

4-2005

## **Purposive Action: The Centrality of Teleology in Kant's Formula of Universal Law of Nature**

Franklin D. Worrell

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



Part of the [Philosophy Commons](#), and the [Religion Commons](#)

---

Purposive Action: The Centrality of Teleology in Kant's Formula of Universal Law of  
Nature

by

Franklin D. Worrell

Undergraduate honors thesis under the direction of

Professor Husain Sarkar

Department of Philosophy and Religious Studies

Submitted to the LSU Honors College in partial fulfillment of  
the Upper Division Honors Program.

April, 2005

Louisiana State University  
& Agricultural and Mechanical College  
Baton Rouge, Louisiana

For

My Parents

Who instilled in me a sense of duty

“Sometimes I lie awake at night and I ask ‘What is the meaning of life?’ Then a voice comes to me that says ‘I before E except after C!’”

- Charlie Brown

## Table of Contents

<b>Abbreviations</b>	<b>iii</b>
<b>Introduction</b>	<b>1</b>
<b>1 An Exegesis of the Preface and Section I of the <i>Groundwork</i></b>	<b>5</b>
I. The Scheme of Things	5
II. The Good Will and the Three Propositions	9
<b>2 The Moral Law</b>	<b>16</b>
I. The Three Types of Imperatives	17
II. The Formulae of Universal Law (FUL) and Universal Law of Nature (FLN)	21
III. The Nature and Structure of Maxims	24
IV. The Contradiction in Conception and Contradiction in Will	29
<b>3 Kant's Explication of the Four Examples</b>	<b>30</b>
I. General Structure of the Four Examples of FLN	30
II. Example One: Suicide	33
III. Example Two: The Lying Promise	35
IV. Example Three: Idleness and Wasted Talents	38
V. Example Four: Beneficence	40
<b>4 Theories of Contradiction</b>	<b>44</b>
I. The Logical Contradiction Interpretation	45
II. The Practical Contradiction Interpretation	52
III. A Critical Note on the Logical and Practical Contradiction Interpretations	59
IV. The Teleological Contradiction Interpretation	62
V. An Objection of Herman's and the Seeds of a New Contradiction Interpretation	72
<b>5 Towards a New Contradiction Interpretation</b>	<b>81</b>
I. Teleology and the <i>Critique of the Power of Judgment</i>	82
II. A New Teleological Contradiction Interpretation	87
III. The Four Examples Revisited	89
IV. The Differences in the Analyzed Contradiction Interpretations	94
V. Some "Casuistical Questions"	99
<b>6 The Supposed Right To Lie</b>	<b>104</b>
I. The Philanthropic Lie and Universal Law	105
II. The Formulas of Humanity as an End in Itself and the Kingdom of Ends	107
III. The Ideal and Non-Ideal Theory	112
IV. The Incorporation of Non-Ideal Elements into Kant's Ethics	114
V. Some Problems and Concerns	115
<b>7 Conclusion</b>	<b>120</b>



<b>Works Cited</b>	<b>122</b>
<b>Curriculum Vitae</b>	<b>124</b>

### Abbreviations

Citations of Kant's works are referenced parenthetically in the text by the abbreviations given below and by volume and page number (volume:page number) of *Kants gesammelte Schriften* published in Berlin by the *Preussische Akademie der Wissenschaften*. This is the standard method of citation for the relevant works of Kant. The exception to this, the *Critique of Pure Reason*, will here be cited in the now traditional fashion of giving the page numbers to both the first and second editions. Quotations will be taken from the translations cited below.

- G** Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 2002.
- MS** Kant, Immanuel. *The Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.
- KpV** Kant, Immanuel. *Critique of Practical Reason*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1997.
- KrV** Kant, Immanuel. *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1998.
- KU** Kant, Immanuel. *Critique of the Power of Judgment*. Edited by Paul Guyer and translated by Eric Matthews and Paul Guyer. Cambridge: Cambridge University Press, 2000.
- SRL** Kant, Immanuel. "On a Supposed Right to Lie from Philanthropy." *Practical Philosophy*. Translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 611-5.

The Categorical Imperative and Its Formulae:

- CI** The Categorical Imperative, as opposed to any other categorical imperatives. The supreme principle of morality as embodied, in part, by the following five formulae.
- FUL** The Formula of Universal Law—“*Act only in accordance with that maxim through which you can at the same time will that it become a universal law*” (G 4:421).
- FLN** The Formula of Universal Law of Nature—“*Act as if the maxim of your action were to become by your will a **universal law of nature***” (G 4:421).
- FH** The Formula of Humanity as an End in Itself—“*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*” (G 4:429).
- FA** The Formula of Autonomy—“*Act only so that the will could regard itself as at the same time giving universal law through its maxim*” (G 4:434).
- FKE** The Formula of the Kingdom of Ends—“Every rational being must at as if he were by his maxims at all times a lawgiving member of the universal kingdom of ends” (G 4:438).

## Introduction

It is unfortunate that the great philosophers cannot return a century, or later, after their deaths to write commentaries on their great works, clarifying any unclear concepts or developing any nascent themes in their writings. Such commentaries could clear up inconsistencies and do away with much controversy over divergent interpretations of their work. Kant is one such philosopher: we would have profited tremendously from his returning exegesis. As it stands, however, certain aspects of Kantian moral philosophy are unclear, with Kant's language suggesting and supporting very different interpretations on several points.

One such point in Kantian ethical thought is the idea of the contradiction in conception and the contradiction in will. In the *Groundwork of the Metaphysics of Morals*, Kant's first major work in moral philosophy, the Categorical Imperative, the supreme principle of morality, is formulated first as:

*Act only in accordance with that maxim through which you can at the same time will that it become a universal law.* (G 4:421)

He continues and very quickly delineates a second version of the same law; the formulation that he feels may be applied to human action for moral evaluation:

*Act as if the maxim of your action were to become by your will a **universal law of nature**.* (G 4:421)

These formulations of the universal law require an agent to determine the maxim, the subjective principle of action, upon which he wishes to act and envision a system of nature in which that maxim was universalized into a universal law of nature.

Maxims which can be conceived of and willed as universal laws are morally permissible. Those that cannot be thought of as a universal law of nature without contradiction violate the

contradiction in conception test, while those that can be conceived of as a universal law of nature, but cannot be rationally willed violate the contradiction in will test. Kant, however, is not clear on what constitutes either a contradiction in conception or a contradiction in will, and three major interpretations—the logical, the practical, and the teleological contradiction interpretation—have emerged to explain these terms.

The logical contradiction interpretation, most notably defended by Allen Wood, suggests that the type of contradiction Kant appeals to is one that follows the guidelines of contradiction in classical logic. Christine Korsgaard's practical contradiction interpretation draws heavily on the efficacy of the agent's action in the world of the universalized maxim. H. J. Paton, who proposes his own version of the teleological contradiction interpretation, maintains that the contradiction arises within a system of human purposes that the agent creates.

The goal of this work is to analyze each of these three theories, evaluate their performances at illuminating the four examples of the Categorical Imperative's first two formulae in the *Groundwork*, and then, from these criticisms derive a new contradiction interpretation, based on a revision of Paton's view, that may explain the contradiction in conception and contradiction in will distinction and which learns from the shortcomings of its three predecessors while preserving some of their successes. Several tasks must be accomplished before the contradiction interpretations can be explained.

First, an exegesis of the opening sections of the *Groundwork of the Metaphysics of Morals* will be necessary to introduce key concepts and Kant's moral theory generally. Chapter One is an exegesis of the Preface and Section I of the *Groundwork*. The Preface lays out the general scheme of Kant's philosophy to show how the moral, practical part fits into the larger whole. Section I of the *Groundwork* introduces and explicates the idea of the good will,

inclination, and duty. It is also in Section I that Kant presents the three famous propositions that describe the good will and lead to a definition of acting from duty:

*Duty is the necessity of an action from respect for law. (G 4:400)*

Chapter Two deals with Section II of the *Groundwork* from the beginning to just after the four examples of the first formulation of the Categorical Imperative. Section II includes the derivation, among other things, of the five formulae of the moral law and four examples that illustrate the formulae. Kant lays out three types of imperatives, hypothetical, assertoric, and categorical, and each of these is discussed, as well as the differences between the first two formulae of the Categorical Imperative. Since the concept of a maxim lies at the very core of the first two formulae of the Categorical Imperative, this chapter will also include a detailed discussion on what constitutes a maxim and a properly formulated maxim. The contradiction in conception and contradiction in will tests are also briefly addressed in Chapter Two.

Chapter Three will consist solely of an exegetical discussion of Kant's four examples of the formula of universal law of nature from the *Groundwork*. The general structure of the examples will be explained, and each illustration will be examined at length as Kant presents them. Thus, all four examples—suicide, the lying promise, the rusting talents, and beneficence—will be examined in turn.

Chapter Four then examines the logical, practical, and teleological contradiction interpretations and evaluates their relative success. How each theory handles each of Kant's four examples will be examined, as well as some of the strengths and weaknesses of each. Adapting a criticism of Barbara Herman, I will show a fundamental flaw in Paton's teleological contradiction interpretation.

In Chapter Five, I will delineate a new contradiction interpretation, springing from the criticism of Herman's that I levy against Paton. Since my reworking of Paton's view will draw heavily on the Kantian notion of teleology, a general discussion thereof will open the chapter. A natural teleology is the idea that everything in nature is purposive and made to fit that purpose, including the reason of human beings. Kant expounded his mature views on teleology in his third great Critique, the *Critique of the Power of Judgment*, and a discussion of its parts dealing with teleology is central to an understanding of Kantian teleology.

The explanation of Kant's teleology is directly followed by a review of the four examples of the formula of universal law of nature under the guidance of this new contradiction interpretation. The contradiction interpretation is expounded in detail, and then various things entailed by each contradiction interpretation are discussed, including what type of maxim is necessitated by each interpretation and what each interpretation dictates to be what a rational agent wills. The work concludes with the new contradiction interpretation being used to solve certain moral puzzles of great weight put forth by Kant in *The Metaphysics of Morals*.

I will conclude in Chapter Six with a detailed discussion of the problem of the philanthropic lie. This discussion will center around Christine Korsgaard's detailed analysis of the problem in her *Creating the Kingdom of Ends*. I discuss her views at length and then evaluate them using the new contradiction interpretation that I have delineated in Chapter Five.

## **An Exegesis of the Preface and Section I of the *Groundwork***

### **I. The Scheme of Things**

The Preface serves as a preliminary to the main text of the *Groundwork of the Metaphysics of Morals*, and, in some respects, it forms a justification for and a representation of the reasoning that leads Kant to his ethical method. Kant opens the Preface to the *Groundwork* by delineating the three divisions of classical philosophy. The Greeks, he says, divided philosophy into three fields or disciplines: physics, ethics, and logic. “This division,” he claims, “is perfectly suitable to the nature of the subject and there is no need to improve upon it except, perhaps, to add its principle, partly so as to insure its completeness and partly so as to be able to determine correctly the necessary subdivisions” (G 4:387). With his pending agenda in mind, he seeks this principle and determines the necessary subdivisions, which will determine the structure of his own ethical theory.

