpretation in arguing that the ‘fact of reason’ is “just as much the fact of the existence of reason in us as the fact of the existence in us of the moral law” (Carnois 120). Though Carnois has not sufficiently

---

<sup>20</sup> Translations of Carnois are my own, unless otherwise indicated.

clarified the relation and difference between reason and the moral law, his attempt to mediate between the two approaches to the moral law remains a good one. Jacob Rogozinski succeeds in this attempt, by reinterpreting the notion of autonomy and the relation of the moral law to reason.

Jacob Rogozinski claims that the categorical imperative should be qualified as an *auto-heteronomy*. Practical reason legislates the law to the will, but reason cannot be considered the author of the law through the act of legislation: “we are indeed legislating members of a kingdom of morals possible through freedom and presented to us by practical reason for our respect; but we are at the same time subjects of this kingdom, not its sovereign” (*KpV* 5: 82). The law imposes itself on reason, yet is legislated to the will by reason, which does not formulate the content of the law: “One who commands (*imperans*) through a law is the *lawgiver* (*legislator*). He is the author (*autor*) of the obligation in accordance with the law, but not always the author of the law. In the latter case the law would be a positive (contingent) and chosen [*willkürlich*] law” (*MM* 6: 227). The interpretation that considers reason as “framing an order of ends or ought to bes” (Allison 40) does not account for Kant’s insistence that reason is not the author of the law.

The subjection of the will to the law results from practical reason legislating a law it does not create:

Autonomy, in its authentic Kantian meaning, would not be the absolute autonomy of a sovereign subject creating the Law (a strictly impossible hypothesis according to Kant) but a “heteronomous” autonomy, the *auto-heteronomy* of a subject submitting himself to a Law of which he is not the author, that he has not created, that he receives and makes his own as if it was his Law (Rogozinski 189. My translation).

The importance granted to the *as if* (*als ob*) by Krüger and Rogozinski should not go unnoticed, for practical reason can thus only legislate the law to the will *as if* reason were the author of the law, but not literally. The moral law cannot be situated in a definite position; its nature as pure *a priori* form defines the law according to where it is not to be found. The law is apophatic: its negative relation to a position (whether it be heaven, reason, human nature) can be established, but a pure form can never be granted a definite position, which assumes empirical *matter* that the pure form excludes. The law can only be understood with respect to where it is not: the law un-penned by anyone but itself, the law whose content is known through its own ‘exposition’, can only be characterized according to its ‘*ex-position*’, its complete lack of position.<sup>21</sup>

Interpretations suggesting or overtly establishing a proximity between the law and reason overlook the *ex-position* of the law: the law is its own sovereign, its own author. Respect, therefore, cannot consist of a motive to follow one’s own reasons for acting because the content of the moral law does not belong to practical reason. The exposition of the law provides the foundation for interpreting Kant’s solution to the problem of radical evil in the *Religion*. Since the content of the law can never be known with certainty, the opposite of the law—institutions and error, which are both empirical—become the means according to which the law can be sought. However, the same dynamic will raise a problem for Kant: if the moral law exists in an *ex-position*, can a person ever be sure that he or she is following

---

<sup>21</sup> The neologism ‘*ex-position*’ is a play on words used by Rogozinski to emphasize his interpretation of Kant as having established the moral law prior and above all other principles that traditionally served as foundations for ethics (such as the Good, happiness, or the will of God). The moral law is good not because it is founded on the Good (or is dictated from God), but because it conforms to the form of the moral Law. The law, therefore, ‘exposes’ itself in the particular laws of the categorical imperative: “Rather than a ‘retreat’ [*retrait*] of the Law... it will be preferable to speak from now on of its *ex-position*. To be understood, first of all, as the deliverance of the Law, as this break [*césure*], this return where it liberates itself from the tutelage of the Good, *situates itself* [*se pose*] *outside* and beyond the principles of an era and of supreme foundations” (Rogozinski 111).

the law? Has Kant, in other words, provided a viable solution to the empiricist philosophy he hoped to overcome? My task for the remainder of the thesis will be to show how Kant's solution to radical evil involves the empirical training of historical institutions to coerce the will into following the moral law. Since the will never entirely overcomes the influence of evil—since the positive force of the law always supposes an opposing negative force—Kant's ideal of duty for duty's sake will never be achieved and the viability of Kant's ethics as an alternative to an empiricist ethics will be called into question.

### Chapter 3: Real Opposition and the Discovery of Radical Evil

In turning to *Religion within the Limits of Reason Alone*, the assumptions concerning negative magnitude in Kant's 1763 essay begin to hinder a philosophical justification for the instantiation of the moral law and its object (the highest good) in the world. For the concept of radical evil is an outcome of Kant's philosophical system and not an antiquarian religious interest on the part of the Prussian philosopher. Radical evil rears its head at the place where Kant's philosophical system is most vulnerable: finding a justified philosophical relation between *noumena* and *phenomena*, reason and history, the moral law and the will.<sup>1</sup> If radical evil prevents Kant from establishing the relation between *homo noumenon* and *homo phenomenon* within the structure of an ethics of autonomy, then Kant's claim that the will can be determined without external influence would appear untenable. My analysis of the *Religion* will keep the following question in mind: can a finite rational being, evil by nature, choose of his own volition to follow the moral law? The logic of real opposition, however, will commit Kant to the position that the moral law cannot become a motivating principle of the will without positing a non-moral influence inclining a person to duty for duty's sake.

#### Why Religion?

The preface to the first edition of *Religion within the Limits of Reason Alone* begins by affirming the autonomy of the will: "So far as morality is based upon the conception of man as a free agent who, just because he is free, binds himself through his reason to unconditioned laws, it stands in need neither of the idea of another Being over him, for him

---

<sup>1</sup> Ricoeur claims that commentators have not paid enough attention to the fact that the whole problem of evil and of religion in Kant's oeuvre is raised at the point in which the moral law is supposed to reach its fulfillment in the world: "In fact, it has not been sufficiently noticed that the doctrine of evil, which opens the philosophy of religion, does not reach its climax in the Essay on Radical Evil, but fully accompanies the philosophy of religion. The true evil, the evil of evil, is not the violation of something forbidden, the subversion of the law, or disobedience, but fraud in the work of moral fulfillment (totalisation)...The search for a complete object of the will is, in essence, 'antinomic' (antinomique). The evil of evil is born instead of this antinomy (Ricoeur 414).

to apprehend his duty, nor of an incentive other than the law itself, for him to do his duty” (*Rel.* 6: 3/3). Kant’s book on religion thus reiterates a point that would land him in trouble with the Prussian authorities<sup>2</sup>: the will of finite rational beings does not need—and *should not* need—to be determined through the influence of a divine being, nor of a church. The moral law, in other words, is the sole incentive required by the will to act. The first sentence, in fact, lays out the game plan for the entire book: Kant must affirm the autonomy of the will against the archetype of heteronomy—religion.

The subject of religion is of interest to Kant because the concept of God is an important one for the *fulfillment* of morality. Kant never claims that the concept of God is necessary for the sake of moral action because such a claim would destroy the foundations of ethics laid out in the *Grounding* and in the *Critique of Practical Reason*. The concept of God, however, is important for the instantiation of the moral law and of the highest good in the world.<sup>3</sup> A person “honouring the moral law” (*Rel.* 6: 5/5) would also attempt to realize the demands of the law in the world “because the moral law demands that the highest good possible through our agency should be realized” (*Rel.* 6: 5/5). And since human beings

---

<sup>2</sup> On the controversy created by the first edition of the Religion, see Kant’s preface to the *Conflict of the Faculties* (CF 7: 5-11). On the scandal that Kant created for his contemporary *Aufklärer* by his discussion of radical evil, see Jean-Louis Bruch, *La Philosophie Religieuse de Kant*, p. 75-76. Goethe, for instance, complained to Herder that “Kant required a long lifetime to purify his philosophical mantle of many impurities and prejudices. And now he has wantonly tainted it with the shameful stain of radical evil, in order that Christians too might be attracted to kiss its hem” (English translation quoted in Allison 270). In a milder tone, Schiller would write that “all that we can expect of the well-known character of defenders of religion is that they accept support but reject philosophical foundations, so much so that Kant has done nothing else than patch up the spoiled edifice of foolishness” (Bruch 76).

<sup>3</sup> The link between the highest good and rational religion has been noticed by a few commentators. Nancy Sherman argues that “the ethical commonwealth becomes the embodiment of the important, albeit problematic notion of the highest good” (Sherman 295). Allen Wood argues that “in my pursuit of the highest good as a social good, I must recognize the moral law not only as pertaining to me as a rational being, but also as a law binding morally on all rational beings, a law which obliges me to see myself as part of a moral unity with such persons, a *member* of a Kingdom of Ends, under a divine *head*” (Wood 191). See also Yovel (70-77) and Bruch (161-66).

cannot fulfill the demands of the highest good on our own, we require “the idea of a powerful moral Lawgiver” to supplement our limitations (*Rel.* 6: 6/5).

In the ‘Dialectic’ of the *Critique of Practical Reason*, Kant introduces a new object<sup>4</sup> of practical reason: the Highest Good (the complete unity of virtue and happiness). It belongs to the nature of reason to seek the unconditioned totality for all conditioned objects (*KpV* 5: 107). Theoretical reason, however, can only know an object in its *phenomenal* appearance, never as an object in itself. The unconditioned totality thus falls outside theoretical cognition, producing an antinomy of theoretical reason through which reason seeks an object that cannot be known through its own powers (*KpV* 5: 107). Practical reason does not fare any better, for in seeking the unconditioned totality of the object of practical reason (the highest good; *KpV* 5: 108), another antinomy of reason results: neither can happiness be the motivating cause of virtue—which would result in a heteronomous determination of the will—nor can virtue be the cause of happiness—which cannot be known, since knowledge of the natural connection of cause and effect surpasses the capacities of theoretical reason (*KpV* 5: 113). The resolution of the antinomy follows a regular Kantian strategy: assert both claims as true, yet cut the knot through appeal to the distinction between *phenomena* and *noumena*. That happiness can be the motivating cause of virtue is “false absolutely”, for the moral determination of the will cannot be based on sensible conditions (*KpV* 5: 114). However, while it may be false for virtue to be the cause of happiness for human beings in their *phenomenal* condition, it is entirely apt for virtue to cause happiness when human beings are considered in their *noumenal* condition:

---

<sup>4</sup> The object of practical reason in the ‘Analytic’ concerns good and evil understood as supersensible objects (*das Wohl* and *das Übel*).

since I not only am authorized to think my existence also as [that of a] noumenon in a world of understanding but even have in the moral law a purely intellectual determining basis of my causality (in the world of sense), it is not impossible that the morality of [one's] attitude should have a connection, and moreover a necessary one, as cause with happiness as effect in the world of sense, if not a direct connection then still an indirect one (*KpV* 5: 115).

The solution rests on Kant's assumption that an ought implies a can; in other words, if the moral law commands my will to a certain act, then I must perform the action as if (*als ob*) I were capable of it. To deny this assumption would amount to subjecting duty to the contingency of *phenomena*, for acting from duty would depend upon whether or not the ordering of natural objects made the attainment of duty possible. If, therefore, the unity of happiness with virtue is a necessary condition for the fulfillment of the highest good, and if such unity depends upon the performance of virtue, then the fulfillment of virtue must be possible for the sake of meeting the demands of the moral law.

Fulfilling the highest good, though not a condition for following the moral law, nonetheless bears a consequence for a person's motivation to follow the moral law—for the highest good is "inseparably linked with the moral law" (*KpV* 5: 114).<sup>5</sup> It is necessary to posit (without ever proving) the existence of God and immortality in order to attain the unity of virtue and happiness required by the moral law (the highest good). For if practical reason did not posit the existence of God and immortality, then a person would not expect to fulfill the highest good. And if a person would be unable to fulfill the highest good, then he could not fulfill the command of the moral law, which demands the unity of virtue and happiness. The denial of the postulates thus leads to a denial of our capacity to follow the moral law.

---

<sup>5</sup> Another passage indicates that practical reason demands the production of the highest good: "Only in this subordination is the *highest good* the entire object of pure practical reason, which must necessarily present this good as possible, because to contribute everything possible to its production is a command of this reason" (*KpV* 5: 119).

God and immortality are necessary for avoiding an *ad absurdum practicum*, a phrase suggested by Allen Wood in order to establish a bridge between Kant's discussion of the postulates of practical reason and his analysis of evil in *Religion within the Limits of Reason Alone*. Wood summarizes the importance of the highest good as follows:

Assume that I deny either the existence of God or of a future life. Now if I deny either of these, then I cannot conceive the highest good to be possible of attainment. If I deny that I can conceive the highest good to be possible of attainment, then I presuppose or imply that I will not pursue the highest good, or commit myself not to pursue it. But if I do not pursue the highest good, then I cannot act in obedience to the moral law. Therefore, by denying the existence of a God and a future life, I have presupposed or implied that I will not obey the moral law, or have committed myself not to obey it (Wood 29).<sup>6</sup>

Although the moral law demands the complete unity of virtue and happiness, the absolute fulfillment of either is impossible for a human being. For the pure form of the moral law is infinite, which implies the incapacity of a finite rational being to ever fully attain virtue; and the complete connection between virtue and happiness cannot ever be judged as fully possible in a finite world that does not necessarily (*apodeictically*) grant the fulfillment of happiness (defined as “the state of a rational being in the world for whom in the whole of his existence everything proceeds according to his wish and will.” *KpV* 5: 124). Practical reason must postulate the immortality of the soul—in order to suppose an infinite moral striving toward the demands of the formal moral law (*KpV* 5: 122)—and the existence of an all-seeing God—who grants the unity between virtue and happiness upon observing the intention of the moral agent (*KpV* 5: 125). If, however, the highest good is “inseparably linked with the moral law”, and if the moral law is meant to be realized in the world<sup>7</sup>, then it

---

<sup>6</sup> For a similar argument concerning the importance of the postulates for the fulfillment of the moral law, see Rogozinski, p. 224 and Reboul, p. 166-168, Ricoeur, p. 408-414.