Rational cognition is divided into formal and material cognition, the former concerning itself with the “form of the understanding and of reason itself and with the universal rules of thinking in general” (G 4:387) while the latter takes up the subject of physical, material objects. The science of formal cognition is the study of logic, while material cognition divides itself between two disciplines: physics and ethics. Physics is the delineation of the laws of nature; ethics is the delineation of the laws of freedom. The discipline of formal cognition, the science of logic, must be purely a priori in its basis, it must lack all empirical elements and any empirical basis for its axioms. Ethics and physics, the divisions of material cognition, however, concern themselves with the way the world should be and the way the world is, respectively. Thus,



certain branches of these fields will contain room for empirical observations. One may generalize through induction to produce the laws of physics, just as one may study certain empirical elements of a morally sticky situation to increase understanding of how to bring a moral action into being after a moral motivation is reached.

There is, despite this concern with the empirical, a side of these material philosophies that draws upon purely a priori principles. This is the field of metaphysics. Metaphysics concerns itself, under Kantian philosophy, with “determinate objects of the understanding” (G 4:388). Since metaphysics is not a variant merely of physics, but of material cognition as a whole, Kant has opened up the way for a metaphysics of morals, a purely a priori delineation of the edicts of morality without any external, empirical influence. The rational part of ethics, the metaphysics of morals, Kant says “might properly be called *morals*” (G 4:388), while practical anthropology constitutes the empirical teachings of ethics. Through an analogy with division of labor and specialization in the workplace, Kant emphasizes the importance of keeping the a priori, rational part of ethics separate and very clearly distinct from the empirical side of ethics, practical anthropology. The importance of this division stems from the desire to “know how much pure reason can accomplish in both cases [metaphysics of nature and metaphysics of morals] and from what sources it draws this a priori teaching of its own” (G 4:388-9).

Kant then proceeds to raise a question, more a statement of intent, regarding “the utmost necessity to work out for once a pure moral philosophy, completely cleansed of everything that may be only empirical and that belongs to anthropology” (G 4:389). He claims that if the concepts of duty and moral laws are examined, then *prima facie* such a moral philosophy must exist. If a law holds for humans, then it must hold for all rational beings. Its moral validity is not contingent upon the nature of humanity but instead arises from the principles and nature of

rationality itself and the a priori concepts of reason alone can make it self-evident that any morally binding law finds its binding source of obligation in the constructions of pure reason only. Empirical grounds for morality are reduced to mere motives or practical rules, and may never be seen as morally binding laws of ethics.

This is the fulcrum, upon which Kant will poise his lever to raise up a moral philosophy devoid, he claims, of all empirical concepts and thus binding to all rational beings, not just humans, but all capable of understanding the concept and “common idea of duty and of moral laws” (G 4:389). It is from this idea that Kant will begin his preparatory work for and derivations of the moral law in Sections I and II.

The “Preface” continues with the assertion that the basis of all moral philosophy is its a priori part and that practical anthropology is merely the application of this pure part to a rational being’s life and behavior. Experience is still a necessary factor, because it allows rational beings “partly to distinguish in what cases they are applicable and partly to provide them with access to the will of the human being and efficacy for his fulfillment of them” (G 4:389). Empirical considerations help show humans the most efficient way to bring about a moral action. For example in saving a drowning child (assuming the motivation is moral) my a posteriori knowledge of the world would tell me that it would be more effective to take my nearby boat than to swim the lengthy distance to the child.

Moral principles, those upon which we act, must have their roots in some larger, more definitive guiding principle, or else the corruption of motivation that plagues the human will may make any “moral” action subject to doubt and dismissal. Furthermore, an action must not be merely in accord with the moral law to be deemed ethically laudable, it must be motivated purely for the sake of the law and the sake of upholding it to be genuinely and philosophically moral

and commendable. This passage reflects the justification for the writing of a metaphysic of morals, which can indicate and describe, as derived from a guiding moral principle, the morally correct action carried out with a moral motivation. The passage indicates that the metaphysic of morals may dictate an action, as derived from a higher axiom, as the correct one, thus allowing the rational subject, through a practical anthropology, to concentrate his efforts on determining the motive of the action and ensure that errors in its commission do not undermine its moral status.

Kant follows with an explanation of the failures of other theories of ethics, and attributes their failures to the mixing of “these pure principles with empirical ones” (G 4:390). He claims that such theories have their roots in common rational cognition as opposed to genuine philosophy: a harsh criticism. He says that, “the metaphysics of morals has to examine the idea and the principle of a possible *pure* will and not the actions and conditions of human volition generally, which for the most part are drawn from psychology” (G 4:390-1).

Moving from criticism to guarantee, Kant promises to write a metaphysics of morals in the future, but for now is content laying out the basis for that longer work. This groundwork may also serve as a critique of pure practical reason and it is separate from the promised metaphysics of morals, because he feels that the teachings of such a metaphysics could be read and understood by the multitude, and this popular appeal maybe undermined by the more complex and difficult critique of practical reason required to reach the axiom from which the metaphysics maybe derived. He reveals that this groundwork will search for and discover “the supreme principle of morality” (G 4:392), a search worthy of its own book.

The Preface concludes with a layout for the remaining sections of the *Groundwork*. Its significance lies in that it shows the stages of thought that lead Kant to construct his moral

philosophy in terms of a metaphysics of morals instead of a practical anthropology as well as alluding, early on, to some of the much larger and very important themes of the *Groundwork* and Kant's moral philosophy as a whole. It is the primer to the derivation of the moral law, the information necessary to understand Kant's approach.

## II. The Good Will and the Three Propositions

With the preliminaries established, Kant begins the project of the *Groundwork* in earnest, "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a **good will**" (G 4:393). While opening this section, Kant challenges the reader to imagine anything that can match the unequivocal goodness of a truly and purely good will. He explores two possible sets of candidates, which he divides into *gifts of nature* and *gifts of fortune*. Gifts of nature are comprised of those individual talents most praised by man, such as courage and perseverance, whereas gifts of fortune are those aspects of one's situation that produce happiness and, for the most part, stability, such as financial resources and health. All of these gifts are rejected as meaningless and devoid of all laudability unless a good will lies behind them and motivates their moral use. One can easily surmise situations in which a villain utilizes his courage and financial resources to immoral ends. Kant claims that a good will "brings [a principle of action] into conformity with universal ends" (G 4:393), and thus makes the use of these blessings moral and praiseworthy.

He continues to say that "an impartial rational spectator can take no delight in seeing the uninterrupted prosperity of a being graced with no feature of a pure and good will, so that a good will seems to constitute the indispensable condition even of worthiness to be happy" (G 4:393). It is this good will which proves to be the groundwork upon which the goodness of such gifts of

nature and of fortune are founded. Kant then goes on to reject qualities conducive to a good will, such as self-control, as not good in themselves, but merely contingently good with their commendation originating their alignment with a prerequisite good will. A villain is then described who has such traits, but uses them for evil, thus proving the limited integrity of such traits.

Continuing with rejections of what may otherwise be considered morally praiseworthy, Kant dismisses consequentialism as a legitimate ethical theory. "Usefulness or fruitlessness can neither add anything to this worth [of a good will] nor take anything away from it" (G 4:394). Such a will is not made moral by what it effects; it is made moral by its will to goodness. One may, acting with this good will as motive, fight all natural inclinations within his character and thus produce little good in the world, but the action would still be considered praiseworthy in that the good will brought it about.

Feeling that he has, for now, satisfactorily established the primacy of the goodness of a good will, Kant seeks to justify his qualification of such a will, and defend its position with an examination of some of our inherent characteristics. He uses the existence of our reason to justify his statements regarding the primacy of morality over happiness. Given the assumption of a teleological principle, he asks whether a being whose purpose is preservation, welfare, and happiness would be better equipped with reason and a will or a will devoid of the influence of reason, acting only on instinct. The obvious answer to Kant is the latter. Only a will would be necessary to see to comfort and preservation, because these can be driven by instinct and do not require reason to mediate with the will for their accomplishment. As Kant puts it, "Nature would have taken upon itself the choice not only of ends but also of means and, with wise foresight, would have entrusted them both simply to instinct" (G 4:395). Reason quite often leads us away

from mere happiness and comfort and drives us to the moral act and the manifestations of a good will which frequently undermine our happiness.

This happens on so many occasions that some people allow themselves to slip into a hatred of reason and an envy of those whose cognitive powers are less acute. Reason can have dominion over our will, and “the true vocation of reason must be to produce a will that is good, not perhaps *as a means* to other purposes, but *good in itself*, for which reason was absolutely necessary” (G 4:396). Reason sets for us the end of moral action out of a good will, and we feel a certain satisfaction in acting to accomplish this end and are gifted with a “far worthier purpose of one’s existence to which therefore, and not to happiness, reason is properly destined” (G 4:396). From the proposition that all things bestowed upon us by nature behave purposively and because reason leads us away from happiness and towards morality, Kant concludes that “the true vocation of reason must be to produce a will that is good, not perhaps *as a means* to other purposes, but *good in itself*, for which reason was absolutely necessary” (G 4:396). Reason may be the sole architect of a good will, but reason is given to us solely to design such a will.

The section proceeds with an attempt to examine what constitutes a good will and describe it through the concept of duty, which Kant claims “contains [the concept] of a good will though under certain subjective limitations and hindrances, which, however, far from concealing it and making it unrecognizable, rather bring it out by contrast and make it shine forth all the more brightly” (G 4:397). So using duty, the good will’s subjective manifestation, Kant will attempt to explicate what it is that a good will does and how it seems. He will ignore any actions that outright conflict with duty, but will also fail to consider any action in accordance with duty the sole motive of which was not its being done only for the sake of duty. Dutiful actions done from self-interest or from immediate inclination are rejected without being considered.

For example, the preservation of one's life, beneficence, philanthropy, and the preservation of one's own happiness are all duties. A good will sets all these as ends, but they are morally praiseworthy only in that they are acted upon with a moral motivation. For Kant, the only moral motivation for these actions is that they are motivated by respect for the moral law and done out of a sense of and respect for duty. Three principles characterize Kant's discussion of the good will.

The first is illustrated with a reference to scripture and the command "to love our neighbor, even our enemy" (G 4:399). Kant illuminates it thus:

For, even as inclination cannot be commanded, but beneficence from duty—even though no inclination impels us to it and, indeed, natural and unconquerable aversion opposes it—is *practical* and not *pathological* love, which lies in the will and not in the propensity of feeling, in principles of action and not in melting sympathy; and it alone can be commanded. (G 4:399)

This beneficence from duty originates in the will and finds its fulfillment not from sympathy or "pathological love" but from the will obeying the practical law. This kind of beneficence springs from duty and not from inclination or instinct, and this is what Kant wants to focus on with this example. His first proposition explaining the concept of a good will is that a good will is commanded, not from inclination, but from duty. Thus, the "pathological love" of inclination cannot be morally commanded, while "practical love" as commanded by duty, and in resistance to natural aversion, can bring about morally praiseworthy action.

The second principle explains that an action done for the sake of duty acquires moral worth from the maxim upon which the agent acts. An action gains its moral worth:

*not in the purpose* to be attained by it...and therefore does not depend upon the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire. (G 4:399-400)

Kant explains that the moral worth of an action depends entirely upon the maxim upon which it was acted, “where every material principle has been withdrawn from it” (G 4:400). This is a sort of purity clause, ensuring that the motivation for moral action remains entirely based on an a priori principle. The maxim which serves as the a priori ground of an action, and not the material ends or effects of an action, is the determining ground of the action’s moral worth.

Flowing consequentially from the two preceding principles:

*Duty is the necessity of an action from respect for law.* For an object as the effect of my proposed action I can indeed have *inclination* but *never respect*, just because it is merely an effect and not an activity of a will. In the same way I cannot have respect for inclination as such, whether it is mine or that of another. (G 4:400)

In synthesizing the first two principles, Kant is able to thoroughly and rigidly define duty as the representation of the law within the agent, an act done solely for the sake of the moral law. From the first principle, he draws the idea that moral action cannot have as its motivation inclination, for many moral actions are directly opposed to our inclinations and desires. From the second he secures the a prioricity of moral choice. It is not respect for the ends that must motivate morally, but instead respect for duty that must move us to action. The maxim and not the material ends of an action can be respected, because acting on a maxim is an activity of the will. The outcome of an action is not. It is not within my will to see to that an action come out as I intended, therefore such chance occurrences cannot be respected, because their coming into being is mere chance or luck. A maxim of action, a subjective principle of volition, is how the moral law manifests itself in the will, so only a will objectively determined by law and subjectively determined by respect for the law, which motivates moral action, is morally praiseworthy. Respect for the law motivates moral action and necessitates in my will duty.



In a long footnote, Kant explains what he means by *respect* in the sense of *respect for the law*. He claims that this respect is “self-wrought by means of a rational concept” (G 4:401n) and different from normal conceptions of respect as fear or any other inclination. Kant continues,

What I cognize immediately as a law for me I cognize with respect, which signifies merely consciousness of the *subordination* of my will to a law without the mediation of other influences on my sense. Immediate determination of the will by means of the law and consciousness of this is called *respect*, so that this is regarded as the *effect* of the law on the subject, and not as the *cause* of the law. (G 4:401n)

His respect for the law demonstrates the binding of his will to the moral without interference from inclination, desire, or material ends. When the will is immediately determined by the law and the agent is aware of this, the agent can be said to respect the law, as he is willfully surrendering himself over to it. The agent who respects the law holds it in esteem and awe, and it is not from this respect that the law arises but from this respect that it holds sway over the agent. The law is the *object* of respect, and it is both “necessary in itself” and something “that we impose upon *ourselves*” (G 4:401n). Our respect for a morally good person springs from our respect of the law as it is represented in her.

Next, Kant sets out to determine what constitutes the moral law. Kant claims that only one set of actions and motives, those that conform to universal law, can actually be moral, so he initially posits the moral law here as:

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

He explicates the law by applying it to a moral problem, that of the lying promise, which recurs thrice in the *Groundwork*. Asking “whether it is prudent or whether it is in conformity with duty to make a false promise” (G 4:402), Kant explains that prudence may lead us to make a false promise, then reconsider this for fear, in both scenarios, of retribution and consequence later;

prudence cannot motivate a moral action. Conformity with universal law produces a moral action motivated by a priori factors that include the examination of the act in the world of the universalized maxim. In such a scenario, a lying promise would not work, because the entire system of promises would break down if all promises were lying ones. Thus, the agent cannot will the maxim of a lying promise into an act of duty, because such a maxim could not be willed into a universal law by which everyone abided. For an action to be moral, concludes Kant, it must be motivated by pure respect for the moral law and the maxim of the action must be able to be willed as a universal law of nature.

Section I concludes with the justification for the transition from common rational cognition to pure philosophy. Kant fears that while common rational cognition has led us so far as to discover the moral law, it may be easily seduced and deceive itself in the process of moral judgment. Without the help of true philosophy, he argues that we may convince ourselves that we act purely for duty's sake while we actually corrupt ourselves with some desire or inclination for the results of the action. If philosophy is not there to assist, Kant worries that no moral action may be done.

### **The Moral Law**

Section II opens with a continuation of the discussion of self-deception that closed Section I. Here Kant admits that quite frequently when we feel that we have acted purely out of duty, we have probably just blinded ourselves to our self-interest. It is here that he admits that no truly moral act may have ever been committed in the history of mankind. Nonetheless, he maintains that this in no way diminishes or tarnishes the laudability of moral action or the purity of the moral law.

Kant continues by once again reaffirming the a prioricity of the moral law, saying that there is no possible experience that can lead us to believe that the moral law holds for all rational beings. The only rational beings that we have any experience of are ourselves, and we have concluded that the purely a priori aspect of ethics, metaphysics of morals, is devoid of all considerations of purely human nature. The a priori nature of the moral law cannot be denied, nor can it be recommended that one teach by examples. While Kant admits that examples are helpful in producing intuitively moral responses and as inspiration, every moral example must have already been judged against the moral law, so it would be simpler to teach the moral law and it's a priori motivation. Kant is continuing to attempt a justification for a shift away from popular philosophy that mixes the a priori and the a posteriori to a pure metaphysics of morals. It is furthermore desirable, because a pure sense of duty and a pure metaphysics of morals holds "an influence on the human heart so much more powerful than all other incentives" (G 4:410-1). To complete this transition, "we must follow and present distinctly the practical faculty of

reason, from its general rules of determination to the point where the concept of duty arise from it” (G 4:412).

The will is practical reason, since only rational animals have wills and “the capacity to act *in accordance with the representation* of laws” (G 4:412), according to which all of nature works. With regard to humans, however, practical reason is not the sole determiner of the will; the wills of humans are also subject to desire and immediate inclination, so that practical reason is not the sole influence over our wills. We are not “by [our] nature necessarily obedient” (G 4:413). Practical reason does not possess sole dominion over our wills, and we are fallible. The holy will, however, acts in accordance with duty because it is pure; no inclinations oppose this, and the representation of the law in a holy agent is perfect and motivates the will perfectly.

### I. Three Types of Imperatives

A command that can motivate our fallible wills is called an imperative, and there are two types of imperatives: hypothetical and categorical. “The [hypothetical imperatives] represent the practical necessity of a possible action as a means to achieving something else that one wills,” whereas “the categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end” (G 4:414). Kant explains:

If the action [represented as possibly good] would be good merely as a means *to something else* the imperative is hypothetical; if the action is represented as *in itself* good, hence as necessary in a will in itself conforming to reason, as its principle, *then it is categorical*. (G 4:414)

Hypothetical imperatives are divided into *imperatives of skill*, problematically practical principles providing only for situations that may possibly arise, and *assertoric* or *prudential imperatives* that have as their purpose happiness. Categorical imperatives stand apart as those that promote an action not as a means to some end but as an end in itself; they promote an action

for the action's own sake and express its inherent worth or goodness. Having delineated these distinctions, Kant then proceeds to demonstrate how each of these types of imperatives is possible.

Imperatives of skill are easily explained, because, simply put, if one wills the ends, then she also wills the means to achieve those ends. The hypothetical imperative of skill simply expresses this relationship, for example: I wish to play the piano well, so I will practice for one hour every day. Logically, hypothetical imperatives of skill are analytic in that the means to the act are simultaneously willed with the ends. Hypothetical imperatives of prudence, those with happiness as their end, are more complex. Allen Wood in his *Kant's Ethical Thought* explains assertoric imperatives more thoroughly. Wood describes such an imperative as "a precept that prescribes universally to all rational beings the actions that are required to achieve this greatest total satisfaction with life" (Wood 65). According to Kant, for all rational beings happiness is a "purpose that they not merely *could* have but that we can safely presuppose they all actually *do* have by a natural necessity" (G 4:415) and is "a purpose that can be presupposed surely and a priori in the case of every human being, because it belongs to his essence" (G 4:415-6).

Kant rejects *assertoric imperatives* because he feels that they reduce to hypothetical imperatives that are analytic like those of skill. His argument goes:

Now, skill in the choice of means to one's own greatest well-being can be called *prudence* in the narrowest sense. Hence the imperative that refers to the choice of means to one's own happiness, that is, the precept of prudence, is still always *hypothetical*; the action is not commanded absolutely but only as a means to another purpose. (G 4:416)

What Kant is explaining is that prudence, skill in choosing those means that lead to one's greatest happiness, is just a means to happiness. Thus the imperative is set up with prudence as the means to the end happiness. The proposition is analytic, with prudence being willed

necessarily as a mean with the willing of happiness as an end. Thus prudence always commands toward another end, happiness, and so cannot command categorically. Thus, the assertoric imperative is not a synthetic a priori proposition, and cannot qualify as the moral law.

Putting a different spin on things, Wood takes the quote from Kant to mean “that it belongs to the essence of *rationality* that a rational being is bound to form the idea of its happiness and make that happiness its end” (Wood 66). Taking this conception and going further, Wood then forms a specific “‘pragmatic imperative’: Form an idea for yourself of the greatest achievable sum of your empirical satisfaction (under the name ‘happiness’) and make happiness your end, always preferring it over any limited empirical satisfaction” (Wood 67).

On Wood’s account, this imperative is that unlike any imperatives of skill, which are simply analytic, such a pragmatic imperative is both a priori and synthetic. So it could potentially compete for the title of supreme moral principle by a priori command and authority over all rational beings. This is, however, not the case. Because the concept of happiness is subjective and entirely empirical with each individual often being uncertain of what would make her happy at any given moment, such imperatives do not command at all, but rely on an invention of the imagination, happiness, for their construction (Wood 68). Wood cites claims by Kant that human beings could not even know the sum total of what actions, and their proportions to other actions, could make them happy (Wood 68). Thus, such empirical and subjective elements entail that this imperative cannot command all rational beings a priori. Wood attributes to Kant the view that such prudential imperatives do not exist at all.

Kant’s rejection of the assertoric or prudential imperative can also be seen as a rejection of prudential moral theories. Under a moral theory dictated by a prudential imperative, “the necessity of moral obligation could be represented as the instrumental necessity of actions

required by a pragmatically or prudentially necessary end” (Wood 69). “Acting from duty...could be represented as acting on the imperative to pursue one’s good on the whole or one’s happiness in the face of contingent desires that threaten to distract one from this end” (Wood 69). Thus, moral action would be promoting one’s happiness while avoiding the pitfalls of actions that may cause temporary happiness at the cost of the long-term good on the whole.

Wood finishes out his discussion:

It is therefore important for Kant that relative to the end of happiness there are no universal precepts but only general counsels, all of which are only contingently valid and admit of exceptions. That means there is no way in which moral obligation can be presented under the guise of a hypothetical imperative telling us how to achieve the sole end that we as rational beings cannot disavow (namely, happiness). From these points and the thesis that moral obligations are universal prescriptions of reason that bind us independently of our contingent ends, it follows that if there is any such thing as moral obligation at all, then it cannot consist in rational imperatives that command only hypothetically (relative to an end, perhaps a necessary end), but rather it must consist in imperatives that command *categorically* (independently of any end, even a necessary one). (Wood 69-70)

Wood’s argument shows that Kant’s rejection of both hypothetical imperatives of skill and prudential or assertoric imperatives is intended to also function as a rejection of moral theories that propound a moral principle of such types. It also establishes concretely the importance of the categorical imperative.

Two characteristics of categorical imperatives are important to understanding the difficulty in seeing their possibility. First, of all types of imperatives only categorical imperatives have the makings of law in that they describe an action as necessary. Second, a categorical imperative is a synthetic a priori proposition. Prolonged discussion on the possibility of a categorical imperative is suspended until the third section of the *Groundwork*. Kant continues by trying to determine what the categorical imperative may express.