<sup>7</sup> Yiramiyahu Yovel identifies two stages in Kant's ethics: the formal stage, which deduces the existence of the categorical imperative, and the material stage, which develops the concept of the highest good and postulates

follows that the highest good itself is meant to be realized in the world, in history. It follows that morality “leads ineluctably to religion, through which it extends itself to the idea of a powerful moral Lawgiver, outside of mankind, for Whose will that is the final end (of creation) which at the same time can and ought to be man’s final end” (*Rel.* 6:6/5-6).<sup>8</sup> It is thus necessary to take stock of the ‘second’ part of Kant’s ethics<sup>9</sup>, in which the highest good is meant to be gradually brought about in history.<sup>10</sup>

The demands of morality lead Kant to the subject of religion because it is religion that proposes the idea of God as governor of the world. Kant will thus attempt to

---

immortality and God necessary for its fulfillment (32-36). The highest good, however, requires that the individual “remodel” the world because the historical good concerns the particular empirical will of each individual (47; 65). Yovel’s analysis depends on the observation that Kant claims an almost erotic function for reason: reason must be self-motivating (out of the concept of interest) in order to bridge the distance between the finite, autonomous self and the world. I believe that Yovel’s observation is correct, for without the claim that reason realizes itself in the world through an interest, how else could Kant maintain the subject’s autonomy as well as the link between a rational being and the *phenomenal* world?

<sup>8</sup> Paul Ricoeur notices the inherent connection between morality and religion in Kant’s ethics, albeit a religion within the limits of human reason: the practical requirement for the fulfillment of the Highest Good leads inevitably from ethics to religion due to religion’s hope of immortality (Ricoeur 412). Ethics does not depend on religion for its content; rather, the ethical requirement for the fulfillment of the demands of the moral law necessitates the postulates of immortality and God so that a finite creature may rationally hope in fulfilling the demands of the law: “Such is the first origin of the question: what can I hope for? It is situated at the heart of moral philosophy, engendered itself by the question: what must I do? Moral philosophy engenders the philosophy of religion while, being conscious of obligation, is added the hope of fulfillment” (Ricoeur 412).

<sup>9</sup> Robert Louden refers to Kant’s ethics as a “two-tiered approach”, the first tier referring to the ‘pure’ ethics of the moral law, while the second tier—referring to ‘impure’ ethics—deals with the anthropological and empirical elements of human nature and human society (Louden, preface, x). Louden claims that this second part, which is “about making morality efficacious in human life” (Louden 13), “has yet to be investigated by Kant scholars” (Louden 6). I do not support Louden’s last claim, since Rogozinski had published a book on this subject a year before Louden and Ricoeur had already hinted at the importance of the *Religion* in Kant’s ethics several decades before Louden’s publication. However, as will become apparent in my thesis, I support Louden’s view (along with that of several other scholars, such as Allen Wood) that an accurate interpretation of Kant’s ethics must take into account the “impure part” of the Kantian system.

<sup>10</sup> Yovel presents the most extensive and cogent account of how the highest good leads to Kant’s writings on history and religion. Since reason governs itself by interests—reason attempts to realize its activity in the world—it follows that practical reason must seek to realize its own end in the world through self-legislation accomplished through the moral law. The highest good thus becomes fully instantiated in the political and social institutions of a historical community, a process that had its beginnings in the *Grounding* in the notion of a ‘Kingdom of Ends’. Yovel concludes that the “duty to promote the highest good no longer means that a person should make himself good, but that he should also make the *world* good” (Yovel 74). See also Yovel 70-77.

circumscribe the practice of religion within the limits of reason alone<sup>11</sup> in order to use a rational representation of God for the sake of providing the grace necessary to fill the gap between the strivings of the finite rational being and the infinite moral law.

it is noticed that the moral law demands the fulfillment of the highest good, then the historical works become more important for understanding Kant's ethics, for they deal with the challenge of instantiating the highest good in history. If, however, the moral law 'materializes' itself in the world, then Kant's later writings, such as the *Religion*, "witness to the limitation, to the incompleteness of the first ethics and lead [the first ethics] to its truth" (Rogozinski 228. My translation).<sup>12</sup>

## **What is Evil?**

### *A priori Deduction of Evil*

If the will is supposed to heed the demands of the moral law for the sake of the law, and if the moral law is able to subjectively influence the will through the feeling of respect, then the will needs no other influence but the moral law to perform an act *from* duty.<sup>13</sup> But even a cursory glance at experience reveals countless cases of people acting against their duty: "That 'the world lieth in evil' is a plaint as old as history" (*Rel.* 6: 19/15); "that such a

---

<sup>11</sup> Jean-Louis Bruch best points out the novelty of Kant's methodology in the *Religion*. Unlike Kant's earlier works, in which morality was deduced *a priori* from concepts, the *Religion* demonstrates an attempt to circumscribe an already existing practice within the limits of reason. Kant, in other words, is not deducing an ideal religion from reason; rather, he attempts to determine which elements of historical religion can conform to the standards of practical reason (see Bruch, chapter 1, especially pages 30-39).

<sup>12</sup> The recognition of the moral law as requiring the instantiation of the highest good in the world also depends upon a certain reading of the law. The law may be completely undetermined, as Rogozinski argues, but it nonetheless is not completely distinct from any form of materiality. It is distinct from matter understood as sensible; it is not distinct from matter understood as pure matter. The law produces the objective final end (the highest good) as its matter because the matter of the law is still distinct from *phenomena*. "*Forma dat esse rei*" (the form gives being to the object) is the phrase through which the moral law should be interpreted. See Rogozinski p. 225-227.

<sup>13</sup> For Kant's distinction between acting in conformity with duty and acting from duty, see *G* 4: 397-99; 406-407 and *KpV* 5: 117.

corrupt propensity [for evil] must indeed be rooted in man need not be formally proved in view of the multitude of crying examples which experience *of the actions* of men puts before our eyes” (*Rel.* 6: 32-33/28). Kant’s examples of “murder dramas enacted in Tofoa, New Zealand” or of “unending cruelty...in the wide wastes of northwestern America” (*Rel.* 6: 33/28) or, finally, of the corruption observed by a member of the British Parliament (“Every man has his price, for which he sells himself” *Rel.* 6: 38/34) point to the fact that the will is regularly led astray from the pursuit of good acts. If nothing in the world can be regarded as “good without qualification, except a *good will*” (*G* 4: 393), then it appears as though we rarely catch a glimpse of unqualified goodness, for experience testifies more readily to the *absence* of a good will. If the analysis of duty set forth in the *Grounding* and in the *Critique of Practical Reason* had firmly established the autonomy of the will, then it appears from the opening lines of the *Religion* that the will of human beings regularly prefers the heteronomy of the inclinations to the autonomy of the moral law.

Kant places two opposing interpretations of experience side-by-side at the beginning of the *Religion*, both in response to the question of whether or not human beings are evil by nature. On the one hand, the moral corruption evident in world-events indicates a constant moral degradation over the course of history; this interpretation is the oldest and most common of all, finding expression through religious doctrines such as original sin, whereby an idyllic beginning gives way to an increasingly corrupt present (*Rel.* 6: 19/15). On the other hand, a small group of philosophers argues that the world is heading in the opposite direction: the moral conduct of human beings has gradually improved over the course of history, which suggests that education can continue “to encourage the sedulous cultivation of that seed of goodness which perhaps lies in us” (*Rel.* 6: 20/16). Experience, however, is an

unreliable guide for interpreting the moral conduct of a human being, for the empirical observation of a person's acts does not reveal the nature of the maxim on which the person acted. Experience is particular in nature, and so precludes the possibility of arriving at a *certain* judgment concerning the goodness or evil of a maxim (*Rel.* 6:20/16). The strategy, then, requires inferring *a priori* a malevolent intention from the observation of several or even one evil act (*Rel.* 6: 20/16).

The first of several paradoxes appears at this point. The question, as it is commonly presented, asks whether or not human beings are evil *by nature*. Kant, however, cannot permit the source of moral corruption to lie in a natural cause, since the chain of natural *phenomena* remains forever determined according to the categories of space and time. Morality and natural causes are antithetical to one another, since natural causation precludes the freedom presupposed by morality (*Rel.* 20-21/16). Fortunately, for Kant, the common meaning of the word 'nature' does not refer to the question of whether or not human beings are evil by nature! The moral sense refers to the "subjective ground of the exercise (under objective moral laws) of man's freedom in general" (*Rel.* 6: 21/16). And since the "subjective ground" of freedom cannot lie in a foreign object determining the will, it must lie in the act of the will adopting a reason for acting—in other words, a maxim (*Rel.* 6: 21/17). It is impossible, however, to inquire into the nature of the ground for the adoption of evil maxims, for if it were possible to discover a ground for the ground of the adoption of evil maxims, then an infinite regress would occur (*Rel.* 6: 25/20-21). Kant concludes that we must remain content with the fact—deduced *a priori*—that there lies in human beings an "inscrutable" ground that causes the will to accept evil maxims. Kant thus concludes his *a priori* deduction of evil in the following paradoxical, yet consistent sentence:

But since the ultimate ground of the adoption of our maxims, which must itself lie in free choice, cannot be a fact revealed in experience, it follows that the good or evil in man... is termed innate only in *this* sense, that it is posited as the ground antecedent to every use of freedom in experience (in earliest youth as far back as birth) and is thus conceived of as present in man at birth—though birth need not be the cause of it. (*Rel.* 6: 21-22/17)

a) *Moral Rigorism: the Re-emergence of Real Opposition*

Though Kant has not yet provided a definition of evil, he can conclude that human beings are evil *by nature*. Common experience, however, seems to suggest otherwise: human beings are both good *and* evil, for it is quite rare to conclude from *every* individual act the presence of evil maxims. Kant, in fact, holds great respect for the opinions of the common man since, in practical matters, the common understanding is just as likely—perhaps even more likely!—to judge correctly concerning proper moral conduct.<sup>14</sup> Surprisingly, on the topic of evil, Kant does not ratify the judgment of the common understanding. Ethics, rather, demands clarity, avoiding “anything morally intermediate, whether in actions...or in human characters; for with such ambiguity all maxims are in danger of forfeiting their precision and stability” (*Rel.* 6: 22/18).

The correct approach in answering the question of whether or not human beings are evil by nature is rigorism: human beings are either good or evil, never both. In defence of his claim, which at first glance appears to contradict his previous laudations of the common man, Kant pens an important footnote outlining the logic of real opposition.<sup>15</sup> The opposite of goodness (the not-good) results either from an absence of goodness or from a positive force opposing goodness. The second scenario is described by Kant as “positive evil” (*Rel.* 6: 22, note 43/18). If the moral law were not a motivating force, then the absence of good

<sup>14</sup> See all of *G* 4: 404 for Kant’s praise of the common understanding in ethical matters.

<sup>15</sup> See *Rel.* 6: 22-23/18, note 43/1.

maxims could be explained as a privation: good fails to exist where it should due to the natural inertia of matter.<sup>16</sup> If, however, the moral law were a motivating force, then the absence of goodness must result from a force acting against the force of the moral law. And since the moral law *is* a motivating force, “the absence of agreement of the will with this law (=0) is possible only as a consequence of a real and contrary determination of the will, i.e., of an *opposition* to the law, = —a, i.e., of an evil will” (*Rel.* 6: 23, note 43/18). Human beings cannot be both good and evil by nature, for the positive influence of the moral law on the will precludes cohabitation with evil.

The influence of the moral law or of evil is not considered, by Kant, as a force pushing or pulling the will in a deterministic fashion.<sup>17</sup> The rigoristic position goes hand in hand with the Incorporation thesis: the freedom of the will (*Willkür*) requires that an influence determine the will “to an action *only so far as the individual has incorporated it into his maxim*” (*Rel.* 6: 24/19. Italics in text.). Since the moral law is considered “in the judgment of reason” (*Rel.* 6: 24/19) as an incentive, it follows that, in cases where the law does not determine the will, a contrary incentive must have been adopted into the maxim of the will (*Rel.* 6: 24). According to the logic of real opposition, the influence of evil is necessary *because of* the moral law; in other words, the absence of goodness as resulting from a force acting contrary to the law is only possible because of the positive force of the law. Without the positive force of the law, goodness would not only be absent from the will; the will would be wholly evil, a position that Kant rejects in qualifying evil as *radical*, not absolute (*Rel.* 6: 32, 37/28, 32). It is, however, only because the moral law is understood as a

---

<sup>16</sup> See my discussion of the difference between Kant and Leibniz on the issue of causation and evil, chapter 2, pgs. 44-47.

<sup>17</sup> See my discussion of Zinkin’s article in chapter 2, p. 44-45.

force that the opposing force of evil is necessary in order to explain the absence of goodness. Evil thus accompanies the moral law; human beings are evil by nature *because of* the law, since Kant has understood the law as a negative magnitude.

b) *How should we interpret Kant's reference to negative magnitude*

Kant's explicit reference to the logic of real opposition, after a long period of dormancy, renders interpretation more difficult. Does Kant use the logic of real opposition to prove that evil exerts a real force on the will? Or does Kant recall the logic of real opposition in order to provide an image for the resistance of evil, without claiming that evil exerts a real force on the will? Andrews Reath and Henry Allison, for instance, interpret Kant's use of negative magnitudes in ethics as providing a sufficient reason for action and not a direct influence on the will.<sup>18</sup> If, however, negative magnitude is merely a form of sufficient reason, then it is hard to see why Kant would take such effort to use the image of negative magnitude in the first place. Why not just describe radical evil as a lack of sufficient reason in adopting a maxim to act?

Melissa Zinkin argues, on the other hand, that Kant's use of imagery suggesting force when discussing respect should be taken at face value: Kant thinks of respect as the force exerted by the moral law on the will.<sup>19</sup> The same interpretation would have to be adopted for the case of radical evil. Zinkin nonetheless concludes that, despite the literal use of force to describe respect, force should not be understood as an attraction 'pushing or pulling' the will. Zinkin seems to adopt two perspectives for interpreting respect: the first interprets Kant's use of force as in a literal manner; the second preserves the freedom of the will by

---

<sup>18</sup> See my discussion of Allison and Reath: chapter 2, p. 36-38.

<sup>19</sup> See my analysis of Zinkin's article on respect: chapter 2, p. 44-45.

claiming that the force in question does not exert an external attraction on the will. The first lines of Zinkin's conclusion indicate the tension between these two perspectives:

It should now be clear that taking literally Kant's use of terms that refer to force is essential for understanding his theory of moral motivation and what it means to have a good will. But by speaking of force, Kant is not referring to a stronger 'force' that the moral law would have to exert against the 'force' of the inclinations that push and pull the will. The will is not something that can be moved by an inclination, or by the moral law, in this way. It is instead practical reason, which can take inclinations to be *reasons* for acting. Still, the will can have its own inner strength, to varying degrees of magnitude, depending on the extent to which it can be exercised" (Zinkin 50).