## II. The Formula of Universal Law and the Formula of Universal Law of Nature

The content of any hypothetical imperative is contingent upon the ends it sets for itself, so that no hypothetical imperative can be known before its need arises. This, Kant claims, is not the case with categorical imperatives:

For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains not condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. (G 4:420-1)

Kant continues:

There is, therefore, only a single categorical imperative and it is this: *act only in accordance with that maxim through which you can at the same time will that it become a universal law.* (G 4:421)

This is one of the most significant passages in all of moral philosophy, the formula of universal law (FUL). Here Kant has derived, once and for all, the categorical imperative, the supreme principle of his a priori morality.

But Kant does not stop with the first formulation of the categorical imperative. He reformulates the moral imperative in the formula of universal law of nature (FLN) as “*act as if the maxim of your action were to become by your will a **universal law of nature***” (G 4:421). The difference between these two statements of the imperative lies in the different types of laws to which Kant alludes. By “universal law,” he means universal law of freedom, a dictate of morality prescribing what should be done in all similar instances. A universal law of nature is different. There is not room to disobey such a law. If nature is set up in accordance with a certain law, then it cannot ever be violated; nature runs according to it at all times, regardless of any contingent considerations. H. J. Paton in his *The Categorical Imperative: A Study in Kant's Moral Philosophy* discusses laws of nature in great depth.



Paton analyzes two types of natural law that may be functioning as universal laws of nature in the second formulation of the categorical imperative: *causal law* and *teleological law*. Causal natural law dictates that the “same cause must always produce the same effects. Hence it may be thought that if we are to find contradiction in a maxim when it is conceived as a law of nature, this must be because it would assert that the same cause could produce different effects” (Paton 148). A contradictory causal law of nature would “admit of arbitrary exceptions” (Paton 148), but it is impossible for us to generalize this into a breach of moral law, and Paton gives examples of “special circumstance” (Paton 149) that illustrate his point. One in particular is a case of poisoned food. Paton claims, “we assert no breach of a law of nature if we say that food, which ordinarily causes life, may in special circumstances—for example, in certain kinds of illness—cause death” (Paton 149). It is not entirely clear from his argument how he would like this to be generalized, as I see no way other than an appeal to common sense. It seems to be no breach of causal natural law in this circumstance. If he is appealing to the impossibility of a breach of causal natural law, then the claim is valid, as no breach of the natural may exist, thus nothing could be generalized from it. He concludes: “we may say generally that any attempt to make the causal law of nature a test of moral law is foredoomed to failure” (Paton 149). I would venture to agree with this claim.

On the other hand, Paton analyzes what he describes as teleological laws nature, and concludes that this is what Kant intends the universal laws of nature of FLN to look like. “Even in the understanding of physical nature we may have to use another concept besides that of causal law—the concept, namely, of purpose or end” (Paton 149). Teleological laws of nature look at individual parts, organs, and actions of nature in order to analyze the purpose of each of these. Whether or not this is the way the world actually is, “we must consider them *as if* they

had a purpose and see whether in this way we can understand them better” (Paton 149). For Paton this provides a type of natural law that *can* be abstracted into moral law, and he feels certain that this is the type of natural law that Kant intends his readers to see utilized in FLN. To defend his claim, he points to the purposiveness Kant assigns to human action and to human nature and claims that “this was apparently so much taken for granted by Kant that he fails to state it explicitly, and so tends to mislead his readers” (Paton 151). Paton’s claims seem legitimate, and he is the only major commenter on Kant to give a detailed analysis of the differences between FLN and FUL, regardless of his overall agenda in doing so. Paton seems to successfully explain the gap left between the two formulae that Kant himself did not fill. The application of teleological natural law to FLN as opposed to causal natural law is also a plausible account, and the parallels Paton cites between such a purposive view of nature and purposive views of human action and human nature entail convenient consequences for him, which will be developed later.

According to Paton, we may will that our maxims become universal laws of freedom “in the sense that we can will our action as an instance of a principle valid for all rational agents and not merely adopted arbitrarily for ourselves” (Paton 146). We cannot, however, will that our maxims become universal laws of nature, which are stronger, and we cannot, because we lack the power to do so. “Hence in this new formula Kant very properly says ‘Act [sic] *as if*...’” (Paton 146). Along these same lines, Paton expresses the need to apply FUL and FLN differently. On the other hand, Korsgaard, in her *Creating the Kingdom of Ends*, acknowledges a distinction between the two formulae, but not a distinction between their respective applications.<sup>1</sup>

---

<sup>1</sup> See Korsgaard in Bibliography page 77 for the acknowledgement of a distinction between the formulae, and the rest of her Chapter Three for the complete application of the moral law over the four examples.

The difference in application will become a significant question in forthcoming discussions of the contradiction in conception and contradiction in will tests.

The difference in the formulae of FUL and FLN can be seen as significant if one considers Kant's own words regarding the relationship between the two. In the *Critique of Practical Reason*, he indicates that FLN is the typic of the moral law as represented in FUL. Kant explains:

This is how even the most common understanding judges; for the law of nature always lies at the basis of its most ordinary judgments, even those of experience. Thus it has the law of nature always at hand, only that in cases where causality from freedom is to be appraised it makes that law of nature merely the type of a law of freedom because without having at hand something which it could make an example in a case of experience, it could not provide use in application for the law of a pure practical reason. (KpV 5:70)

Thus, FUL, which prescribes consistency with universal laws of freedom may only be applied through its typic, FLN, which prescribes consistency with universal laws of nature, which as Paton says, are teleological in this context. With the experiential primacy of FLN in mind, the first four examples of the *Groundwork* must be analyzed in terms of FLN and not FUL.

Kant continues by delineating four controversial examples of the application of FLN that will be discussed in detail in the next chapter.

### **III. The Nature and Structure of Maxims**

Given that FUL and FLN require us to act as though our maxims were to become universal laws and universal laws of nature, respectively, by our wills, it seems that the concept of a maxim is a crucial one in Kant's ethical theory. He is very clear on what a maxim is, and he restates this several times in his writings on moral theory. In the *Groundwork* Kant explains that "a *maxim* is the subjective principle of acting," and it "contains the practical rule determined by

reason conformable with the conditions of the subject (often his ignorance or also his inclinations), and is therefore the principle in accordance with which the subject *acts*” (G 4:421n). This is perhaps the clearest and most thorough of several similar definitions.<sup>2</sup> So a maxim is a subjective “principle of volition,” a motivating ground of the will that “the agent himself makes his principle on subjective grounds” (MS 6:225). An agent acts on a maxim with the maxim functioning as the motivation for an action.

Under Kantian guidelines, all actions are performed on such subjective principles of actions, whether such a maxim is explicit or implied. Thus, according to Kant, a man who instinctually flees from a lion chasing him acts just as much on a maxim as does a man who laboriously debates the morality of a “philanthropic” lie. In his discussion of this aspect of a Kantian maxim, Wood adds:

Adopting a maxim means subjecting one’s action to self-given norms. This includes such things as setting an end, choosing means to it, and selecting actions of a certain description to be performed under various circumstances. (Wood 78)

A maxim is, thus, a sort of tool of rationality and a demonstration thereof. A rational agent binds his actions to his wish to accomplish certain goals, and a maxim is a demonstration of that rational ability, because it requires intellect to subject oneself to the self-commanded regulations of setting an end, choosing a mean to accomplish it, and deciding what actions best suit every set of circumstances.

Kant expands his concept of a maxim saying that maxims “merely *qualify* for a giving of universal law, and the requirement that they so qualify is only a negative principle (not to come into conflict with a law as such)” (MS 6:389). The quality that allows a maxim to be generalized into universal law is its not being in conflict with the proper idea of a law as such. Here Kant is

---

<sup>2</sup> The reader is referred to several passages from G and MS such as G 4:401n, MS 6:389, and MS 6:225, to name a few.

explaining that the form of the maxim is what qualifies it for being thought a universal law. In the second Critique, the *Critique of Practical Reason*, Kant states:

Either a rational being cannot think of *his* subjectively practical principles, that is, his maxims, as being at the same time universal laws or he must assume that their mere form, by which *they are fit for a giving of universal law*, of itself and alone makes them practical laws. (KpV 5:27)

This passage is explained by Kant as assuming the principle of non-contradiction. The form by which a maxim is unfit for a universal law is one by which its universalization is contradictory. The form by which a maxim *is* fit for a universal law is one by which its universalization is consistent.

Kant does not provide a detailed or even a very clear description of what a properly formulated maxim looks like. One can infer at least some type of structure from the four examples of FLN in the *Groundwork*,<sup>3</sup> but given the prominent place of maxims in moral evaluation and the necessity of finding contradiction in immoral maxims, a clear analysis of what constitutes a properly formulated maxim appears necessary. Onora Nell provides an analysis of what constitutes a properly formulated maxim in her *Acting on Principle: An Essay on Kantian Ethics*. Her agenda in this section of the book is to derive the “Universal Principle of Justice” and the “First Principle of the Doctrine of Virtue,” which is irrelevant to the topic of this work, but she nonetheless provides insights into the proper constructions of maxims with regard to ethics, and also the proper form of maxims as imperatives.

Nell begins her discussion by analyzing maxims that may become universalized into categorical imperatives. Her simplified explanation of a maxim that may become an imperative runs as follows:

Since maxims are often adopted for personal use only, they may be very elliptically formulated. All specification of the agent and even of whether the act

---

<sup>3</sup> See the below discussion of the four examples for such an attempt.

is thought to be obligatory or permitted or to have some other status may often be eliminated as “understood,” and the maxim rendered schematically simply by (3) To - - - - if . . . . , or by (3') I will - - - - if . . . . , where “- - - - ” and “. . . . ” are as before filled, respectively, by some act description and some agent description and form together a composite act description. (Nell 35)

Nell sets up a two part maxim that describes both the act that the agent seeks to perform and the circumstances of the agent in seeking to accomplish that act. Relevance between the two parts is something that Nell assumes, and she thus names the combination of these two descriptions “a composite act description.” She claims that the categorical imperative derived from a maxim of form (3) would be expressed in “(3'') - - - - if . . . . ” (Nell 36). I agree with her when she maintains that the categorical form prescribed by FUL is the form represented in her (3''). Such an imperative gives the circumstances under which a prescribed action must take place, and it also expounds the action that must be performed.

Appealing to the first example of FLN in the *Groundwork*, she expands the two-part maxim:

As the first of the specimen maxims shows, the agent's maxim may include not only an act and an agent description, but also a *purposive component*. This does not show that not all maxims fit the schemata suggested, for the purposive component can be regarded as an amplification of the act description. In some contexts it is essential to consider an agent's purposes explicitly, and in these cases the agent's maxim may be rendered schematically by (4) To - - - - if . . . . in order to \_\_\_\_\_, where “\_\_\_\_\_” is filled by some specification of the agent's purpose in acting. Any element of the form “- - - - if . . . . in order to \_\_\_\_\_” will be called an *amplified act description*. Since Kant believes that all human action is purposive any maxim may be stated in sufficient detail to fit into the schematic form (4).<sup>4</sup> (Nell 37)

She goes on to dub a maxim “in amplified act description, as in (4)...a *complete maxim of ends*” (Nell 38). Such a complete maxim of ends expounds three things: a description of the agent, which can be the circumstances under which the proposed action would be performed, a description of the act itself, which is what the agent intends to do given her circumstances, and a

---

<sup>4</sup> The symbol “\_\_\_\_\_” is substituted for one which I cannot replicate on my word processor.

purposive element which defines the agent's motive in acting and what she hopes to accomplish by the action. This is a relatively full concept of what a maxim should contain, but Nell is careful to qualify, maintaining that such purposive content is not always a necessity for a maxim, though she acknowledges that under the Kantian rubric, all actions are purposive ("Every action, therefore, has its end" [MS 6:384].).<sup>5</sup>

Nell concludes with one final type of maxim:

When the composite act description is suppressed and we know the agent's proposed ends, but not his means, I shall call his maxim an incomplete maxim of ends. Incomplete maxims of ends have the form (4") To do/omit what is needed in order to \_\_\_\_\_. (Nell 38)

In this type of maxim, neither the agent description, nor the act description is present, only a purposive content. In the process of completing her argument, she also gives a name to maxims expressed in the form of (3) above, "*maxims of action*" (Nell 38).

Thus, Nell delineates three types of maxims that may have relevance here: *complete maxims of ends*, *maxims of action*, and *incomplete maxims of ends*. Each of these is formed of different combinations of act or agent descriptions and purposive content or motivation. This discussion of maxims is fairly complete, and it seems to me that Nell has exhausted the possible combinations that create plausible and informative maxims. Also helpful is her nomenclature, which will now be used throughout this work. This tripartite division of maxims may prove very useful in future discussions of the contradiction in conception, contradiction in will, and the three main theories explicating these concepts.

---

<sup>5</sup> Purposiveness of action is, according to Kant, inherently tied to freedom. "Since no one can have an end without *himself* making the object of his choice into an end, to have any end of action whatsoever is an act of *freedom* on the part of the acting subject, not an affect of *nature*" (MS 6:385).

#### **IV. The Contradiction in Conception and Contradiction in Will**

After the examples, Kant explains another problematic aspect of his moral philosophy: his theory of the “contradiction in conception” and “contradiction in will” (G 4:423-4). Moral appraisal under the rubric of the first two formulae of the categorical imperative focuses on the ability to will a maxim in its universalized form—indeed, not just in its universalized form, but as a universal law of freedom in the case of the first formulation and as an unconditionally binding law of nature under the second formulation. Contradictions emerge when this is impossible. Kant sees two types of contradiction, the first being one arising in thought. This applies only to perfect duties, and the test of contradiction in conception versus contradiction in will can be seen also as a test for perfect and imperfect duties, imperfect duties being tied to contradictions in will. Under the contradiction in conception, the maxim cannot even be thought of as a universal law of nature without a contradiction arising. Under the contradiction in will dichotomy fall those maxims that can be conceived of as universal law, but cannot be sensibly willed as such, “because such a will would contradict itself” (G 4:424). This is more than a little unclear. Its lack of clarity has created several problems in the realm of Kantian ethical studies, hence this thesis. It is of extreme interest that in his discussion of these two different types of contradictions, Kant appeals only to FLN, not the statement of universal law.

“We shall now enumerate a few duties in accordance with the usual division of them into duties to ourselves and to other human beings and into perfect and imperfect duties” (G 4:421).



### **Kant's Explication of the Four Examples**

Immediately after expounding the second formulation of the categorical imperative, Kant sets about explaining a set of examples, one from each type of duty: a perfect duty to oneself, a perfect duty to others, an imperfect duty to oneself, and an imperfect duty to others. Many scholars have found these examples and their respective explications to be extremely problematic, and they are. The purpose, for now, is merely to examine them and seek out the inconsistencies and similarities among them and the discrepancies from the illustrations of the examples and the rubrics regarding moral judgment that Kant defines.

#### **I. General Structure of the Four Examples of FLN**

The demarcation of duties into four classes affects the structure of each example as it is present in the *Groundwork*. Those examples which are perfect duties, and thus fall under the contradiction in conception test to determine their moral worth show a more streamlined structure than those imperfect duties that demonstrate a contradiction in the will when an immoral action is committed contrary to them. The basic structure of each of the four examples is the same: posing the problem and formulating the maxim, imagining a nature that had that maxim as a law thereof, and declaring the act immoral.

For the perfect duties, suicide (to oneself) and the lying promise (to others), a four part structure can be demonstrated. In the first section, Kant explicates the agent's situation and poses the problem. In this section he provides, using Nell's terminology, an act description and an agent description and, thus, a composite act description for the maxim he formulates in the

second step. In the composite act descriptions of both the suicide and lying promise examples, Kant is careful to explain that both agents are in full possession of their reason, thus ensuring that the action cannot be immediately eliminated from the moral sphere.

He goes on to formulate the maxim of each agent in second steps for the examples of perfect duties. Both of these maxims contain the composite act descriptions provided in the first step, but both also go further to contain purposive content regarding the action. Kant includes the motivations of both men in their maxims creating complete maxims of ends for both examples. Given Kant's explication thus far, if these examples are typical and we are to generalize from them, it can be assumed that perfect duties, whether to the self or to others, require maxims containing motivational content to be judged moral or immoral.

The third step in the analysis of these two examples is the attempted conception of a nature in which the agent's maxim held as a universal law of nature therein. In each case, Kant claims the inability to conceive of such a nature, and he proceeds to give reasons why. In each case, Kant references the purposive aspect of the complete maxims of ends he has set up. Suicide is rejected based on the self-love that motivated it. In the lying promise example, Kant outright states "it would make the promise and the end one might have in it itself impossible" (G 4:422), thus arguing that the motivation for the action is in itself undermined. Since the contradiction in conception is what is being portrayed here, it might be helpful to ask if it is essential to such a contradiction for the maxim being willed to universal natural law to contain the agent's motivation or a purposive content. It would certainly provide more material for a contradiction to be found lurking: a three part maxim leaves more room for some element to contradict another than does a two part maxim of action.

The violations of imperfect duties that Kant describes, the last two examples, also show a structural similarity. While the first two steps of each of the latter two examples mirror those of the first two, there are crucial differences. The explications of both problems are the same as above; they provide a composite action description in both instances. The first key difference between these examples and the first two lies in the absence of clearly defined or clearly stated maxims. While maxims can easily be inferred in both cases, Kant chooses not to explicitly indicate what the maxim of each agent is with regard to the imperfect duties.

The next difference lies in the type of contradiction embodied in violations of imperfect duty. In each case, the system of nature in which the maxim is a universal law can be conceived, and Kant explicates this system in each example. The actions are rejected as immoral in that a contradiction occurs in the will of the agent. Each contradiction involves the agent willing two contrary things. In each case, the willed action that contradicts the maxim as a law is introduced later, and not found in what maybe assumed to be the maxim in each case. In the case of idleness and rusting talents, Kant appeals to the idea that “a rational being...necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes” (G 4:423). To reject the actions of the man neglecting his duty to beneficence, Kant claims that

A will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (G 4:423)

Given the manner in which these are rejected and Kant’s option to exclude a well-developed maxim from his discussion, it is more difficult to try and devise a relationship between the type of maxim employed and the manner in which it is rejected in the case of imperfect duties.

In each case, the immorality of the action examined is derived from the impossibility of conceiving a nature in which the respective maxim was a law or the lack of rationality in willing such a nature. In these four examples, universal law is shown as a negative test of duty. The rubric it lays out defines permissible action in a negative fashion, as that which is non-contradictory. It is much harder to take this formulation of the CI and use it to positively derive duties than it is to condemn those actions which are contrary to duty, which Kant has demonstrated for us. Kant has shown us how to use the moral law to determine which actions are morally prohibited, but he has not yet given us a method by which FLN can give us prescriptive commands of morality. The prescription that can be devised at this point in the text is that any action committed on a maxim that does not lead to contradiction when universalized as a law of nature is at least morally permissible. So actions that are permissible can be devised as can actions that are prohibited, but no direction is given as to how to derive those actions which are morally praiseworthy.

## **II. Example One: Suicide**

In the first example, a suffering man feels that his pain outweighs the goodness of his life and contemplates whether his duty to preserve himself no longer applies. Kant takes special care to add that he is “in possession of his reason” (G 4:422), because if the agent is not in possession of his reason, then his action cannot be considered within the moral sphere; action solely from inclination, desire, or passion always proceeds from an a posteriori principle, never from an a priori principle, and is either amoral or immoral but never moral. The suffering man wishes to commit suicide and Kant defines his maxim as: “from self-love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness” (G

4:422). Kant attempts to apply the formula of universal law of nature, and he tries to conceive of a system of nature in which the man's maxim of suicide from self-love is upheld as a law of nature. Kant determines that the purpose of self-love is the preservation of life and using it to destroy life would form a contradiction in conception and violate the categorical imperative. A nature where the drive of self-preservation also caused suicide could not exist; the "maxim could not possibly be a law of nature" (G 4:422). This is a contradiction in conception. As he explains it, Kant cannot even conceive of the system of nature of which the law resulting from the suffering man's maxim was a part.

It is very clear in this example that Kant is applying FLN, and he is appealing to teleology. He claims that the natural purpose of self-love is to ensure the preservation and continuation of life. The man would be "[destroying] life itself by means of the same feeling whose destination is to impel toward the furtherance of life" (G 4:422). The teleological appeal supports Paton's thesis regarding the causal vs. teleological laws of nature. Wood agrees that the contradiction Kant attempts to derive is a teleological one. "Thus [the universalized maxim] is held to contradict an alleged (empirical) feature of the natural world as it actually exists, namely, the naturally purposive function of the feeling of self-love" (Wood 85).

In his discussion of the four examples, H. J. Paton claims that "this is the weakest of Kant's arguments" (Paton 154). It is not an uncommon objection that Kant misinterprets the purpose of self-love in his objection to the man's plan, and Paton attacks the example here. Some have argued that self-love might be intended to promote agreeableness in life, instead of mere self-preservation, so that the man's suicide is not merely permissible, but morally praiseworthy. Paton manifests this argument as such:

It might be maintained as against him that the principle of self-love, which he usually regards as a rational principle of reason at the service of desire, would be

in contradiction with itself if it did not vary in its effects according as pleasure exceeded pain or vice versa. (Paton 154)

Paton claims that Kant is mistaken as to what self-love would dictate in such a situation. There may also be problems with the maxim itself. Though Kant insists that the man is acting with his own reason, the motivational part of the maxim is self-love, which seems to be something more primal than a dictate of reason. The man may be rational on the surface, but his desire to commit suicide stems from his instinct to minimize pain and maximize happiness. Paton picks up on this weakness as well. He claims that

Kant here calls it a 'sensation', [sic] meaning by that presumably some sort of instinct of self-preservation. But this is not one of the powers which distinguish man from the brutes, where Kant's method of argument is more plausible. (Paton 154)

If this is the case, the appeal to rationality is, in this example, just a cover for action motivated by inclination, not a priori principles. Kant may be genuinely mistaken on what self-love is exactly. On the other hand, self-love may be capable of rationally motivating an agent to a morally permissible and, indeed praiseworthy, action. It seems as though Kant has not given his readers sufficient preparation for his elucidation of this example. It seems that self-love needs a lengthy discussion as Kant reveals it to us. The appeal to teleology is also problematic, because Kant does not reference teleology in his definitions of either the contradiction in conception or contradiction in will.