The dichotomy between understanding negative magnitude either as a sufficient reason or as a literal force is unnecessary, for Kant's use of the concept of negative magnitude in the *Religion* directly precedes the passage in which he defends the freedom of the will according to what has been dubbed the 'Incorporation Thesis'.<sup>20</sup> Kant maintains that any incentive—whether it be the moral law or the inclinations—determines the will to a particular act “*only so far as the individual has incorporated it into his maxim*” (*Rel.* 6: 24/19). It follows that the influence of radical evil—even though understood through the logic of negative magnitude—does not exclude the freedom of the will. The fact that Kant employs the language of negative magnitude alongside the 'Incorporation Thesis' does not require that negative magnitude be diluted to sufficient reason.

The claim that an incentive can determine the will only upon being taken up into a maxim (*Rel.* 6: 24/19) depends upon a distinction between two conceptions of the will articulated for the first time in the *Religion*. *Wille* can be defined as the will legislating the demands of the moral law; and since the moral law stands above the chain of determined

---

<sup>20</sup> For Allison's discussion of the Incorporation Thesis, taking as its starting point *Rel.* 6: 24/19, see *Kant's Theory of Freedom*, p. 40.

*phenomena* characteristic of nature, *Wille* is free due to its obedience to the moral law. *Willkür*, on the other hand, can be defined as the capacity of the will to choose between opposing inclinations. The distinction between *Wille* and *Willkür* allows Kant to maintain unopposed the constant influence of the moral law on the will (which, Kant had already claimed, was a ‘fact of reason’) with the capacity of the will to choose against the law (*Willkür*). The distinction is important in order for Kant to maintain the possibility of autonomy and to account for the possibility of deviation from the law: if the will must always be able to legislate the moral law (*Wille*), then the will would not be able to choose against the law if there were not another aspect to freedom (*Willkür*). But evil implies the possibility of choosing against the moral law. Therefore, another aspect of freedom must be possible, distinct from the will’s capacity to legislate the moral law. The challenge, for Kant, in confronting the problem of radical evil, will be to prove that *Willkür*, corrupted by the influence of radical evil, can move *itself* to heed the moral law.

Michaël Føessel provides an interpretation of negative magnitude that navigates between positing the force of the law as “ontological” and claiming that it consists of a reason for action. Føessel recognizes that the use of negative magnitude to explain the positive influence of evil does not commit Kant to attributing substance to evil (Føessel 96). The use of negative magnitude, according to Føessel, acts as a “quasi-phenomenological approach for the choice in favour of evil” (Føessel 96. My translation) The language of real opposition illustrates the reciprocity between evil and the moral law, since the free will is able to choose evil *because* it is able to choose good:

The rapport between evil and the law is a reciprocity without a necessary implication: *in fact, there is evil because there is the law...* Evil as ‘negative magnitude’ is none other than evil not as it results from the liberty of indifference,

but from an intelligible act by which the free will (*l'arbitre*) renders itself indifferent to the law (Föessel 97. My translation).<sup>21</sup>

The influence of evil is always present; it may be outweighed by the influence of the moral law,<sup>22</sup> but the presence of a good influence implies the opposing influence of an equally ‘positive’ ground of evil. Kant, nevertheless, maintains the freedom of the will because the influence of negative magnitude is not a substance; it is a logical structure helping to explain the co-implication of evil with good—the influence of evil can never be completely corrupting *because* we always retain a ‘seed of goodness’, nor can we ever become entirely free of the influence of evil.<sup>23</sup>

c) *Definition of Evil: Subordination of the Incentives*

As a finite rational being, the human being feels the incentives of both reason and the inclinations. Furthermore, human beings manifest a capacity to choose the good and a capacity to choose evil. The problem, for Kant, lies in finding the source of evil.

Human beings, by nature, are predisposed to goodness. Kant notes three levels in the disposition to follow the good principle: the predisposition to animality, whereby the human being is predisposed to fulfill his physical needs through self-preservation and the sexual instinct; the predisposition to humanity, whereby the human being’s instinct of self-love brings him into comparison with other human beings, making him fit for life in society; and the predisposition to personality, whereby the human being is properly considered a *person*, a being capable of “simple respect for the moral law” (*Rel.* 6: 27/23). Only the third

---

<sup>21</sup> See also Rogozinski, pg. 274: “evil exists *because* there is the law”.

<sup>22</sup> In the ‘Analytic’ of the second Critique, Kant uses the image of counterweight to compare the hindering effect of the inclinations on the moral law: see *KpV* 5: 75-76; 88.

<sup>23</sup> See also Rogozinski, p. 274-275 for the co-implication of evil with good required by the logic of negative magnitude.

predisposition cannot become entirely corrupt, for it is impossible for a rational being to destroy his own reason. (*Rel. 6: 35/31*).

Human beings also manifest three degrees of the propensity to evil, which correspond to the three levels of the predisposition to good. First of all, human nature is weak because human beings regularly succumb to the subjective incentive to follow the influence of the inclinations (*Rel. 6: 29/25*). Kant expresses this unfortunate truth through the complaint of Saint Paul: “What I would, that I do not!” (*Rel. 6: 29/25*). The second degree of the propensity to evil consists of “the impurity of the human heart” (*Rel. 6: 30/25*): human beings adulterate the pure motive of the moral law by mixing respect with spurious motives stemming from the inclinations (*Rel. 6: 30/25*). Alas, *duty for duty’s sake* is rarely the aim of human beings! Finally, the “corruption of the human heart” (*Rel. 6: 30/25*) illustrates the worst propensity to evil in human nature: the will downright neglects “the incentives springing from the moral law in favour of others which are not moral” (*Rel. 6: 30/25*). The third degree of evil downright “reverses the ethical order [of priority] among the incentives of a *free will*” (*Rel. 6: 30/25*). The third propensity, however, would imply the complete corruption of reason, a choice that Kant could not accept as a definition of evil because reason must always remain capable of following the moral law.

Kant claims that evil cannot lie in the senses or in reason, since the senses do not permit freedom and since reason cannot legislate universal maxims against itself—an act that would be remarkably un-universal (*Rel. 6: 35/30*). A human being cannot “repudiate the law in the manner of a rebel....The law, rather, forces itself upon him irresistibly by virtue of his moral predisposition” (*Rel. 6: 36/31*). Human beings retain a ‘seed of goodness’ (*Rel. 6: 20/16*) because the moral law is simply a ‘fact of reason’. That evil cannot be attributed to

the senses appears unsurprising, since their *phenomenal* nature precludes the possibility of free moral action. That Kant could not situate evil in reason, permitting a total corruption of the predisposition to personality, is surprising given his positing the existence of evil through the logic of real opposition (negative magnitude). For, if evil exists, it exists in virtue of the moral law; an opposing force, in other words, only exists insofar as the first force acts on the object. Kant could not admit this possibility, despite the fact that the logic of real opposition—which creates the problem of evil in the first place—leads to it. For reason itself—the legislator of the moral law—cannot be held responsible for the problem of evil, a scenario that would also involve an incapacity to overcome evil and to be motivated solely by the moral law.

The result is that Kant surreptitiously renounces his initial claims that evil cannot be attributed to the senses; unlike his initial claim, Kant now argues that evil may attach itself only to the first two predispositions, not to the third (*Rel.* 6: 35/31). *Why* a human being may not legislate against his own reason, however, is never answered. Yet, an answer seems necessary, for if the problem of evil is the result of negative magnitude, then the opposing forces cannot be heterogenous from each other, otherwise the presence of one force could not imply the presence of another (as real opposition requires). If this were the case, however, then evil would have to be attributed to reason, a possibility that Kant repudiates several times in the *Religion*, most notably when he claims that the disposition to personality cannot become entirely corrupt (*Rel.* 6: 35/31).

Kant thus backtracks on his initial claim and argues that evil is the result of the human being's sensuous nature, a move that brings forth the logic of privation that had been abandoned years before. Human beings, due to their sensuous nature, are 'innocently

predisposed' (*Rel.* 6: 36/31) to adopt the influence of the senses into their maxim for action. Human beings do not consider the influences of the senses as "*in themselves wholly adequate*" (*Rel.* 6: 36/31) for determining the will because, through reason, human beings always retain a 'seed of goodness'. Human beings thus adopt *both* sorts of influence—through the moral law and through the senses—into their maxim, which amounts to succumbing to the second propensity to evil. But a human being cannot both be good and evil, since the logic of real opposition demands a rigorous separation of the two qualities.<sup>24</sup> Evil, therefore, must consist of 'subordination' of the incentives of the moral law to the incentives of the senses:

"[the human being] adopts, indeed, the moral law along with the law of self-love; yet when he becomes aware that they cannot remain on a par with each other but that one must be subordinated to the other as its supreme condition, he makes the incentive of self-love and its inclinations the condition of obedience to the moral law" (*Rel.* 6: 36/31-32).

Evil is even more insidious than a conscious subordination of maxims: many people *unwittingly* subordinate the moral law to the senses, convinced they are acting morally, yet unaware of subjectively disregarding the supreme principle of moral goodness. Such people "picture themselves as meritorious, feeling themselves guilty of no such offenses as they see others burdened with" (*Rel.* 6: 38/33); they even project a false image of moral conduct, leading to "falsity and deception of others" (*Rel.* 6: 38/33).

Evil, in other words, does not consist of wilful malice, of a positive refusal to follow the moral law that would correspond to the third propensity to evil. Rather, evil is an unconscious *privation* of goodness: the maxim of the will *lacks* the universal form necessary

---

<sup>24</sup> Hence Kant's adoption of "rigorism" and the subsequent appeal to the logic of negative magnitude to justify his position. See *Rel.* 6: 23/18, note 43/1.

to qualify for moral action according to the standard of the moral law. The logic of real opposition that created the problem of evil in the first place has been replaced by the logic of privation in the definition of evil. As is normally the case in philosophy, the solution to a problem depends on the definition of the problem itself. In defining evil as subordination, Kant has opened the door to a solution through reversion that re-subordinates the incentives of the inclinations to the incentives of the moral law. The logic of privation, in other words, precludes the possibility that evil will reappear each time the will manifests the possibility of following the moral law for its own sake. The cost of finding a solution, however, comes at the price of inconsistency, for Kant quietly abandons the logic of real opposition that created the problem of radical evil in the first place.

### **Consequences of Defining Evil as Subordination**

Kant's definition of evil as subordination has been clearly noticed and regularly commented on by scholars. Olivier Reboul, for instance, recognizes that the "subordination" of the incentives of the moral law to the incentives of the senses is that "which constitutes radical evil" (Reboul 97). For Reboul, the subordination described by Kant is characteristic of a lie,<sup>25</sup> since radical evil is actually a wilful consent "to fragility and above all to impurity" (Reboul 99). A human being stands guilty before the moral law because "he knows that he *could* be virtuous and that he *could*, through his own powers [*de lui-même*], deliver himself from evil at any time" (Reboul 160). Radical evil, according to Reboul's interpretation, always depends upon the will's choice for ignorance. Likewise, Henry Allison correctly notices that evil cannot lie in practical reason or in *Wille*. Evil, therefore, can only lie in *Willkür*, which gives "priority to the nonmoral incentive or, equivalently, the principle

---

<sup>25</sup> Jean-Louis Bruch characterizes evil in the first book of the *Religion* as a "*tromperie*" (deception), an unwitting capacity to trick oneself into subordinating the moral law to the inclinations (Bruch 71).

of happiness, even in those cases where it conflicts with the dictates of morality” (Allison 151). Our nature as sensible creatures tempts us to “subordinate moral considerations to our needs as sensuous beings” (Allison 157). Carnois observes that evil, for Kant, means that a human being “no longer obeys the moral law except to the extent that such obedience is useful to satisfy his sensuous [*sensible*] inclinations” (Carnois 171). Radical evil results in “reason itself subordinating itself to the sensuous” (Carnois 172).<sup>26</sup> Finally, Robert Louden concludes that “all three levels of the propensity to evil involve a reversal of priorities, insofar as agents who exhibit each kind of evil have non-moral rather than moral reasons as the sufficient motive of their actions” (Louden 138).

The previous interpretations notice quite rightly that Kant characterizes evil as subordination. The same commentators, however, do not notice that Kant’s definition of evil depends upon a logic that differs from the one that gave rise to the problem of evil. Kant had to avoid drawing the ultimate conclusion produced by the logic of real opposition: that reason itself—through its capacity to legislate the moral law—is the reason for the existence of evil. Evil, therefore, must consist of a privation of goodness: only the first two predispositions to good can result in evil; the human being always retains his “personality” (*Rel.* 6: 28/23) and the complete corruption of reason is deemed impossible, for reason cannot contradict itself (*Religion*, p. 30/25). But, if evil consists of an unwitting subordination of the law to the inclinations, then there does not seem to be anything particularly bad about evil at all; for, that a person may *fail* to perform his duty for duty’s sake does not appear particularly malicious, since the dutiful act has still been performed.

---

<sup>26</sup> Zinkin’s interpretation, in the conclusion of her article, of the failure of the will to follow the moral law as the result of ‘weakness’ also runs along the same lines as the interpreters above.

Avoiding the corruption of reason by dropping the logic of real opposition ends up leading Kant to propose a definition of evil that does not correspond to either the horrific examples of evil proposed by Kant in book 1 or that we observe in everyday life. Jacob Rogozinski notices the consequence of Kant's definition of evil:

As long as the thesis of a real opposition of evil to the Law, which, in the *Essay* [on radical evil], establishes the concept of radical evil, is surreptitiously identified with the hypothesis of diabolical evil, any real opposition of an evil will to the Law will be declared impossible and we then return to the conception of the *Grounding*, in which evil is nothing but a *logical contradiction*, the particular transgression of a universal Law for which the guilty individual 'recognizes the authority [of the Law] while nevertheless violating it.' From the beginning, the dice were loaded: we are left with nothing but a choice between *Bosheit* and *Bösartigkeit*, between the 'impossible' perspective of a diabolical rebellion and the reduction of evil to this ordinary human badness [*méchanceté*] that has nothing really bad about it" (Rogozinski 283. My translation).

Kant's surreptitious slip from a logic of real opposition into a logic of privation allows him to evade the necessary conclusion that the will can become entirely corrupt without hope. Since evil is characterized by Kant as subordination, he can begin work on a type of 'conversion' for the individual from bad to good. The conversion from evil to good forms the subject of the second book of the *Religion*. However, with the re-emergence of radical evil in a new form at the outset of book III, Kant's efforts in book II become a moot point and the work of helping human beings overcome the social aspect of evil will remain Kant's preoccupation for the rest of the *Religion*.

#### Chapter 4: Moral Training and the Need for Non-Moral Incentives

Kant assures the reader that there is “no greater advantage than freedom from the *sovereignty of evil*” (*Rel.* 6: 93). The definition of radical evil as subordination—and its concomitant abandonment of the logic of real opposition—assured Kant of a way to avoid the consequences of negative magnitude by simply qualifying evil as a reversal of incentive. The door should have been open to re-establishing the autonomy of the will through re-subordination of the incentives. Book III, however, unsettles Kant’s assurance of an autonomous *Willkür*, since evil re-emerges in the form of other people<sup>1</sup>—prompting a more ‘communitarian’ solution to the latest face of evil. Since the presence of other people incites the individual to feel greater motivation from the inclinations rather than from the moral law, a new kind of community will be proposed to remedy the harmful influences of other people. The task of establishing this new community is not a simple chore for people easily inclined to moral corruption. Arriving at a community of individuals united in their respect for the law will involve using contingent, historical institutions to train individuals into feeling respect for the law. In other words, the opposite of the law—contingent institutions and their practices—is an integral part of learning to respect the law alone.

---

<sup>1</sup> The ‘social dimension’ of evil corresponds to a shift in emphasis in Kant’s *Religion*. The analysis for the first two books of the *Religion* focused mainly on the individual’s penchant for evil and the possibility of his conversion. The last two books, however, take on a communal focus: book III diagnoses evil as resulting from the presence of other people, and the latter part of book III and the rest of book IV attempt to propose a solution to evil through involvement in another type of community—the ethical commonwealth. For the two stages of analysis in the *Religion*, see Gordon E. Michalson “The Problem of Salvation in Kant’s *Religion within the Limits of Reason Alone*”, especially pages 321-324.

### **The Social Character of Evil: a Crisis of Autonomy**

As a social being, the human being always lives in relation to a community. Life in community, however, raises the same problem of subordination discovered in book I, since the presence of another person invites comparison, evaluation, desire, envy:

[Man's] needs are but few and his frame of mind in providing for them is temperate and tranquil. He is poor (or considers himself so) only in his anxiety lest other men consider him poor and despise him on that account. Envy, the lust for power, greed, and the malignant inclinations bound up with these, besiege his nature, contented within itself, *as soon as he is among men.* (*Rel.* 6: 93-94)

The presence of other people does not assume their *intention* to corrupt each other; envy and the subsequent desire to eschew universal morality for the sake of personal gain depends upon the simple presence of another person, whose difference incites rivalry:

and it is not even necessary to assume that these are men sunk in evil and examples to lead him astray; it suffices that they are at hand, that they surround him, and that they are men, for them mutually to corrupt each other's predispositions and make one another evil. (*Rel.* 6: 94)

If the mere presence of another person corrupts our predisposition to good, then morality cannot consist of the strength of an *individual* to resist temptation from his rivals, for the individual does not have such strength in a community. Presuming to legislate the moral law to oneself amounts to living in an "ethical state of nature", which needs to be overcome if human beings are to escape the corrupting influence of envy.

The ethical state of nature, however, appears to correspond closely to the vision of autonomy portrayed in the *Grounding* and in the *Critique of Practical Reason*: "In both [the juridical state of Nature and the ethical state of Nature] each individual prescribes the law for himself [*gibt...sich selbst das Gesetz*], and there is no external law to which he, along with all others, recognizes himself to be subject" (*Rel.* 6: 95). The problem of evil, however,

precludes the possibility of absolute autonomy because the individual in society would choose to ignore the influence of the moral law even though he retains the capacity to prescribe it to himself. The ethical state of nature is one in which

Men...mutually corrupt one another's moral predispositions; despite the good will of each individual, yet, because they lack a principle which unites them, they recede, through their dissensions, from the common goal of goodness and, just as though they were *instruments of evil*, expose one another to the risk of falling once again under the sovereignty of the evil principle. (*Rel.* 6: 97)

Human beings *ought* to leave the ethical state of nature—a state corresponding to absolute autonomy—because the presence of other people assures their incapacity to perform duty for duty's sake.<sup>2</sup> If the ethical state of nature is insufficient to support a person's predisposition to good because the members of society are not united according to the common goal of following the moral law, then they must choose to enter an ethical commonwealth, which consists of “a union of such individuals into a whole toward the same goal” (*Rel.* 6: 97).<sup>3</sup>

The insufficiency of an ethical state of nature—according to which an individual prescribes the moral law to himself—marks a change in Kant's ethics. If, in an ethical state of nature, the individual cannot resist subordinating the influence of the moral law to the influence of the inclinations, then the individual requires the help of a certain form of community to perform duty for duty's sake. The ‘Analytic’ of the *Critique of Practical Reason* consistently attacked traditional forms of morality precisely with the claim that they made respect for the moral law conditional upon the fulfillment of another principle (such as the good or happiness). If *Willkür* requires the influence of an ethical commonwealth to

---

<sup>2</sup> Kant's observation in the *Religion* of the corrupting influence of other people runs parallel to his claim in the *Anthropology* that people never appear as they truly are as soon as they realize that they are being watched (*A.* 7: 121). Performing duty for duty's sake becomes difficult in community because it can never be determined whether a person has truly acted dutifully *out of respect for the moral law*.

<sup>3</sup> “eine Vereinigung derselben in ein Ganzes zu eben demselben Zwecke” (*Rel.* 6: 97/89).

choose the proper order between the moral law and the inclinations, then it might appear that Kant's initial outline of autonomy in the 2<sup>nd</sup> *Critique* comes under pressure from the reality of radical evil. For Rogozinski, the *Religion* represents

*a crisis of autonomy*, since the possibility of triumphing over evil by oneself is at once reaffirmed and compromised... It should be deemed highly significant that Kant here defines a state of nature, whether it be juridical or ethical, as one where "each individual prescribes the law for himself, and there is no external law to which he, along with all other, recognizes himself to be subject."<sup>4</sup> The same power to give to oneself the Law, which seemed to define the autonomy of the Subject, now characterizes the state that is most exposed to evil: from the point of view of the *Religion*, such autonomy is radical evil.<sup>5</sup> (Rogozinski 292).

Similar to Rogozinski, Sharon Anderson-Gold observes that the introduction of evil into Kant's ethical framework rules out an individualistic view of autonomy: "models of moral perfection which view the achievement of virtue as a result of 'individual' decision making concerned only with the 'purity' of intention, detached from all 'purposes' are rendered futile by Kant's view of the human condition in *Religion*. To maintain such a posture is to remain in an 'ethical state of nature'" (Anderson-Gold 125).<sup>6</sup>

Despite the pessimistic outlook that the rediscovery of radical evil may present for moral autonomy, Kant proposes a new solution in the form of the ethical commonwealth, which maintains the possibility of autonomy and respect for the law through a gradual transition from the historical churches to the pure religion of reason.

---

<sup>4</sup> Translation taken from *Rel.* 6: 95/87.

<sup>5</sup> The bulk of Rogozinski's analysis in this passage appears correct to me; however, he exaggerates slightly near the end of this passage, for the autonomy characteristic of the 'ethical state of nature' leads to radical evil (the subordination of the law to the inclinations) but cannot be considered radical evil itself.

<sup>6</sup> On the novelty of the concept of an ethical commonwealth for Kant's ethics, see Jean-Louis Bruch *La philosophie religieuse de Kant*, p. 161-62, 167; and Sharon Anderson-Gold "God and Community: An Inquiry into the Religious Implications of the Highest Good", p. 126-127.

### *The Ethical Commonwealth*

#### a) *Need for an Ethical Commonwealth*

Although the introduction of the ethical commonwealth places a new concept at the center of the fight to save the autonomy of *Willkür*, the scaffolding on which the concept is built has remained the same. The ethical commonwealth is deliberately placed in tandem with an important ethical concept of the *Critique of Practical Reason*: the Highest Good. Attaining the unity of virtue and happiness—the Highest Good—is specified by Kant as a social—and therefore historical—goal (*Rel. 6: 97*). Since the corrupting influence of other people prevents the individual from reaching the highest good on his own, the moral duty to attain the highest good requires the creation of an ethical commonwealth to mitigate the nefarious attraction of evil (*Rel. 6: 97-98*). The ability to create an ethical “whole”, however, cannot be guaranteed on the basis of the finite volition of human beings; the Idea of a “higher moral Being” is necessary so that “the forces of separate individuals, insufficient in themselves, are united for a common end” (*Rel. 6: 98*). The ethical commonwealth, then, stands as the historical representation—or schema—of the highest good first presented in the *Critique of Practical Reason*. The formal viewpoint of the 2<sup>nd</sup> *Critique* and the historical viewpoint of the *Religion* consists of a difference in tone; the first represents the ‘formal’ stage of Kant’s ethical system, while the second represents the ‘material’ stage, in which society is remodelled through reason according to the universal *a priori* demands of the moral law:

The material imperative becomes a source of positive initiatives, enriching the concept of freedom. It no longer suffices to respond to circumstances; one has to change them in light of an *a priori* moral scheme, thereby creating new orders and systems, such as moral education, free political institutions and a “rational” church. (Yovel 47).

The ethical commonwealth has religion as its form, for the act of conceiving of “all duties as divine commands”<sup>7</sup> requires, through the supernatural aid of God, the unification of all rational beings obedient to the moral law. The divine being—whose actual existence remains unknown to finite rational beings—unites the diversity of human beings into the state of religion, gradually achieved through the historical churches. God’s role is important only insofar as he makes up for the incapacity of human beings to accomplish the duty of uniting into an ethical commonwealth.<sup>8</sup> The *can* demanded by the *ought* of the moral law must be performed apart from the influence of God: “[the ethical community], according to Kant, is the concrete historical state of mankind, and in it, as we have seen, it is man himself who has the duty to act while God’s existence only guarantees his capacity to do so” (Yovel 111).

The problem of the transition from *ought* to *can* lies with the *can*, for the limitation of finite rational beings prevents their fulfilling all demands of the moral law. Yovel is not alone in noticing that God—though not invoked as a heteronomous principle overpowering the autonomous human will—is integral to the historical fulfillment of the Highest Good in the form of the ethical commonwealth. Sharon Anderson-Gold perceptively notes that, in the *Religion*, the will (*Willkür*) needs the social union of the ethical commonwealth:

---

<sup>7</sup> This phrase stands as Kant’s definition of religion. *Rel.* 6: 153/142.

<sup>8</sup> The interpretation of Rogozinski on this claim, though accurate in its diagnosis, is not quite so significant in its consequences. Since God must be postulated as actively guaranteeing the unification of human beings into an ethical commonwealth, it follows that the pure autonomy of the will can no longer be considered a tenable thesis, since a divine being is required to guarantee the passage from *ought* to *can*. Rogozinski thus thinks it best to speak of an auto-heteronomy, whereby practical reason postulates a foreign being helping the finite rational being attain the moral end laid out by his reason. Though Rogozinski is right in claiming that the classic definition of autonomy appears untenable once Kant admits the problem of evil into his philosophy, it is an exaggeration to claim that postulating God for the fulfillment of an ethical commonwealth amounts to renouncing the theory of autonomy wholesale (Rogozinski 292-293). I agree, rather, with Yovel, who reminds the student of Kant that the intervention of God remains an insoluble problem for theoretical reason and human beings furthermore always retain the duty to act, while God only guarantees their capacity to do so (Yovel 111). Acting out of duty for duty’s sake remains the premise from which the intervention of God may be deduced.

“the social significance of the propensity to evil is not often noted even though Kant takes up this theme quite explicitly and draws from this predicament the conclusion that unless individuals deliberately unite into a society ‘for the sake of the laws of virtue’ they cannot hope to *remain free* from bondage [from evil]” (Anderson-Gold 125).

Finite rational beings, in other words, are unable to attain moral unity on their own (Anderson-Gold 128-129), a situation which renders the absolute autonomy of the will difficult to maintain and which establishes the ethical commonwealth as a concept integral to the coherence of Kant’s ethics as a whole through victory over the evil principle.<sup>9</sup> The result is a shift in the *Religion* concerning Kant’s thought on God and on the ethical commonwealth: “In *Religion* one finds a different (and a clearly constitutive) relationship between the idea of God and the ethical commonwealth; the unity of the latter and the purity of its principles are seen as derived from the holiness of the will which serves as its ‘foundation’” (Anderson-Gold 126).

Though it is often noted that the historical churches act as vehicles for the attainment of religion (the final stage of the ethical commonwealth), commentators rarely note the ramifications of this concept for the autonomy of the will. Thus, after highlighting the importance of the ethical commonwealth for the historical stage of Kant’s ethical philosophy, Yovel does not go into detail concerning the use of the historical churches as means for moral training, nor into its implications for interpretations of respect and

---

<sup>9</sup> I cannot agree with Reboul’s claim that the “doctrine of the Church is too late, too isolated in Kant’s corpus—it by far lacks respondents (répondants) in his other works—so that we may grant it a first-rate importance” (Reboul 205. My translation.). Although relegated to the *Religion*, the interpretation of the ethical commonwealth by Yovel, Rogozinski, Bruch, and Anderson-Gold, all of whom set the concept of the ethical commonwealth in relation to Kant’s earlier ethical writings, illustrate that the commonwealth belongs to another stage—and not another kind or nature—of Kant’s ethics. Due to the social nature of evil, the ethical commonwealth should be interpreted as a concept for successfully overcoming the challenge of evil. It is thus hard to explain the silence of Henry Allison and Bernard Carnois on the concept of the ethical commonwealth, especially in their books attempting to establish a coherence to Kant’s ‘theory of freedom’!

autonomy. Instead, Yovel restricts much of his discussion to Kant's biblical hermeneutics. Certain passages indicate that Yovel notices the importance of the *Religion* as a text outlining a form of moral education, but he does not go into significant detail.<sup>10</sup> Bruch devotes a whole chapter to the "true and false worship (*culte*)", but generally pursues a historical investigation, setting Kant's thought on the ethical commonwealth and the historical churches in the context of the writings of other Enlightenment philosophers on religion.<sup>11</sup> Anderson-Gold rightly points out the change to Kant's conception of autonomy that is brought about through the concept of the ethical commonwealth. Her article, however, devoted as it is to establishing one particular claim, cannot pursue the ultimate significance of the ethical commonwealth for an ethics of autonomy as it unfolds through the need to overcome the corruption of the historical churches. I will attempt to fill this lacuna in the following pages of this chapter.

b) *Corrupt Human Beings cannot Establish the Ethical Commonwealth*

Corrupt rational individuals must enter an ethical commonwealth for the sake of following the good principle—for performing duty for duty's sake. Despite the consequence for the pursuit of autonomy, the need to enter an ethical commonwealth is unequivocal: the proclivity to ignore the moral law through jealousy provoked by the mere presence of other people receives a salve through a common existence (*ethischen gemeinen Wesen*) in which all strive to fulfill their duty. Though the plan is clear, the execution remains faulty, for corrupt individuals are unable to establish an ethical commonwealth on their own:

---

<sup>10</sup> Yovel, for instance, claims that Kant's *Religion* aims to use "suggestive educational means rather than formal logical arguments" (213) in order to introduce morality among the populace, but he does not perform much textual analysis of book 3 and 4 of the *Religion* in order to demonstrate Kant's argument. See Yovel 201-223 for his discussion of Kant's biblical hermeneutics and the educational training of the historical churches.