### **III. Example Two: The Lying Promise**

The second example is a proposed violation of a perfect duty to others in which a man is hard-pressed to borrow money that he knows very well he will not be able to pay back. He wishes to promise his potential lender to repay the money and give a lying promise; he feels that

this is the only way that he will be able to get the money he needs. This too should be an instance of contradiction in conception, and his maxim is explained thus: “when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen” (G 4:422). The man understands that such an action will be exceedingly beneficial to his future welfare, “but the question now is whether it is right” (G 4:422). Kant mentions self-love in this example as well; it seems that this is more clearly a case thereof. Self-love would encourage the man to acquire the money in order to prevent hardship and poverty to himself.

He turns the maxim of the man’s proposed action into a universal law of nature, and attempts to conceive of the system of nature produced by it. He concludes that such a system would be contradictory, because such a system “would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses” (G 4:422). In this refutation Kant relies heavily on the world of the universalized maxim—the imagined world where the maxim is a universal law, in this case, of nature. In such a world, everyone would give a lying promise whenever he was in urgent need of money. Everyone would expect all promises to repay debts to be lying promises, and no one believe them or issue them. The system of promises breaks down completely in the world of the universalized maxim. Promises would hold no water in this world and thus the lying promise gets the agent nowhere and nothing. “The law would cancel itself out and could not even be conceived, let alone willed, as a law of nature” (Paton 153).

There is an obvious inconsistency between the first perfect duty example and this one. A natural teleology of an innate human faculty is invoked as creating the contradiction in one and the world of the universalized maxim in the other. While the two are not necessarily mutually

exclusive, the styles of contradiction are strikingly different. While a system of nature in which self-love dictates suicide is contradictory, according to Kant's explanation, it is contradictory in that a faculty is being used in a manner opposed to its purpose in such a world. On the lying promise case, the natural purpose of a promise, the establishment of trust, is undermined, but the reason the resultant system of nature cannot be conceived is that the whole institution of promises is undermined in the world in which the agent's maxim is a natural law. The difference seems to come from what is being emphasized in each style of contradiction. The suicide case emphasizes the misuse of a faculty of some sort, whereas the lying promise case relies on the misuse of an action. The contradiction arises in the first case from the motivational or purposive content of the maxim, and from the act description part of the maxim in the second, lying promise, example. Paton gives a potential reason for the discrepancy between the two styles of contradiction:

Perhaps [Kant's] emphasis on the relatively external side of the action is due to the fact that he has in mind a contract which can be enforced by State law, since perfect duties to others were traditionally regarded as falling under the law of the State, and the State must necessarily take an external view of action. (Paton 154)

Paton thus attributes part of the inconsistency to convention and the expectation that perfect duties to others should be externally enforceable. This makes sense given that later in *The Metaphysics of Morals* Kant changes keys and explicates lying as a violation of a perfect duty to oneself, not merely to others (MS 6:429). This seems to support Paton's hypothesis that Kant's association of this duty with perfect duties to others is provisional and at least partly based on the convention of his contemporaries. Paton claims that Kant's case for the rejection of the lying promise, while not prudential, is not satisfactory, while others, such as Christine Korsgaard, find it suitable and acceptable, but this will be discussed at length later.



There is, however, a parallel between the way the maxims in the first and second examples are formulated. Both maxims include a motivation; both include the intended outcome, an end to suffering or the acquiring of money; and both include an action by which this will be accomplished, suicide or the lying promise. Thus from these two examples, we can infer the structure of a properly formulated maxim for perfect duties: that of the complete maxim of ends as delineated by Nell. It may be that we are to infer from this example and others what a maxim should contain, but this is problematic, as different interpretations of what the contents of a maxim should be exist.<sup>6</sup>

#### **IV. Example Three: Idleness and Wasted Talents**

The third example is of a man who prefers to give himself over to leisure even though he “finds in himself a talent that by means of some cultivation could make him a human being useful for all sorts of purposes” (G 4:422-3). This man has the means to live comfortably without developing his talents and prefers pleasure over “troub[ling] himself with enlarging and improving his fortunate natural predispositions” (G 4:423). He wonders if this is morally permissible or if it violates an imperfect duty to himself. It becomes apparent to this person that a nature could subsist in which a law of this nature is formed from his maxim. A universe can be envisaged in which all men allow their talents to rot for the sake of pleasure and idleness, but Kant appeals to contradiction in the will saying, “he cannot possibly **will** that this become a universal law or be put in us as such by means of natural instinct” (G 4:423). Kant’s claim is that “as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes” (G 4:423). Here he expounds that all rational beings will the development of all their talents, and this man should

---

<sup>6</sup> See above for a full discussion of the contents of a properly formulated maxim on one such view.

also, “since they serve him and are given to him for all sorts of possible purposes” (G 4:423). This is a contradiction in the will, because the idle man wills one thing that directly contradicts another willed action within him. In this case, his willed idleness contradicts his willed development of innate talents. Kant claims that he wills the development of these talents, because it is a necessary condition of all rational beings that they will the furtherance of their useful capacities.

Many have seen this last bit as Kant smuggling in empirical, a posteriori motivations and concern for ends as justification of immorality in this case. This is not the case, and Paton illustrates it:

If I conceive myself as having created men with all sorts of talents, I should certainly feel myself to be willing inconsistently if I willed it to be a law of nature that these talents should never be developed or used. This argument is miles apart from the argument that I ought to develop my talents because I shall find this profitable or advantageous to myself. (Paton 155)

The argument is not prudential, because Kant is not appealing to the good caused by the development of these talents. He is appealing to the rational necessity of their development if one sees himself as authoring a system of nature and putting them into humans. Instilling humans with such gifts would indicate the rational wish to develop them, which is undermined by the desire for idleness and neglect of talents, thus creating a contradiction in the will.

The first clear problem with this example is Kant’s failure to formulate any sort of clear-cut maxim on the man’s behavior—we don’t know what this person was attempting to will into a universal law of nature. A potential maxim is the neglect of talents as the benefit of my leisure, but there is no maxim that follows the structural precedent established by the previous two examples. Unfortunately, he has left the contradiction in will test very obscure. On this example it seems that the contradicting wills need not be a part of the same maxim in order for the

contradiction to emerge and show the action to be immoral. If this is indeed the case, then the lack of a precisely formulated and clearly delineated maxim is no problem. One need only have the gist of the man's maxim, namely the act description, to see the conflict in wills once the respective system of nature is authored in the agent's mind. The will to neglect talents is undermined by the necessary will of a rational agent to develop those qualities innate to her.

### **V. Example Four: Beneficence**

The fourth and final example is of a man pondering the imperfect duty to others of beneficence. This moral agent decides to cease helping and end all personal philanthropic charges. The universal law of nature seems as though it can be conceived without contradiction. Indeed, Kant admits that the world in which the man's maxim becomes a universal law of nature would be better than the actual world where "everyone prates about sympathy and benevolence and even exerts himself to practice them occasionally, but on the other hand also cheats where he can, sells the right of human beings or otherwise infringes upon it" (G 4:423).

This example can be conveniently explained if three different worlds are imagined. The first is the one cited directly above in which people lie, cheat, steal, and otherwise infringe upon the rights of others while hypocritically speaking of the goodness and necessity of benevolence and beneficence towards other men. This is no doubt intended by Kant to represent the actual world and the way the vast majority of humans handle themselves in their dealings with others; let this world be referred to as  $W_A$ . The second world is the world of this man's willing, in which his maxim holds as a universal law of nature. In this realm, people do not lie, cheat, steal, or encroach upon the rights of others, but on the other hand neither do they help others or extend charity to those in need. Let this be called  $W_M$ , for the World of the Universalized Maxim. But

there is also a third world in which no one lies, cheats, steal, or sells the rights of other. In this third world, however, everyone helps those who need it, and no one's need goes unaided, for all mankind, when able, give to those whose needs go unmet. This will be referred to as  $W_I$ , for the Ideal World. It should be clear to any rational agent that  $W_I$  is superior to both  $W_M$  and  $W_A$ , though  $W_A$  is still a morally worse world than  $W_M$ . It should also be clear that a rational agent would will  $W_I$  over the other two worlds, and it is on this that Kant bases his moral judgment.

Kant claims that this law cannot be willed without contradiction:

For a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (G 4:423)

His justification for the immorality of such an action is that the man himself would never find assistance should he be in need and desire the help of others. But the conclusion is not prudential. It is based on the consideration of the three worlds delineated above, and Kant's contention that a rational agent would will the world in which all help all those in need.

The agent can very well will such a natural system as dictated by his maxim. The contradiction is a contradiction of the will, appropriately so, since this is an imperfect duty to others. The man as a rational agent must will the existence of  $W_I$  when it comes to beneficence, but this man wills the existence of  $W_M$ , which is a morally inferior world. The man cannot possibly will both worlds as a rational agent, so that his maxim of neglect, but not abuse, of other human beings is an immoral one that must be rejected.

As Paton explains, there is no prudential concern in the condemnation of a lack of beneficence. With reference to the way Kant words his argument, Paton claims:

This way of putting the matter stresses too much the need of the individual agent for help and so gives rise to the mistaken view that Kant is concerned merely with self-interest. It is, however, true to say that our kindness will be less mixed with

condescension if we remember that we also are subject to human needs; and it is important to emphasise [sic] that we can reasonably claim help from others only if we are willing to regard the principle of helping others without thought of reward as a law which is binding equally upon ourselves. (Paton 152)

The key feature is this: a rational agent cannot will that no one help another at any time. This is based on the empirical fact of human nature that everyone requires assistance at some time or another, but the moral justification remains a priori, in that reason dictates that a will neglecting this duty is in contradiction with itself.

As can be seen in this example also, the lack of a clearly stated maxim is not damning to Kant's argument. He has convincingly shown a contradiction in wills as emerging from the neglect of fellow man and the neglect of beneficence. The only purpose a maxim would serve would be to locate the two contradicting wills in one place, so that they may be seen as contradictory with greater ease. In both the third and fourth examples, Kant has given maxims that immoral, and has declared them as such based on the ideal of something necessarily willed by a rational agent. The neglect of talents is contradictory to the necessarily present will to develop innate capacities, whereas the failure to help fellow humans contradicts the necessary will to provide assistance amongst humanity. These two contradictions are explained without the aid of well delineated maxims, and rely on the notion of necessary willing. It may be possible to, at least tentatively, hypothesize that for imperfect duties the necessary rational will that contradicts that of the agent need not be included in the maxim, but this has the effect of making any argument for the immorality of an action appear circular. It seems that Kant has already decided that both the development of talents and the aiding of others are morally required actions, so that declaring the agents of the examples as behaving immorally relies on the preconceived notion of these actions and not the form of the law as such. Thus the form of the law has not been used to determine the morally correct thing before the morally correct thing has

been invoked. I feel that this apparent circularity shows the need for a well developed theory of what Kant means by “contradiction in will,” different ideas of which will be developed in the next chapter.

### Theories of Contradiction

Three rival theories exist to explain and clarify Kant's contradiction in conception and contradiction in will. Christine Korsgaard discusses all three at some length in her *Creating the Kingdom of Ends*, and Allen Wood discusses two of these interpretations in *Kant's Ethical Thought*. The three principle interpretations of the contradiction in conception and the contradiction in the will tests are the logical contradiction interpretation which Wood, Kemp, and others have defended; the practical contradiction interpretation of Korsgaard and Onora O'Neill; and the teleological contradiction interpretation of Paton. An alternative teleological account is mentioned by Korsgaard, but not given serious consideration as a viable theory.