<sup>11</sup> See Bruch, chapters 6 and 7.

The sublime, yet never wholly attainable, idea of an ethical commonwealth dwindles markedly under men's hands. It becomes an institution which, at best capable of representing only the pure form of such a commonwealth, is, by the conditions of sensuous human nature, greatly circumscribed in its means for establishing such a whole.<sup>12</sup> How indeed can one expect something perfectly straight to be formed out of such crooked wood? (*Rel.* 6: 100).

Despite their striving, finite and corrupt human beings will never attain the ultimate—and infinite<sup>13</sup>—goal of living in a common existence in which (subjectively) all duties are conceived as divine commands. This incapacity stems from the corruption of evil and from finite human nature (“the conditions of sensuous human nature”). As in the *Critique of Practical Reason*, an all-seeing God must be postulated in order to forgive the distance separating the finite individual and the infinite moral law—but only on the condition of the individual striving to perform duty for duty's sake. The foundation of a “moral people” (*moralisches Volk Gottes zu stiften*) can only be expected from God (*Rel.* 6: 100). God's foundation of an ethical commonwealth, on the other hand, depends upon the individual's moral striving for its own sake: “he must, on the contrary, proceed in such a way as if everything depended on him, and only on this condition may he hope that a higher wisdom will bestow completion upon his well-intentioned effort” (*Rel.* 6: 100-101).<sup>14</sup> The role of

---

<sup>12</sup> The first two sentences in the English translation actually form one sentence in the German. Through a rise and fall in tone, the first part of the sentence illustrates the sublimity of the commonwealth, while the second part illustrates the corruption of human nature: “Die erhabene nie völlig erreichbare Idee eines ethischen gemeinen Wesens verkleinert sich sehr unter menschlichen Händen...” (*Rel.* 6: 100). The rise and fall in tone encapsulates the constant problem for Kant: the lofty ethical ideal of the ethical commonwealth or the moral law can never be obtained by a finite and corrupt humanity—the opposite of the law's purity.

<sup>13</sup> The incapacity of human beings to reach the ethical commonwealth results from the same reason that human beings are incapable of perfectly fulfilling the moral law: human finitude (see the ‘Dialectic’ of the *Critique of Practical Reason*). The solution will be the same in both cases: the all-seeing God—postulated, but never known theoretically—pardons the distance between the finite rational subject and the infinite standard upon observing the subject's intention.

<sup>14</sup> Though I grant that grace is an important concept for the attainment of the ethical commonwealth, I cannot agree with Despland's claim that “grace itself...like revelation, is no parergon at all and is to be believed in with moral certainty” (Despland 231). Despland claims that, once Kant noticed the problem of radical evil, he had to “believe that there is hope for mankind only if there is grace, a general providential goodness that supplements our efforts, and only if there is the availability of symbolic presentations of specific aid at given

God in founding the ethical commonwealth consists of granting the completion of justice, an infinite ideal that a finite being cannot hope to achieve. God, therefore, accords to each individual “whatever his actions are worth (*was seine Taten wert sind*)” (*Rel. 6: 99*). The conclusion to be drawn is that “an ethical commonwealth can be thought of only as a people under divine commands, *i.e.*, as a *people of God*” (*Rel. 6: 99*).

Gathering corrupt human beings into an ethical commonwealth founded by a moral God involves expediently using the instruments that exist for the sake of leading people to the divine: the historical churches. Since the political commonwealth does not require reference to ethics or to the divine, and since the ethical commonwealth requires the positing of God as founder and ruler, then the idea of a ‘church invisible’ serves as the “archetype for what is to be established by men” (*Rel. 6: 101*). Sensuous human beings, however, require a physical representation of the invisible church in order to gradually reach the state of living in an ethical commonwealth (also known as religion). The historical churches, in other words, serve as the visible churches whose purpose is to gather human beings into a community that resembles as closely as possible the invisible church established by God.

The Bible, for instance,—a book containing numerous violent stories antithetical to morality—must be read and interpreted according to the universal standard of morality.

Thus, churches founded on a scripture become the most important vehicles for pure religion,

---

moments of time to bring about our conversions” (Despland 232). Though Despland has noticed the importance of grace, he tends to put the cart before the horse: the hope engendered by the certainty of grace is not necessary for fulfilling one’s duty, a position which Kant could never condone, since it would abnegate his commitment to autonomy. Despland interprets Kant as positing *faith* within the limits of reason alone, which corresponds establishing a new form of religious experience (Despland 145-46; 153). Pace Despland, Kant advocates *religion* within the limits of reason alone, which is not a matter of religious experience, but the completion of the ethical commonwealth, in which all members subjectively conceive of all duties as divine commands (*Rel. 6: 153*). Since Despland does not place enough emphasis on the role of the historical churches in forming individuals into the autonomous state of religion, he cannot help but interpret grace in terms that approach a renunciation of autonomy.

since the written text, interpreted according to the criteria of the moral law, is considered the most useful instrument for the universal dissemination of religion.<sup>15</sup> The particular historical practices of the churches do not retain value in and of themselves; their value stems from their usefulness in leading people to performing duty for its own sake: “the theoretical part of ecclesiastical faith cannot interest us morally if it does not contribute to the performance of all human duties as divine commands (that which constitutes the essence of all religion)” (*Rel. 6*: 110). The goal of the ethical commonwealth—a universal community of members following duty for duty’s sake—requires using a relatively common practice that sensuously inclined human beings have come up with: the diverse practices of historical religion.

### **Historical Churches as Means to the Ethical Commonwealth**

A problem, however, stands out for Kant in leading people to religion through membership in historical churches: the different churches tend to make themselves their own end, exalting historical rites and statutes as the most important element of church practice and thereby failing to act as a schema for the invisible church of pure religion: the historical churches, in other words, are the institutions of heteronomy *par excellence*, training people to follow the will of God through the practice of contingent rites, statutes, and ordinances.<sup>16</sup> Kant, therefore, interprets the value of church practice according to the criteria of the invisible church of pure religion; an historical church is legitimate and indeed valuable

---

<sup>15</sup> See *Rel. 6*: 105-107/96-98.

<sup>16</sup> “Thus it happens that men will regard neither union into a church, nor agreement with respect to the form which it is to take, nor yet *public* institutions, as in themselves necessary for the promotion of the moral element in religion, but only, as they say, for the service of their God, through ceremonies, confessions of faith in revealed laws, and observance of the ordinances requisite to the form of the church. . . . All these observances are at bottom morally indifferent actions; yet, just because they are to be performed merely for His sake, they are held to be all the more pleasing to Him” (*Rel. 6*: 106/97). Similar passages in which Kant deplores the tendency of human beings to make church practice an end in itself can be found throughout books III and IV.

insofar as its contingent—and sometimes odd—practices train people into fulfilling duty for duty’s sake.

a) *Why use the Historical Churches?*

An ethical commonwealth is universal in scope, for the moral law presents itself to all rational beings. Religion, however, is commonly grounded on a particular act of revelation granted to a particular people in a given historical moment. Religion should thus stand as an unworthy tool for the formation of an ethical commonwealth. Kant elides the particularity of religion by noting the distinction between natural and revealed religion. In contrast to revealed religion, which places revelation in order of precedence before duty, natural religion requires that “I must know in advance that something is a divine command in order to acknowledge it as my duty” (*Rel.* 6: 154/143). The subordination of the revealed element to the moral element (known through reason) in religion makes natural religion “*capable of being shared widely with others*” (*Rel.* 6: 155/143. Italics in text). The particularity of revealed or learned religion—which amounts to all historical religion to date—exempts finite rational beings from being duty-bound to enter this type of religion.

If finite rational beings are duty-bound to practice natural religion in order to reach pure religion, and if historical religions are characterized by the contingency of revelation or learning, then how can finite rational beings reach the goal of pure religion through the means of historical churches? The answer lies in interpreting historical religion according to the universal criteria of natural religion. Human beings are composed of both a rational and sensuous nature, natural religion appealing to the first and revealed religion appealing to the second. Universal religions can nonetheless be considered both natural and revealed, as long as human beings understand and practice the revealed element according to the exigencies of

the moral law known through practical reason (*Rel. 6: 155/143*). The appeal of the revealed component of religion to the senses of human beings—through rites and ordinances—retains legitimacy as long as the sensuous elements of religion act as vehicles for introducing the universal, natural element of religion to the members of the religious community:

a revelation thereof at a given time and in a given place might well be wise and very advantageous to the human race, in that, when once the religion thus introduced is here, and has been made known publicly, everyone can henceforth *by himself and with his own reason* convince himself of its truth. In this event, the religion is *objectively* a natural religion, though *subjectively* one that has been revealed; hence it is really entitled to the former name. (*Rel. 6: 155-56/144*. My italics in penultimate sentence.)

Once the revealed element of religion has assisted in the introduction of the natural element of religion, then the rites and practices related to revelation “can be entirely forgotten without the slightest loss to that religion either of comprehensibility, or of certainty, or of power over human hearts” (*Rel. 6: 156/144*). Concession is made to revelation because human nature resembles more “crooked wood” (*krummen Holze; Rel. 6: 100/92*) than sublime morality.

Human beings thus need to be wooed to natural religion through the prestige and mystery of religious rites: “it follows that unless there are added to the natural laws, apprehensible through unassisted reason, certain statutory ordinances attended by legislative prestige (authority), that will still be lacking which constitutes a special duty of men, and a means to their highest end, namely their enduring union into a universal church” (*Rel. 6: 158/146*). Kant asks us to fight fire with fire! Finite and corrupt human beings, readily inclined to follow their sensuous inclinations, can be led to overcome their servitude to the inclinations through appeal to the aura of sensuous rites and statutes. Overcoming evil—the

subordination of the moral law to the influence of the inclinations—requires using the means of evil—the particular rites and statutes of the historical churches—to the advantage of reason.

Kant's strategy is consistent not only with respect to rites and ritual, but also to the key beliefs of the churches. Kant, for instance, interprets the figure of Christ as the "Teacher of the Gospel", laying down a pure natural religion of moral ordinances suited to reason alone, even "in the face of a dominant ecclesiastical faith" based on statutory observance (*Rel. 6: 158/146*). His teachings, however, would be corrupted by later generations, capitulating once again to the lure of rites and ritual. The mistake made by the disciples of the Teacher of the Gospel was, once again, to turn particular rites into an end in themselves. The Teacher, on the other hand, demonstrates the correct stance toward religious ritual: he uses statutory rites and observances as means to introduce among a wider audience the duties which people have the ability to know through their own reason.<sup>17</sup>

The function of sensible rites—and the value of historical religion in general—lies in allowing natural religion to gain universal access:

For what is easier than to take in so sensuously depicted and simple a narrative and to share it with others, or to repeat the words of mysteries when there is no necessity whatsoever to attach a meaning to them! How easily does such a faith gain universal entrance....Such a faith, therefore, is indeed suited even to the commonest human capacities. (*Rel. 6: 181/169*)

With natural religion further entrenched into the lives of individuals, the rites and statutes of religion may gradually be abandoned, since pure religion is the state of living a common

---

<sup>17</sup> Thus, in describing the preaching of the Teacher to the adherents of the Jewish Law, Kant carefully explains that "when appeals are made here to older (Mosaic) legislation and prefiguration, as though these were to serve the Teacher as means of confirmation, they are presented not in support of the truth of his teachings but merely for the introduction among people who clung wholly, and blindly, to the old." (*Rel. 6: 162/150*)

ethical existence (*ethischen gemeinen Wesen*), where the particular wills of individuals following the moral law are united by God into a common goal, which they could not achieve on their own. The key instrumental value of historical faith in the road to the autonomy characteristic of an ethical commonwealth leads inevitably to the recognition that historical faith is “universally binding” (*Rel. 6:181-182/170*). The obligation to practice historical faith, however, is a result of its value as a *means*, not as an end in itself: “the moral believer is ever open to historical faith so far as he finds it furthering the vitality of his pure religious disposition” (*Rel. 6: 182/170*). The preceding analysis leads to Kant’s fundamental principle for evaluating historical, ecclesiastical faith:

such a faith must contain within itself, along with the statutory articles with which it cannot as yet wholly dispense, still another principle, of setting up the religion of good life-conduct as the real end, in order, at some future time, to be able entirely to dispense with the statutory articles. (*Rel. 6: 175/163*).

*b) Corruption of the Historical Churches*

If religious ritual ought to be shed in favour of pure religion, then the disappearance of contingent religious practice seems all but assured. However, like the social dimension of evil, which hindered the conversion of the individual from the evil principle to the good principle, so evil finds its way into the historical churches, corrupting the vehicle necessary for creating the ethical commonwealth.

Religious practice keeps the free will (*Willkür*) from heeding the moral law because the lustre of rites and the certainty of ordinances act as sirens for sensuous human nature—they keep the will focused on the *means* to the end, rather than on the end itself. Religious practice reveals a propensity to religious illusion, not unlike the propensity to evil uncovered in book I: the sensuously constituted human being finds cult and liturgy satisfying enough to

forgo the vocation of rationality that forms the second key element of his nature. The illusion in ritual consists of assuming that God is pleased through the performance of devotional acts than through the performance of rational moral duty:

To deem this statutory faith...as essential to the service of God generally, and to make it the highest condition of the divine approval of man, is *religious illusion* whose consequence is a *pseudo-service*, that is, pretended honouring of God through which we work directly counter to the service demanded by God Himself (*Rel. 6: 168*).

Religious illusion stems from an innocent human tendency to anthropomorphism (*Rel. 6: 191*). It belongs to sensuously constituted beings to grasp the invisible deity through recourse to images related to human life. The danger in anthropomorphism, for Kant, lies in substituting the image for the object itself; the creation of an anthropomorphic image of God impedes practical reason when the image of, say, an all-loving father substitutes for moral exertion. In resting content with the image of a benevolent father, human beings create God in the “form in which we believe we shall be able most easily to win Him over to our advantage and ourselves escape from the wearisome uninterrupted effort of working upon the innermost part of our moral disposition” (*Rel. 6: 168-69*). Religious illusion occurs as the result of the *wilful* adoption of a maxim granting to a means “an intrinsic value instead of the value deriving from the end” (*Rel. 6: 170*).<sup>18</sup>

The evil of religious illusion stems from distorting respect for the moral law; in resting content with religious ritual, a person unwittingly turns the means of religious practice into an end *out of respect for the end*. In other words, the contingent practices and

---

<sup>18</sup> A similar passage describing the danger of religious illusion: when it comes to pilgrimages, or prayers uttered with one’s lips or through the means of a prayer-wheel, “what matters here is not a difference in the external form; everything depends upon the adoption or rejection of the unique principle of becoming well-pleasing to God—upon whether we rely on the moral disposition alone, so far as this disposition exhibits its vitality in actions which are its appearances, or on pious playthings and on inaction” (*Rel. 6: 173/161*).

beliefs of the historical churches are mistaken for the ethical commonwealth itself. The members of the historical churches, in succumbing to radical evil, nonetheless do so out of respect for the moral law—out of an intention to exist in an ethical commonwealth. Just as in book I, however, their acts are unwittingly founded on a sensuous motive, despite the inevitable respect they feel for the moral law in virtue of their predisposition to personality.<sup>19</sup>

The corruption of the historical churches resurrects the problem that the definition of evil as subordination in book I had hoped to leave behind: the moral law is a condition for evil, since the choice in favour of evil is made as a consequence of respect for the law. The logic of real opposition has returned. The problem of evil returns at every step closer to the law because a positive principle implies its (equally positive) negative opposition. Kant's commitment to the logic of real opposition explains the constant return of evil at the tail-end of every 'conversion' to good. The definition of evil as subordination could not—and cannot—elide the structure of real opposition; for good and evil satisfy the conditions for negative magnitude, by which two objects, distinct from each other, nonetheless imply their opposite:<sup>20</sup>

Where, in the rapport of the Law with radical evil, is found the two conditions of real opposition, the two requirements of any *a priori* synthesis: the distance of otherness, of an obstacle—an irreducible conflict between evil and the Law— [is also found] a capturing of the Law within evil, and a return of evil within the Law. If one wants to understand the enigmatic proximity of the Law and of evil, then it is necessary to return to the “onto-logic” of real opposition (Rogozinski 297-98).

Since the social dimension of evil has already precluded “the possibility of triumphing over evil by oneself”, it follows that the problem of evil allows the interpreter to “finally decide

---

<sup>19</sup> “The law, rather, forces itself upon him [the human being] irresistibly by virtue of his moral predisposition” (*Rel.* 6: 36/31).

<sup>20</sup> *NG* 2: 171-172.

between the two opposing conceptions of autonomy—as the absolute self-determination of the Subject or as auto-heteronomy (Rogozinski 291). The need for an ethical commonwealth for the sake of escaping an ethical state of nature—in which the individual prescribes the moral law to himself!—runs against the interpretation of autonomy as self-determination. Neither is the ethical commonwealth an imposition by a foreign will of a social structure on an unwilling subject. Rather, the entrance into a state where the presence of morally obedient beings and the idea of an all-seeing “Moral World-Ruler” is itself a dictate of one’s own reason. The gradual process of reaching the state of pure religion—the ethical commonwealth—through involvement in a historical church reveals that non-moral practices are bound up with respect for the moral law. In other words, the autonomy of the will to follow the good requires the influence of evil (i.e. the contingent practices of historical religion). However, the corruption of the historical churches—religious illusion—spoils the gradual movement from unwitting heteronomy to wilful autonomy running through the *Religion*. At each stage of evil, Kant had found a solution; with the return of evil in the form of religious illusion, it appears that even the process of using non-moral means (church rites and statutes) to reach autonomy falls apart.

### **Solution to the Corruption of the Churches: A Final Attempt**

If the individual requires the practice of historical religion to follow the moral law made visible through the symbol of statute and ritual, then it appears that Kant eventually abandons the claim of autonomy; the prodding of *Willkür* through religious rite seems akin to heteronomy, by which the motive for moral action stems from a source other than respect for the law. Moral training through historical religion, however, should not be mistaken for heteronomy, since the end goal remains the practice of duty for duty’s sake, once

dependence on contingent religious ritual has been overcome. The problem is that, once the lure of rites makes the historical churches an end in themselves, the challenge of heteronomy returns once again. It is at this point that an unexpected solution to heteronomy appears in the text, a solution that Kant was perhaps not consciously aware of: the mere *possibility* of error can incite conscience to scrutinize the will's efforts to follow the demands of the moral law.

a) *The Problem of Uncertainty*

Can a human being, striving to follow the moral law, ever know whether or not he has performed an action for the sake of the law? If he gives money to charity on the conclusion that it is his duty to do so, can he ever be sure that he acted out of duty or out of a contingent reason, such as the natural goodness of his character? Kant notices that the difficulty in distinguishing between acting *from* duty and in conformity with duty becomes acute in cases where the inclinations accord with the demands of duty: “this difference is far more difficult to note in the case where the action accords with duty and the subject has in addition an immediate inclination to do the action” (*G* 4: 397). Kant seems to have provided a pessimistic answer to this problem a few pages later in the *Grounding*:

When we pay attention to our experience of the way human beings act, we meet frequent and—as we ourselves admit—justified complaints that there cannot be cited a single certain example of the disposition to act from pure duty; and we meet complaints that although much may be done that is in accordance with what duty commands, yet there are always doubts as to whether what occurs has really been done from duty and so has moral worth. (*G* 4: 406).

Kant's claim concerning the uncertainty of moral motivation is mirrored by his observation in the *Anthropology* that human beings always dissimulate their behaviour once they become

aware that they are being watched (*A. 7: 121*).<sup>21</sup> And, since the majority of human beings live in community—creating an opening for evil to corrupt the will through the mere *presence* of other people—it should appear equally uncertain whether people can *ever* be confident of performing duty for duty’s sake. It is unsurprising, therefore, that Kant’s use of examples in the *Grounding* to illustrate the performance of duty makes use of an opposition between the demands of the moral law and the desires of the inclinations; for the distinction between the influence of the inclinations and the influence of the moral law becomes less precise in the case of a person disposed to the performance of good acts through a cheerful or charitable disposition.<sup>22</sup>

The recurrence of moral uncertainty has not been lost on certain of Kant’s commentators. In an insightful article on the subject’s phenomenological experience of the moral law, Jean Grondin claims that a person can never know if his choice has been determined by the moral law. Since the moral law “never manifests itself” in human action, “from a phenomenological point of view, the moral law is therefore not at all a law of human action” (Grondin 52).<sup>23</sup> In fact, in contrast to the categorical imperative, the moral law is never given a definitive definition by Kant (Grondin 53). Kant’s main concern is to establish the *causality* of the moral law on the will (Grondin 59), but this causality is not, in fact, proven through the Typic of the *Critique of Practical Reason*, since the image of a “law of nature” does not adequately express the freedom of a will that must be distinct from any deterministic order of nature (Grondin 63-64). As a result, the bridge between the

---

<sup>21</sup> The same dishonesty applies to self-scrutiny, for the act of watching oneself carries with it the desire to distort the image of oneself. (*A. 7: 121*)

<sup>22</sup> For Kant’s use of tension between the demands of the moral law and the desire of the inclinations, see the four examples of acting from duty in the *Grounding* (*G 4: 421-423, 429-430*).

<sup>23</sup> Translations of Grondin are my own.

representation of the moral law and the sensible world cannot be established (Grondin 63-64). Since the causality of the moral law on the will cannot be proven through a rational deduction, it is not surprising to find Kant writing about the importance of the sublime as a symbol for the elevation of the individual to the supersensible level of the moral law (Grondin 55).

For Daniel Dumouchel, Kant's final writings "seem to pay more attention to the feelings and to the subjective aspect of the moral law as a condition for receiving duties demanded by the moral Law" (Dumouchel 110-111). Like Grondin, Dumouchel also notes the importance for ethics of the sublime and of the beautiful, both of which begin to take on greater importance from the writing of the *Critique of Judgment* onwards:

The experience of beauty furnishes a sensible image of moral freedom, which includes the anticipation of a state of harmony between subject and object that is overcome through moral determination; the experience of the sublime permits [*opère*], in a "negative" mode, the presentation of morality as a submission of the sensible inclinations to the demands of moral duty. (Dumouchel 120)

Since the direct determination of the will by the moral law cannot be definitively established, Kant must begin to place more emphasis on subjective experiences of elevation above the sensible as symbols of the law. The final sections of book IV present the possibility of another subjective experience acting as a symbol of the moral law. In similar fashion to the increased usage of beauty and the sublime as symbols for the following of the moral law, the experience of error—typified through the cases of the Inquisitor and of Abraham—act as non-moral conditions for the following of the moral law. The insights of Grondin and Dumouchel, grounded on the difficulties of moral certainty in Kant's ethical writings, point to another way to reach autonomous obedience to the moral law.

b) *Conscience*

The difficulty and complexity of moral decision-making is not lost on Kant, for, as previously noted, his examples of moral decision-making often involve tension between morality and inclination. The same tension is even truer for members of a historical church who would be tempted to make of rituals and statutes the end of moral practice, not its means. It is thus incumbent on a finite rational being to properly distinguish between duty and its opposite in complex cases of moral decision. This requirement leads to the importance of conscience.

The case of conscience is raised by Kant as an outcome of the corruption of the historical churches, since sincerity in following the will of God threatens to jeopardize the will's desire to follow the moral law. Conscience is a "state of consciousness" responsible for probing the willingness of a subject to perform a moral action (*Rel. 6: 185-86/173-74*). Conscience does not evaluate the moral quality of a proposed action, for that capacity belongs solely to the understanding, which judges whether or not a particular act falls under the scope of the universal law (*Rel. 6: 186/174*). Conscience, rather, concerns the *certainty* of the morality of an action: the subject must be certain that he has exhausted all possible means of determining whether or not an act conforms to and is motivated by the moral law (*Rel. 6: 186/174*). Kant concludes his definition of conscience by explaining that "reason here judges itself, as to whether it has really undertaken that appraisal of actions (as to whether they are right or wrong) with all diligence, and it calls the man himself to witness *for* or *against* himself whether this diligent appraisal did or did not take place" (*Rel. 6: 186/174*).

As an example of erring conscience, Kant raises the case of the inquisitor, who unknowingly lapses into moral error. An exemplar of heteronomy, the inquisitor turns the means of historical faith into an end, choosing to listen to the arbitrary—and uncertain—will of God instead of the necessary demands of the moral law, burning heretics as an act of righteousness. The will of God *could* override the certainty dictated by the moral law that it is wrong to kill a human being on the basis of his religious beliefs (*Rel. 6: 186-87/175*). The command of God, however, can never itself be known with certainty, since the will of God has only been made known through contingent historical documents passed on and interpreted by fallible human beings (*Rel. 6: 187/175*). The error of the inquisitor lies not in acting on the will of God, but in assuming that he has understood God's will *with certainty*. All historical faith is prone to the error of the inquisitor: "This is the case with respect to all historical and visionary faith; that is, the *possibility* ever remains that an error may be discovered in it" (*Rel. 6: 187/175*). As long as there is a *possibility* for error, "it is unconscientious to follow such a faith with the possibility that perhaps what it commands or permits may be wrong, i.e., with the danger of disobedience to a human duty which is certain in and of itself" (*Rel. 6: 187/175*). Since, in the case of the inquisitor, knowledge of the will of God cannot be obtained with the same certainty as knowledge of the moral law, he should have listened to the law of reason and desisted from burning heretics for their beliefs.

There is another way of coming to know the will of God, one that does not involve the fallible mediation of historical documents and their interpretation: direct revelation. God, in other words, could choose to make his will known directly to an individual or to a group. Such is the case for Abraham, the paragon of *Schwärmerei* (religious enthusiasm) according to Kant, for Abraham chooses to ignore the certainty of the moral law for the sake of an

imagined divine voice. Kant mentions Abraham in the *Religion* only in passing, as he chooses to focus primarily on the case of the inquisitor. Nonetheless, Abraham stands as the representative of unconscientious obedience to direct revelation, choosing to listen to an uncertain command “to slaughter [*zu schlachten*] his own son like a sheep” (*Rel.* 6: 187/175). The significance of Abraham for an interpretation of conscience, however, can be developed more fully through a brief analysis of a footnote dedicated to this biblical figure in the *Conflict of the Faculties*.

Shortly before the footnote on Abraham in the *Conflict of the Faculties*, Kant raises the same problems of certainty as in the discussion of the inquisitor in the *Religion*. The Bible, claims Kant, would be the “most effective organ for guiding men and citizens to their temporal and eternal well-being” (*C.* 7: 63; 115). The authenticity of the Bible as the “Word of God”, however, cannot be attested, since historical documents do not provide such certainty (*C.* 7: 63). Direct revelation is also ruled out, for it is impossible for the empirically conditioned senses of human beings to grasp the signs of an infinite being (*C.* 7: 63).

The discussion of certainty in matters of divine revelation mirrors almost exactly the same discussion in the *Religion*, except for one novel element—the certainty that a revelation *cannot* come from God. Though we may be uncertain at some times as to whether a voice belongs to God, we may be absolutely certain *at all times* in cases where the command of the divine voice contradicts the demands of the moral law (*C.* 7: 63). It is at this point that Kant pens his footnote on Abraham, to offer the worst example of flagrant contradiction between the moral certainty of the law and the terrifying uncertainty of the divine voice:

We can use, as an example, the myth of the sacrifice that Abraham was going to make by butchering and burning his only son at God's command (the poor child, without knowing it, even brought the wood for the fire). Abraham should have replied to this supposedly divine voice: 'That I ought not to kill my good son is quite certain. But that you, this apparition, are God—of that I am not certain, and never can be, not even if this voice rings down to me from (visible) heaven (C. 7: 63/115, note.)