These three interpretations represent the majority of serious and scholarly work on the contradiction in conception and contradiction in will tests. In my opinion, the logical and practical views are inherently flawed, relying on either unsound or incomplete assumptions or deviating too far from Kant's theory and explication of the four examples, however inadequate it maybe. Paton's view seems the most plausible of the three, and I shall initially defend it. In the course of defending Paton's view, I wish to refine it and offer, perhaps, a more plausible interpretation of Kant's contradiction tests. Each theory will be tested against Kant's four examples of the Formula of Universal Law of Nature, and what the plausibility of its explanation as well as its correlation with Kant's texts will be put to the test. In addition to this, a theory must adequately explain the distinction between the contradiction in conception and the contradiction in will. For a theory to be correct, it must also explain, as mine will try to, many

other difficult examples and corroborate with Kant's other writings, for instance *The Metaphysics of Morals*, on these examples.

My discussion of these theories will follow the discussion of Korsgaard's *Creating the Kingdom of Ends*, but will include criticism from Wood, Nell, and myself. My intent here is to creatively construct, and then destruct the theories typified and presented by Kemp, Nell, and Paton in order to pave the way for a new and more convincing contradiction interpretation, which is both true to what Kant actually writes and what Kant seems to imply in his writings. My hope is that my own theory will answer some standard objections which the other three theories have trouble in handling.

### I. The Logical Contradiction Interpretation

Korsgaard cites three commentators on Kant that expound (at least something very close to) what she dubs the logical contradiction interpretation: Dietrichson, Kemp, and Wood (Korsgaard 81). Kemp<sup>7</sup> explains his version of the theory in his discussion of the lying promise example and Kant's explanation of it, saying

What Kant actually says...is not that the maxim of making false promises could not exist as a universal law of nature for very long<sup>8</sup>..., but that it could not exist at all as a universal law of nature without contradiction. And if he really means this, it is clear that the relation designated by the verb 'machen' must be one of logical or quasi-logical, not causal consequence. The argument has the effect of a *reductio ad absurdum*. If, *per impossible*, there were a universal law to this effect, then there would not be and would never have been any promises (the 'if...then' indicating an entailment-relation); but the statement that there is such a law of nature also entails that there are promises. Hence it has contradictory implications—i.e. it is self-contradictory. (Kemp 67)

He continues in a footnote:

---

<sup>7</sup> Kemp's article is actually a response to and critique of an earlier article by Jonathan Harrison, which is not significant for my intention here.

<sup>8</sup> This is a critical remark directed towards Harrison's article.



People might have used the expression 'I promise', but they could not (logically) have used it for the purpose of making a promise; for you cannot (again logically) make a promise if nobody will believe you. Although you could say 'I promise to repay the money', it would be only a statement of intention, not a promise, which requires the existence of a promise as well as a promisor. (Kemp 67)

Kemp maintains that the contradiction in conception of attempting to universalize the maxim of the lying promise is a logical one, in that the action is impossible in the system of nature of the universalized maxim. Thus, the agent attempting the lying promise would be acting on his maxim in a world in which everyone acted on his maxim. There could be no promise in this case, because the promise could not be believed in such a system. He appeals to the form of a *reductio ad absurdum* argument as an illustration of the type of contradiction he interprets Kant to mean. If a lying promise has destroyed the system of promising in the world of the universalized maxim, then the agent attempting a lying promise would find his ends completely impossible because of the non-existent system of promising. The contradiction is logical, because the universalized form of the maxim contains what Kemp interprets to be a logical contradiction by virtue of the above argument.

Korsgaard claims that lying promise example is the one most conducive to the logical contradiction interpretation (Korsgaard 81), because Kant's own words point to an impossibility of the commission of the lying promise in the world of the universalized maxim:

For, the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses. (G 4:422)

Pointing to several other instances in which Kant's language seems to support this interpretation, she states, "he says that universalizations of immoral maxims destroy themselves (G [4:]403), annihilate themselves ([KpV] [5:]27), are inconceivable or cannot be thought, and so on"

(Korsgaard 81). These descriptions all seem to suggest an internal logical contradiction lurking in universalized maxims that are immoral, thus supporting the claims of the proponents of the logical contradiction interpretation.

Korsgaard generally says that this interpretation theory looks for “something like a logical impossibility in the universalization of the maxim, or in the system of nature in which the maxim is a natural law” (Korsgaard 78). Contradictions may arise in one of two ways. In the first,

We are trying to conceive a world in which the agent (and everyone with his purpose) is making a certain sort of false promise, but at the same time we are necessarily conceiving a world in which no one can be making this sort of promise, since you cannot make a promise (of this sort) to someone who will not accept it. (Korsgaard 82)

In this first possible venue for contradiction, the agent is conceiving a world in which she is acting on her maxim, but this world is also conceived of as one in which the action she is attempting is impossible. This contradiction arises between two conceptions of one (numerically the same) world, the world of the universalized maxim.

Korsgaard gives another method for finding a contradiction:

Perhaps the clearest way to bring out a logical contradiction is to say that there would be no such thing as a promise (or anyway a repayment-promise) in the world of the universalized maxim. The practice of offering and accepting promises would have died out under stress of too many violations. Thus we are imagining a world in which the agent and everyone with his purpose is making a certain sort of promise, but also a world in which there is no such thing. And this is logically inconceivable. If universalizing a maxim makes the action proposed inconceivable, then, we can get a logical contradiction. (Korsgaard 82)

This explanation of how to find a logical contradiction in a universalized maxim involves looking at the world of the universalized maxim and imagining the agent trying to act on her maxim in this world. The result is a contradiction, because the promise will not exist in the world of the universalized maxim.

Allen Wood, in his *Kant's Ethical Thought*, explains such a logical contradiction as a contradiction in the universalized maxim when it is viewed as a law of nature:

For if [the universal law resulting from the agent's maxim] were the case then promises themselves would be impossible, but since [the universal law resulting from the agent's maxim] is a law involving the making of promises, it follows that the very conception of [the universal law resulting from the agent's maxim] as a law of nature is self-contradictory. (Wood 88)

Wood characterizes the contradiction as arising in the law of nature itself, not in the agent's two conceptions of the world of the universalized maxim as Korsgaard does. On Wood's account, the natural law is self-contradictory, because it legislates over an institution that would no longer exist if it were in place. Thus, the law of nature that results from the immoral maxim of the lying promisor, legislates promising but destroys promising and the ability of lying promisor to promise, so that it contradicts itself. He is careful to note that, should this theory be correct,

Kant would not be claiming that the very *concept* of promising would then become self-contradictory—whatever that might mean—but only that such a natural law would render it impossible under the circumstances to engage in behavior that either the agent or others could regard as an act of promising. (Wood 88)

The contradiction arises in the natural law formulated, not in the conception of what action the natural law commands, in this case promising.

While Korsgaard intends the two ways of finding a contradiction to be merely methods to discover one contradiction seen in two different lights, Wood it seems offers a different view of the logical contradiction all together. Korsgaard seems to discuss a contradiction arising from the conceptions of the world of the universalized maxim showing the impossibility of the accomplishment of the maxim due to the natural law it dictates. Wood's discussion focuses entirely on the universal law of nature stemming from the agent's maxim and its absurdity as legislating over a practice which it dictates into non-existence. On Korsgaard's account the

emphasis is placed on the impossibility of the action; on Wood's account the emphasis is placed on the impossibility of the universal law of nature. There is a key difference here.

If we examine what Kant says about the difference in contradiction in conception and contradiction in will, we see that Wood's theory is in line with Kant's claims and Korsgaard's interpretation of logical contradiction theories is not. Kant says,

Some actions are so constituted that their maxim cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* that it *should* become such. In the case of others that inner impossibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself. (G 4:424)

Kant is quite clear that the contradiction in conception occurs solely in the willing of a maxim into a universal law of nature, not in the possibility of the action prescribed in the agent's maxim, which is closer to, but in no way exactly, a contradiction in the will. Since a lying promise is a violation of a perfect duty to others, its immorality should be apparent from a contradiction in conception. With this understood, and the sources of contradiction that Wood and Korsgaard explain considered, it should be apparent that Wood's reading of the logical contradiction interpretation is more consistent with Kant's thought than is the view that Korsgaard credits him and others as having. The rest of my discussion regarding the logical contradiction interpretation will rely solely on Wood's view of how the contradiction arises, as it already seems more plausible than the view Korsgaard attributes to Wood.

The logical contradiction interpretation works best with respect to Kant's example of the lying promise in the *Groundwork*, and few accounts of this interpretation's explications of other examples are given. I will attempt to evaluate the remaining three examples of FLN from the *Groundwork* according to the precepts of the logical contradiction interpretation and then evaluate its success in explaining these examples.

In the first example, the suicide case, Kant states the maxim of the man as being “from self-love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness” (G 4:422). Wood formulates the resultant universal law of nature as such: “It is a universal law of nature that from self-love, all make it their principle to shorten life when by a longer duration it threatens more ill than agreeableness” (Wood 85). Kant assumes that the system of nature in which this is a law cannot even be conceived without contradiction. He claims

It is then seen at once that a nature whose law it would be to destroy life itself by means of the same feeling whose destination is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature; thus that maxim could not possibly be a law of nature and, accordingly, altogether opposes the supreme principle of all duty. (G 4:422)

Thus, contradiction emerges in the law that forces the faculty meant, as Kant defines it, to further life to be responsible for its premature termination. This universal law of nature can be seen as containing a contradiction. This works pretty well for the logical contradiction interpretation. The contradiction may not be exactly that described by Wood, but it can be seen as legislating the misuse of a faculty.

The complication in fitting this contradiction into the model Wood provides arises as a result of the teleological considerations that Kant writes into this example. Here Kant argues that the contradiction in conception emerges from the misuse of the faculty of self-love, which normally prescribes the furtherance of life. The contradictory forces in the maxim are the purpose of the faculty of self-love, the furtherance of life, and the purpose of the agent’s action, the shortening of life. So Wood’s account seems to work, at least partially in the suicide case, in that it corroborates at least one element of Kant’s analysis: the law itself is contradictory. It may be the case, however, that the contradiction Kant indicates lies in the concept of self-love, the

motivational aspect included in the maxim. Such a contradiction would differ from any of the three types described above, but the contradiction is not in two conceptions of the same system of nature, not in the concept of the agent's action, nor in the universal law resulting, as such, from the maxim. In this case the contradiction lies in the concept of the motivational force behind the action.

The next test of Wood's logical contradiction interpretation is the evaluation of the two imperfect duties given in the *Groundwork*. Evaluation of such imperfect duties is, however, a problem for this theory. The logical contradiction it seeks in proposed universal laws of nature works fine for the contradiction in conception, but the contradiction in the will test explicitly does not allow anything like a logical contradiction, because the universal law can be conceived of without contradiction, just not willed. Thus, a second step must be added to this argument after the contradiction in conception test has been passed. No satisfactory attempt at producing a contradiction in will theory has been produced by advocates of the logical contradiction interpretation. I will attempt at least an outline of one. An example will help here.