The sparseness of the footnote—Kant has not even mentioned the journey of Abraham and his son to the mountain, nor the outcome of the story after Abraham's display of obedience—allows Kant to focus on the contradiction between the universal law and the divine voice. Unlike the brief note in parentheses in the *Religion*, Kant confidently affirms that Abraham "*should have replied...*" Abraham's certainty concerning the treatment of his son should have been unwavering (it "is quite certain"); his un-certainty concerning the command of God should have been unfaltering: "but that you, this apparition, are God—of that I am not certain." One certainty, one necessary judgment is clear through the conflict between the law and the voice: Abraham could not be certain that the voice belongs to God, even if "this voice rings down...from the (visible) heaven." The limits to theoretical reason, creating uncertainty with respect to the origin of the voice, engender certainty with respect to the unknown origins of a command that contradicts the demand of the universal law. When confronted with a course of action that contradicts the command of the moral law, the understanding can be sure that the action *is not* moral.

The apophatic character of Kant's ethics has become perceptible again: though, as Grondin and Dumouchel point out, the determination of the will by the law can never be established with certainty, the judgment concerning what *does not* conform to the universal law is able to be made with certainty. The experience of uncertainty, however, should lead the subject to seek the certainty necessary for making a judgment concerning the morality of

his act because, in order to be sure of the immorality of an act, the agent must be *sure* that the law forbids the act. The fact of uncertainty, therefore, spurs the understanding to continue seeking just what the moral law demands. Conscience is integral to the last part of the *Religion*, since its dissatisfaction with uncertainty forces the member of the historical church not to rest content with a possibly immoral means, but to act for the sake of the end. Making the means the end—typified by the inquisitor, Abraham, and any member of a historical church in thrall to evil (i.e. contingent revelation, rituals, and liturgies)—is rectified through the voice of conscience that informs the understanding of the *possibility* of not keeping duty for duty's sake. With the *possibility* of failing to follow the moral law comes the necessity of not resting on a means as if it were an end.

The description of conscience becomes even more dramatic after the publication of the *Religion*. In the *Metaphysics of Morals*, Kant describes conscience as a threatening judge, which elicits awe and fear on the part of the subject standing before the tribunal:

Every man has a conscience and finds himself observed, threatened, and, in general, kept in awe (respect coupled with fear) by an internal judge; and this authority watching over the law in him is not something that he himself (voluntarily) *makes*, but something incorporated in his being. . . . He can indeed stun himself or put himself to sleep by pleasures and distractions, but he cannot help coming to himself or waking up from time to time; and when he does, he hears at once its fearful voice. (*MM*. 6: 438).

As in the *Religion*, Kant does not claim that conscience evaluates the moral quality of acts, since that capacity is fulfilled by the understanding. Unlike the *Religion*, Kant emphasizes the comparison of conscience with the role of a judge. Though a faculty of the subject, conscience must nevertheless be *thought* as distinct from the subject, for the feeling of falling under the judgement of conscience is best expressed through the image of standing

before a judge foreign to oneself (*MM*. 6: 439). The best image for conceiving of conscience as judge is that of God, an “omnipotent moral being”, almighty and capable of playing the role of “inner judge of all free actions” (*MM*. 6: 439). God is only postulated as a helpful analogy for conscience, since the *existence* of God does not require proof nor can it be proven (*MM*. 6: 439). The image of God as *symbol* for conscience is appropriate as a result of divine power; the power and knowledge possessed by God, on the one hand, demonstrates the inescapable scrutiny of conscience and, on the other hand, indicates the posture of accountability and submission that the subject must adopt before his own conscience, as if he were “submitting to the will of this [divine] Being, as the rule of justice” (*MM*. 6: 440). The use of fear and awe to describe the feeling of falling under the scrutiny of conscience become clearer through the analogy with God, since the power of an omnipotent, all-seeing deity elicits admiration of his majesty, yet fear at his unbounded power.

In focusing on later Kantian texts, Thomas Hill has argued that, toward the end of Kant’s career, the role of non-moral incentives becomes consistently greater for moral action. The increasing presence of non-moral incentives in Kant’s later words, however, does not necessarily result in heteronomy. The pain experienced through a violation of the moral law does not motivate the will to act morally in and of itself; the pangs of conscience pinch the subject *because* he has violated a law that he already respects (Hill 65). Furthermore, the subject’s fear at violating the moral law is a worthy kind of fear, since the moral agent fears this violation “through an unwillingness to betray our moral standards” (Hill 67).

Hill’s distinction between morally worthy and unworthy fear in the role of conscience, along with his observation that non-moral incentives begin to make a return

from the periphery in Kant's later works, is certainly right. My analysis already makes reference to a similar observation made by Grondin and Dumouchel. Furthermore, Hill is right in noticing that Kant does not intend to make of fear and non-moral incentives heteronomous springboards for moral action. Hill's analysis of conscience, however, does not highlight the fact that non-moral motives act as boundary-markers between acts motivated by the moral law and acts motivated by sensuous inclination. Conscience, in other words, takes on greater importance in Kant's later work because it permits clearer judgment of the morality of an act through the tension between the moral command of the law and the contingent incentives of a case motivated by inclination. Conscience leads to respect by taking advantage of the tension between the law and the inclinations; as long as the possibility of unwitting motivation by the inclinations exists, the prodding of conscience may serve to incite a person to determine just what the law requires. The logic of real opposition demands the possibility of error as a condition for obedience to the moral law. Hill's analysis, in the end, does not go far enough in pointing out the significance of conscience and non-moral incentives for Kant's ethics, nor the logical reasons for the increasing use of non-moral incentives in the second, 'historical' stage of Kant's philosophy.

### **Consequences of this Analysis of Radical Evil**

The conflict between the divine will and the moral law manifest in the case of Abraham mirrors the conflict between the demands of the moral law and the pull of the inclinations in the examples of moral action in the *Grounding*. Unlike the earlier work, however, by the time that Kant wrote the *Religion*, incentives outside the moral law (such as conflict, contradiction, and error) had become means for grasping the command of the *a priori* law. Since the moral law should not be counted as equivalent to reason, the challenge

of judging with certainty whether or not reason has properly received the law becomes acute for the sake of accurately representing duty to the will. As long as the possibility of error is present, the subject obedient to the moral law must continue his attempt to conform the maxim of his will to the universality of the law. The possibility of error thus leads to a greater effort to provide a solution to the perennial Kantian problem in ethics: finding a synthesis between the will and the moral law.

The interpreter of Kant's ethics, at this point, faces a challenge, for he must ask himself whether or not a Kantian ethics must imply a sharp separation between the moral law and the influence of the sensible inclinations. If the success of Kant's ethics depends on Kant's ability to make such a sharp distinction, then he seems to have failed, for he never provides an example of a finite rational being having fully overcome the influence of radical evil. If, on the other hand, it is seen that Kant remains committed to the logic of real opposition—which implies the co-implication of evil with the good—then it appears that the interpreter will have to ask if this counts as a true form of autonomy. I will not answer this final question, however, since it depends upon more foundational assumptions that extend beyond the scope of this thesis (such as, 'what constitutes morality'? Or 'what is autonomy'?).

The analysis presented in this thesis points to the conclusion that the will, for Kant, requires the influence of factors extraneous to the moral law in order to be able to follow the law—and this requirement is never overcome in the *Religion*. Whether this view counts as a legitimate alternative to empiricist ethics—which Kant himself thought he was providing—is another matter. These deeper questions, however, can only be addressed once the work has been accomplished of revealing the fact that Kantian autonomy (at least in the *Religion*)

cannot be equated with autarky. If this thesis makes a small contribution toward correcting a mistaken interpretation of Kantian autonomy and thereby opening up avenues for further discussion about the nature and viability of Kantian ethics (which is important for interpreting the distinction between Kantian and empiricist ethics), then this thesis will have accomplished its task.

## Conclusion

### The Ideal of Autonomy

In 1784, eight years before he published the essay “On the Radical Evil in Human Nature”, Kant defined Enlightenment as “the human being’s emancipation from its self-incurred immaturity” (*Auf.* 8:35). The short, yet influential essay, defines human immaturity as “the inability to make use of one’s intellect without the direction of another” (*Auf.* 8:35). To be autonomous, in other words, is to be able to make a public use of one’s understanding without the guidance of another person.

Kant, of course, recognizes that society cannot operate without obedience, since citizens must submit to a ruler or other authorities who maintain social order. Hence, Kant makes a distinction between a private and a public use of reason. The private use of reason is necessary for fulfilling a “*civil* post or office with which one is entrusted” (*Auf.* 8:37); this use of reason, in other words, is placed at the service of a *particular* state, fulfilling a *particular* office. The public use of reason, on the other hand, is *universal* in nature: it is typified by the scholar, who subjects his ideas to the universal agreement of a learned community (*Auf.* 8:37). The priest, for example, is bound to teach the particular doctrine in the service of his church. As a scholar, however, he must be allowed complete freedom to test—and even disagree with—the tenets of his church (*Auf.* 8:38). The same distinction found throughout Kant’s critical works between *noumena* and *phenomena*, the *a priori* and the empirical, the universal and the particular, is at work in Kant’s definition of Enlightenment and of autonomy.

Since the understanding is a natural faculty of rational beings, the universal, public use of the understanding cannot be curtailed, unless through external force or by the personal choice of individuals. Rational beings will pursue freedom as long as the authorities in place allow them to do so: “it is much more likely than an entire public should enlighten itself; indeed it is nearly unavoidable if one allows it the freedom to do so” (*Auf.* 8:36). Despite Kant’s optimism that a populace will become free as long as the political conditions are propitious to a public use of reason, there remains a pessimistic outlook in the essay: rational beings themselves choose to follow the guidance of another, rather than their own rational understanding.

“*Sapere aude!* Have the courage to make use of your own intellect” (*Auf.* 8:35). The public use of reason requires courage because it involves making use “of one’s intellect without the direction of another” (*Auf.* 8:35). Hence, alongside Kant’s bold claim that Enlightenment involves freedom from the guidance of another (in the form of political or religious institutions) comes the fear that politically liberated human beings may prefer docility to courageous reasoning. The fault of docility lies less with the guardians who control their subjects and more with the subjects themselves who, in being “content to remain immature for life”, make it “easy for others to set themselves up as their guardians” (*Auf.* 8: 35). Immaturity (or “the inability to make use of one’s intellect without the direction of another”<sup>1</sup>) is the result of “idleness and cowardice” by rational beings who find it comfortable to remain under the authority of guardians in their public use of reason.

---

<sup>1</sup> *Auf.* 8:35.

Kant, therefore, cannot claim that his time was an age of Enlightenment; rather, he lives in an “enlightened” age, in which the “obstacles to universal enlightenment, or to the emancipation from one’s self-incurred immaturity,...[are] gradually becoming fewer” (*Auf.* 8: 40). Though freedom from the guidance of tradition and institutions in the public use of reason was gradually being achieved, Kant readily admits that his era has not yet reached the goal of complete Enlightenment. In claiming that his era is enlightened, he remains confident, however, that political freedom is gradually being opened and that autonomous use of the understanding will naturally follow suit.

### **Limits of Autonomy**

Political freedom is certainly difficult to achieve, especially as rulers of all kinds work hard to maintain authority and control over their subjects. On the other hand, Kant may have discovered around the time of writing the *Religion* that the *desire* to be free may be even harder to produce. The discovery that rational human beings willingly choose to subordinate the moral law to the influence of the inclinations (a tendency that Kant names radical evil) indicates that the political conditions for freedom may not be enough to entice sensuously inclined human beings to listen only to their own reason in fulfilling duty for duty’s sake.

The political and ethical state of nature corresponds closely to the definition of Enlightenment offered by Kant several years earlier: in both forms of the state of nature, the “individual prescribes the law for himself, and there is no external law to which he, along with all others, recognizes himself to be subject” (*Rel.* 6: 95; 87). The realization that the presence of other people invites the will to once again subordinate the moral law to the inclinations necessitates an ethical commonwealth in which people are united by and

protected in their following of the moral law. Autonomy, therefore, may be freedom from the guidance of another in “What is Enlightenment?”; however, by the time of the publication of the *Religion*, autonomy *requires* the presence of others in a special community in order to follow duty for duty’s sake.<sup>2</sup>

Despite the need for a community in order to act autonomously, the formation of the ethical commonwealth is a work in progress that requires the influence of finite historical institutions motivating rational human beings to heed the moral law. Tradition, institutions, and the guidance of another are *required* in order to learn duty for duty’s sake through the symbols of the moral law acted out in the historical churches. Once the members of an historical church are built up in their respect for the moral law, the contingent rites and statutes of the church—which always threaten to substitute for respect for the moral law—may be left behind.

As was the case in “What is Enlightenment”, the state of autonomy is never fully achieved in the *Religion*. The reader is regularly promised that duty for duty’s sake is possible, but, at several successive stages in the *Religion*, a foreign influence outside the moral law is necessary for overcoming evil and for training the will to respect the law alone. Book IV ends with an examination of the figure of the inquisitor, an example that demonstrates what *should not* be done to follow the moral law; the book, however, does not end with an example of someone who actually *does* follow the law (solely out of motivation for the law). The recurrence of elements outside the law—such as the historical churches, error, or conscience—along with the absence of a real example of duty for duty’s sake

---

<sup>2</sup> Though, as I have already indicated, the need for the presence of an ethical commonwealth does not release the individual from the responsibility of following the moral law simply for the sake of the law.

suggests that radical evil is not fully solved, since individuals on their own cannot be trusted to act out of respect for the law. Furthermore, if the analysis of the structure of real opposition is correct, then it follows that the moral law never determines the will without the absence of an influence external to the law. A pure *a priori* law, in other words, does not act as the sole source of motivation for sensibly constituted human beings.

### **Kant's Ethics as a Solution to Empiricism?**

Kant critiques empiricism on the grounds that it does not offer universal theoretical concepts, nor does it permit a universal determination of the will.<sup>3</sup> Hume's critique of causality—which Kant singles out in several texts—is more than a theoretical interest in the domain of epistemology. The fact that the concept of a cause cannot be validly established as a fact about the world means that universal concepts cannot be validly established in metaphysics or ethics. It is a truism for Hume that all knowledge begins and must be resolved into impressions of the senses. Ideas, for instance, are the result of a combination of various impressions through the relation of contiguity, resemblance, and cause and effect.

A universal concept in ethics, however, is not entirely possible on this model. The concept of virtue and vice emerges from feelings of pleasure and pain. Though human beings feel pleasure and pain at similar objects, these common feelings result from a shared human nature that is discovered through a process of induction; human nature, in other words, is not strictly universal, since the discovery of different behaviour in a select group of people would lead to a revision of our concept of human nature. Artificial virtues demonstrate even more so the contingent nature of empiricist ethics. Since human beings do not naturally feel motivation to follow the rules of justice, it is necessary for society to

---

<sup>3</sup> See *KrV* 2:496. See also my analysis of this passage in the introduction (pg. 2).

demonstrate the utility of acting justly by reorienting self-interest through social formation. The feelings of pleasure and pain moving the will to act are influenced by the opinions of parents, politicians, and neighbours—all of whom regard certain acts and qualities of character as useful or otherwise. The perception of utility, therefore, is the result of the contingent influence of society training its members to regard just acts and characters as virtuous. In the case of Hume, the influence of institutions in forming moral character does not appear optional or able to be left behind.

Kant objects to any empirical basis for ethics (whether it be happiness or a moral feeling) because empiricism lacks the universality required to form moral judgment:

Empirical principles are wholly unsuited to serve as the foundation for moral laws. For the universality with which such laws ought to hold for all rational beings without exception (the unconditioned practical necessity imposed by moral laws upon such beings) is lost if the basis of these laws is taken from the particular constitution of human nature or from the accidental circumstances in which such nature is placed. (*G.* 4: 442)

The moral law is the universal *a priori* principle that empiricism—due to its reliance on contingent empirical impressions—cannot establish. The contingent foundations of empiricism ensure that the problem of the sensible knave will repeatedly occur because the feeling of sympathy—intended to erect moral universality on the resemblance of ideas—dissipates as the distance between two ideas becomes larger. Kant’s solution to the contingency of empiricism is achieved by “removing the empiricism” (*KpV* 5: 53). The moral law thus universally determines the will in virtue of its *a priori* nature and respect supplies the subjective motivation necessary to encourage a person to fulfill his duty for its own sake.

As I have argued in this thesis, however, the problem of radical evil represents an important challenge to Kant's theory of respect and to the ideal of autonomy founded on the principle of duty for duty's sake. Since history demonstrates that people often subordinate the moral law to the influence of the inclinations, it follows that sensuous human beings do not readily feel sufficiently motivated by respect for the moral law. The ultimate solution to the problem of evil—in the form of the ethical commonwealth—does not appear to be fully achieved, since the historical churches needed for training corrupt human beings into following the moral law are never left behind. If this analysis is correct, then Kant's critique of and response to empiricism appear impractical and unable to provide the solution to contingency that Kant had hoped to achieve.

After this analysis of the impasse represented by radical evil, it can be asked whether or not Kant's writings on morality and autonomy should provide a framework for contemporary ethical theories that seek to avoid the assumptions of empiricism? The question is too large to be discussed here. However, several questions for further research could be proposed. First, to what extent is autonomy really possible, if autonomy is defined as freedom from the guidance of another? Second, since the individual must resort to historical institutions for the sake of moral education and does not seem able to leave that education behind, then to what extent can Kant's transcendental philosophy be considered a successful answer to empiricism? Lastly, if Kant perceptively noted a real problem in empiricism, and if his solution to empirical contingency turns out to be impractical, then perhaps it is worth exploring the key assumptions of empiricism regarding nature and knowledge in order to find a different solution to the issue of contingency and the sensible knave that recurs in empiricist ethics. If these questions are indeed legitimate and open

fruitful avenues for further research in ethics and the history of philosophy, then this thesis will have accomplished its primary task.

## Bibliography

### Primary Sources

- Hume, David. *An Enquiry Concerning the Principles of Morals*. in *Hume's Ethical Writings*. Ed. Alasdair MacIntyre. New York: Collier Books, 1965. 21-156.
- Hume, David. *A Treatise of Human Nature*. New York: Dover Philosophical Classics. 2003.
- Kant, Immanuel. "An Answer to the Question: What is Enlightenment?" in "Toward Perpetual Peace" and *Other Writings on Politics, Peace, and History*. Ed. Pauline Kleingeld. New Haven: Yale University Press, 2006. 17-23.
- Kant, Immanuel. *Anthropology from a Pragmatic Point of View*. Trans. Robert B. Louden. Cambridge: Cambridge University Press, 2006.
- Kant, Immanuel. "Attempt to Introduce the Concept of Negative Magnitudes into Philosophy" in *The Cambridge Edition of the Works of Immanuel Kant: Theoretical Philosophy, 1755-1770*. Trans. David Walford. Cambridge: Cambridge University Press, 1992. 203-241.
- Kant, Immanuel. *The Conflict of the Faculties*. Trans. Mary Gregor. New York: Albaris Books, Inc. 1979.
- Kant, Immanuel. *Critique of Practical Reason*. Trans. Werner S. Pluhar. Indianapolis: Hackett Publishing Company, 2002.
- Kant, Immanuel. *Critique of Pure Reason*. Trans. Werner S. Pluhar. Indianapolis: Hackett Publishing Company. 1996.
- Kant, Immanuel. *Grounding for the Metaphysics of Morals*. Trans. James W. Ellington. Indianapolis: Hackett Publishing Company, 1993.
- Kant, Immanuel. *The Metaphysics of Morals*. Trans. Mary Gregor. Cambridge: Cambridge University Press, 1991.
- Kant, Immanuel. *Religion Within the Limits of Reason Alone*. Trans. Theodore M. Greene and Hoyt H. Hudson. New York: Harper & Row Publishers, 1960.

### German editions of Kant's works

- Kant, Immanuel. *Der Steit der Fakultäten*. Cambridge: Cambridge University Press, 1991.

- Kant, Immanuel. *Die Religion innerhalb der Grenzen der blossen Vernunft* in *Immanuel Kant: Werke in Zehn Bänden*. Vol. 7. Ed. Wilhelm Weischedel. Darmstadt: Wissenschaftliche Buchgesellschaft, 1975. 649-879.
- Kant, Immanuel. *Grundlegung zur Metaphysik der Sitten*. Ed. Karl Vorländer. Leipzig: Felix Meiner, 1925.
- Kant, Immanuel. *Kritik der Praktischen Vernunft*. Ed. Karl Voländer. Hamburg: Felix Meiner, 1993.
- Kant, Immanuel. “Versuch den Begriff der Negativen Grössen in die Weltweisheit einzuführen” in *Immanuel Kant: Werke in Zehn Bänden*. Vol. 2. Ed. Wilhelm Weischedel. Darmstadt: Wissenschaftliche Buchgesellschaft, 1983. 779-819.

### **Secondary Sources on Kant**

- Allison Henry E. *Kant's theory of freedom*. Cambridge: Cambridge University Press, 1990.
- Anderson-Gold, Sharon. “God and Community: An Inquiry into the Religious Implications of the Highest Good” in *Kant's Philosophy of Religion Reconsidered*. Eds. Philip J. Rossi and Michael Wreen. Indianapolis: Indiana University Press, 1991. 113-131.
- Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: University of Chicago Press, 1960.
- Beck, Lewis White. “A Prussian Hume and a Scottish Kant” in *Essays on Kant and Hume*. New Haven: Yale University Press, 1978.
- Bruch, Jean-Louis. *La Philosophie Religieuse de Kant*. Paris : Aubier, 1968.
- Carnois, Bernard. *La cohérence de la doctrine kantienne de la liberté*. Paris : Édition du Seuil, 1973.
- Despland, Michel. *Kant on History and Religion*. Montreal: McGill-Queen's University Press, 1973.
- Dumouchel, Daniel. “Kant et la ‘part subjective’ de la moralité” in *Kant actuel. Hommage à Pierre Laberge*. Eds. François Duchesneau, Guy Lafrance and Claude Piché. Montréal: Bellarmin, 2000. 109-125.
- Foessel, Michaël. *Le scandale de la raison: Kant et le problème du mal*. Paris : Honoré Champion, 2010.

- Grondin, Jean. "La phénoménologie de la loi morale » in *Kant actuel. Hommage à Pierre Laberge*. Eds. François Duchesneau, Guy Lafrance and Claude Piché. Montréal: Bellarmin, 2000. 51-65.
- Guyer, Paul. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000.
- Heidegger, Martin. *Kant and the Problem of Metaphysics*. Bloomington: Indiana University Press, 1997
- Heidegger, Martin. *Kant und das Problem der Metaphysik*. Frankfurt am Main: Vittorio Klostermann, 1951.
- Hill, Thomas E. "Punishment, Conscience, and Moral Worth" in *The Southern Journal of Philosophy*. Vol. 36, Supplement (1997): 51-71.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.
- Krüger, Gerhard. *Philosophie und Moral in der Kantischen Kritik, zweite Auflage*. Tübingen: J.C.B. Mohr (Paul Siebeck), 1967.
- Lebrun, Gérard. *Kant et la fin de la métaphysique. Essai sur la « Critique de la faculté de juger »*. Paris : Armand Colin, 1970.
- Lebrun, Gérard. *La patience du concept. Essai sur le discours hégélien*. Paris : Gallimard, 1972.
- Louden, Robert. *Kant's Impure Ethics: From Rational Beings to Human Beings*. Oxford: Oxford University Press, 2000.
- Michalson, Gordon E. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. Cambridge: Cambridge University Press, 1990.
- Michalson, Gordon E. "The Problem of Salvation in Kant's *Religion Within the Limits of Reason Alone*" in *International Philosophical Quarterly*. Vol. 37, n. 3 (Sep. 1997): 319-328.
- Mulholland, Leslie A. "Freedom and Providence in Kant's Account of Religion: The Problem of Expiation" in *Kant's Philosophy of Religion Reconsidered*. Eds. Philip J. Rossi and Michael Wreen. Indianapolis: Indiana University Press, 1991. 77-102.
- Mulholland, Leslie A. *Kant's System of Rights*. New York: Columbia University Press, 1990.

- Reath, Andrews. "Kant's Theory of Moral Sensibility: Respect for the Law and the Influence of Inclination" in *Kant Studien*. Vol. 80, no. 3 (January 1989). 284-302.
- Reboul, Olivier. *Kant et le problème du mal*. Montréal: Les presses de l'université de Montréal, 1971.
- Ricoeur, Paul. *Le conflit des interprétations : essais d'herméneutique*. Paris : Seuil, 1969.
- Rogozinski, Jacob. *Le don de la Loi : Kant et l'énigme de l'éthique*. Paris : Presses Universitaires de France, 1999.
- Stratton-Lake, Philip. *Kant, Duty and Moral Worth*. New York: Routledge, 2000.
- Walker, Ralph. "What he really means" in *The Times Literary Supplement*. Issue 4892 (January 3, 1997): 10-11.
- Wood, Allen W. *Kant's Moral Religion*. Ithaca: Cornell University Press, 1970.
- Yovel, Yirmiyahu. *Kant and the Philosophy of History*. Princeton: Princeton University Press, 1980.
- Zinken, Melissa. "Respect for the Law and the Use of Dynamical Terms in Kant's Theory of Moral Motivation" in *Archive für Geschichte der Philosophie*. Vol. 88 (2006): 31-53.

### **Secondary Sources on Hume**

- Aiken, Henry David. « The Originality of Hume's Theory of Obligation » in *Philosophy and Phenomenological Research*. Vol. 42, No. 3 (March 1982): 374-383.
- Árdal, Páll S. *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press, 1966.
- Baier, Annette. *The Cautious Jealous Virtue: Hume on Justice*. Cambridge, Massachusetts: Harvard University Press, 2010.
- Brown, Charlotte. "From Spectator to Agent: Hume's Theory of Obligation" in *Hume Studies*. Vol. 20, n. 1 (Apr. 1994). 19-34.
- Brown, Charlotte. "Is Hume an Internalist?" in *Journal of the History of Philosophy*. Vol. 26, n. 1 (Jan. 1988). 69-87.
- Cohon, Rachel. *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press, 2008.
- Coleman, Dorothy. "Hume's Internalism" in *Hume Studies*. Vol. 18, n. 2 (Nov. 1992). 331-347.

Darwall, Stephen. "Motive and Obligation in Hume's Ethics" in *Noûs*. Vol. 27, n.4 (Dec. 1993): 415-448.

Gauthier, David. "Artificial Virtues and the Sensible Knave" in *Hume Studies*. Vol. 18, n. 2 (Nov. 1992). 401-427.

Haakonssen, Knud. "Hume's Obligations" in *Hume Studies*, vol. 4, n. 1 (April 1978): 7-17.

Kalt, Stefan. "David Hume on the Motivating Effect of Moral Perception: Internalist or Externalist?" in *History of Philosophy Quarterly*. Vol. 22, n. 2 (Apr. 2005): 143-160.

Krause, Sharon. "Hume and the (False) Luster of Justice" in *Political Theory*. Vol. 32, n. 5 (Oct. 2004): 628-655.

Le Jallé, Éléonore. *L'autorégulation chez Hume*. Paris : Presse universitaires de France, 2005.

Mackie, J.L. *Hume's Moral Theory*. London: Routledge & Kegan Paul, 1980.

Pitson, Antony E. "Hume on Promises and their Obligation" in *Hume Studies*. Vol. 14, n. 1 (April 1988): 176-190.

Radcliffe, Elizabeth S. "How does the Humean Sense of Duty Motivate?" in *Journal of the History of Philosophy*. Vol. 43, n. 3 (July 1996): 383-407.

Radcliffe, Elizabeth S. "Hume on Motivating Sentiments, the General Point of View, and the Inculcation of Morality" in *Hume Studies*. Vol. 20, n. 1 (April 1994): 37-58.

Tate, Margaret Watkins. "Obligation, Justice, and the Will in Hume's Moral Philosophy" in *Hume Studies*. Vol. 31, n. 1 (April 2005): 93-122.

### **General Works**

Agamben, Giorgio. *Le sacrement du langage: archéologie du serment*. Trans. Joël Gayraud. Paris : Vrin, 2009.

Darwall, Stephen. *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge: Cambridge University Press, 1995.

Guérout, Martial. *Leibniz: Dynamique et Métaphysique*. Paris : Aubier-Montaigne, 1967.

Korsgaard, Christine. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge, Cambridge University Press, 1996.

Leibniz, G.W. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*. Trans. E.M. Huggard. Chicago: Open Court, 1985.

Nussbaum, Martha. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, Massachusetts: Harvard University Press, 2006.

Rawls, John. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press, 1971.

Rawls, John. *Lectures on the History of Moral Philosophy*. Cambridge, Massachusetts: Harvard University Press, 2000.