The third example from the *Groundwork* is that of the failure to develop one's talents. An interpretation of the example on the logical contradiction view could be provided as follows. The maxim would be the neglect of one's duties in order to promote one's own idleness. The universal law of nature, when the maxim is universalized, could be stated as: 'It is a universal law of nature that all remain idle and lazy at the expense of the development of their natural talents and abilities.' In the world of this universal law of nature, the project of idleness is not destroyed by the law that promotes it, nor is the existence of talents compromised, merely their development. No contradiction in conception as prescribed by Wood can be found. Korsgaard's misinterpretation can come in handy here, however. If we look for a contradiction by the first

method she outlines, then a contradiction in the will that follows an at least quasi-logical guideline may be able to be produced. We are conceiving of a world in which everyone wills their own idleness, but also of one in which everyone still possess talents. Kant maintains that any rational being necessarily wills the development of her talents, therefore the world is conceived of, simultaneously, as being one in which everyone neglects their talent but also as one where everyone develops their talents, thus producing something like a logical contradiction in the will. Kant concludes his discussion, “For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes” (G 4:423). The personal growth he necessarily wills “logically” contradicts the neglect of his talents and idleness that he also wills.

There is a potential rebuttal to this explanation. The maxim of the agent could simply be to promote his own idleness. The law of nature would simply be that all promote their own idleness at the expense of anything opposed. If the maxim and the universal law do not contain the idea of talents, then it seems no contradiction can emerge. The case is no doubt parallel with the forth example of the failure of beneficence.

## II. The Practical Contradiction Interpretation

An option for those dissatisfied with the logical contradiction interpretation is the practical contradiction interpretation defended by, most notably, Korsgaard and Nell. Korsgaard’s defense can be found in Chapter Three of her *Creating the Kingdom of Ends*, on which this exposition is based, and Nell’s is embodied in her *Acting on Principle: An Essay on Kantian Ethics*. Nell discusses her view only in connection with the contradiction in conception. In any case, she says, we must ask “whether the agent can consistently *simultaneously* hold his

maxim and will its [universalized form as a law]" (Nell 69). She gives an example of a bank robber:

If I intend to, say, rob a bank, I intend also some sufficient set of conditions to realize my ends and the normal, predictable results of the success of my intended action. For instance, I intend the continued existence of the bank I plan to rob, that I be neither discovered nor interrupted during the theft, and that I shall use or enjoy the fruits of the theft. These are not separated intentions which a person who intends to rob a bank may or may not have; they are part and parcel or normal intentions to rob banks. Similarly if I intend *qua* universal legislator that everyone should rob some bank, then I must also intend some conditions sufficient for them to do so and the normal and predictable results of their succeeding in doing so...and the normal result of everyone's stealing from banks is that banks will take ever greater precautions to impede and discover thieves and to prevent them using or enjoying their loot. Failing successful prevention, banks, as we know them, would close down. (Nell 70-1)

The thief necessarily wills the means to his end, the robbing of the bank, as well as the foreseeable consequences of doing so. But in judging the morality of this action, the thief must also will that all in similar circumstances follow the law of his universalized maxim and the foreseeable consequences of all these thieves acting on such a law. One of these foreseeable consequences is that banks take all necessary precautions to prevent robbery until the whole system of banking breaks down and there are no more remaining banks as a result of all these robberies. The initial thief cannot continue robbing banks because there are no more.<sup>9</sup>

From this example of the bank robber, she explains how the contradiction is manifested and generalizes:

In such cases we intend that we should be an exception to the universal law, and that the law be not really universal. In the example given, we intend both that banks continue to exist in their present form, as part of the necessary conditions for the robberies we intend, and that banks do not continue to exist in their present form, as part of the normal and predictable results of the robberies we intend. Rational beings cannot intend a society of bank robbers. (Nell 71)

---

<sup>9</sup> It could be objected against Nell that the initial bank robber's action is by no means impossible, because the closure of banks takes places a significant amount of time after his robbery has occurred. Thus, while a later bank robber would have a much more difficult time robbing a non-extant bank, the initial bank robber's action is quite efficacious, and he walks away wealthy.



The agent in such a case, wills both the universal law and that he may be an exception to the universal law that he dictates. Korsgaard explains this element of the practical contradiction interpretation,

The test is carried out by imagining, in effect, that the action you propose to perform in order to carry out your purpose is the standard procedure for carrying out that purpose. What the test shows to be forbidden are just those actions whose efficacy in achieving their purposes depends upon their being exceptional. (Korsgaard 92)

The practical contradiction interpretation takes the maxim of the action, applies it as a universal law and then examines the efficacy of the agent's maxim in world of the universalized maxim. If an action, such as bank robbing is immoral (and it is probably exceptional as well), then in the world of the universalized maxim, the system upon which the action depended would break down. The agent simultaneously willing his maxim and the universalization of his maxim would find his maxim thwarted by the break down of the system that the universalization of his maxim caused.

Kant's language most closely supports this reading of the contradiction in conception test during the discussion of the lying promise example, the second example of the *Groundwork*. The man's maxim is given as: "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen" (G 4:422). Kant claims:

I then see at once that [the maxim] could never hold as a universal law of nature and be consistent with itself, but must necessarily contradict itself. For, the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses. (G 4:422)

Under Kant's explication the system of promises breaks down under universal abuse, and the man who would lie to acquire his quick money would not have his lying promise believed. This is entirely in line with what Korsgaard and Nell view Kant's meaning to be for the contradiction in conception. The man continues to will his own maxim to falsely promise when he needs money, but he also wills the universalization of his maxim, a world in which no one believes any such promise, because all such promises have turned up false, and no one expects such promises for the same reason. His maxim is undermined by the universal law he wills his maxim to be. He cannot falsely promise anyone anything if no one accepts any promises. This interpretation has shown itself to be in line with Kant's own reasoning for this example.

It is an interesting consequence of the practical contradiction interpretation that exceptional actions are just those that are easily seen to be immoral on this account. It is almost as though an additional criterion for the morality of actions can be devised from this contradiction interpretation: Act only in that manner that your action does not depend on all others acting in the manner opposite your own. Indeed, it seems that this interpretation can be a rewording of the moral law to fit perfect duties to others such as keeping one's promises. Such an idea, and indeed the practical contradiction interpretation of the contradiction in conception, breaks down for the case of suicide.

The suicide case is more problematic for the adherent of the practical contradiction interpretation to explain. The man's maxim is: "from self-love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness" (G 4:422). In the world with the universalized maxim as a law, from self-love, all end their lives when they would be happier dead. The agent's maxim is to end his life when he would be happier dead. No practical contradiction emerges here. So everyone kills himself from self-

love, how does that affect the agent? The universalized maxim does not undermine the maxim of the suffering man. This action, on this account, is not immoral. Indeed, its maxim can be consistently willed as a universal law of nature under the contradiction in conception test, and the action could even be morally praiseworthy. Not only is this line of reasoning not corollary to Kant's in the example, it is diametrically opposed to one stringent element of his moral philosophy.

Korsgaard goes a step further than Nell and attempts to formulate at least a sketch of a response to the problem of the contradiction in the will. The practical contradiction interpretation is, Korsgaard claims, the only theory that holds the same type of contradiction applying for both the contradiction in conception and the contradiction in the will, but allowing for a difference between the two (Korsgaard 95-6). She claims:

We must find some purpose or purposes which belong essentially to the will, and in the world where maxims that fail these tests are universal law, these essential purposes will be thwarted, because the means of achieving them will be unavailable. Examples of purposes that might be thought to be essential to the will are its general effectiveness in the pursuit of its ends, and its freedom to adopt and pursue new ends. (Korsgaard 96)

She explains the contradiction in will as being a contradiction between the agent's purpose (again, the agent's purpose is essential to include in the maxim), and some purpose that is necessarily willed by a rational agent. When such purposes contradict, the agent has violated an imperfect duty either to himself or to others. She sketches a hint towards an explanation of the rusting talents and the failure of beneficence examples of the *Groundwork*:

The arguments for self-development and mutual aid will then be that without the development of human talents and powers and the resources of mutual cooperation, the will's effectiveness and freedom will be thwarted. (Korsgaard 96-7)

I will now attempt to expand these sketches.

The third example of the man who gives himself over to idleness at the expense of his talents can be explained by Korsgaard's proposed solution and, so it seems, be in line with Kant's own explanation of the example. Kant says of the man whose maxim is to neglect his talents to promote his own idleness,

[O]nly he cannot possibly **will** that this become a universal law or be put in us as such by means of natural instinct. For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes. (G 4:423)

Kant claims that the man's will contradicts itself by willing the neglect of his talents in the face of the necessary will to develop certain innate qualities present in every rational being. The interpretation that Korsgaard provides seems as though it will be right in line with Kant's. The agent here wills the neglect of his talents. This is one part of the maxim he may universalize without contradiction into a law. However, rational agents necessarily will some ends. One of these is the furtherance and development of one's own abilities and gifts. This necessary will to develop innate gifts contradicts the will of the man to neglect these gifts, thus producing a contradiction in the will perfectly in line with what Kant claims in this example.

The next case awaiting explication is that of the man withholding assistance to those in need. This man will neither harm nor help others and makes his maxim "let each be as happy as heaven wills or as he can make himself; I shall take nothing from him nor even envy him; only I do not care to contribute anything to his welfare or to his assistance in need!" (G 4:423). Kant claims that the contradiction in will arises:

For, a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (G 4:423)

The man necessarily wills that he receive help when in need, but under the law created by his maxim, he will receive none. This perfectly fits into the model created by Korsgaard. It relies on the idea of the necessity of a rational being willing assistance when in need and the world of the universalized maxim for its thwarting. When this man's maxim is made a universal law, he is deprived of any assistance he may need should the need arise, but he necessarily wills that he receives assistance in the event that he needs it. His will has contradicted itself in exactly the way Korsgaard and Kant describe the situation.

It is interesting to return, now, to the suicide example and see how this new contradiction in the will test works for that case. No contradiction in conception could be found for such a case, but the contradiction in the will test may provide a way for the suicide case to be declared immoral such that even if the arguments for it do not coincide, at least the judgment of immorality can be brought into line with Kant's own. The agent here wills his own death, but any rational agent necessarily wills the furtherance of humanity in his person and the development of himself as a rational agent. The action can be seen as immoral by the contradiction in the will test. But this is counter-intuitive on Kant's view. This would make the duty not to commit suicide an imperfect duty, whereas it is very obviously a perfect negative duty to oneself. Korsgaard admits that this case is problematic for her theory, and indeed claims that many such "natural actions" (Korsgaard 84) are problematic for both the practical contradiction theory and the logical contradiction theory, but this will be discussed later. The practical contradiction theory cannot be seen as consistent with Kant's moral theory as a whole if it cannot account for the derivation of the types of duties based on the contradiction in conception and the contradiction in will tests.

### III. A Critical Note on the Logical and Practical Contradiction Interpretations<sup>10</sup>

One may notice that there are some striking similarities between the logical and the practical contradiction interpretations as they have been explained. One of the most compelling similarities is the almost identical manner in which they analyze the lying promise example. In her discussion of conventional actions, in which the lying promise figures, Korsgaard explains the distinction in this way:

On the Logical Contradiction Interpretation, the contradiction arises because the agent wills to engage in a conventional action, but he also wills a state of affairs in which that kind of action will no longer *exist*. On the Practical Contradiction Interpretation, the contradiction arises because the agent wills to engage in a conventional action, but he also wills a state of affairs in which the action will no longer *work*. (Korsgaard 97)

She acknowledges the similarity between the two theories on such cases:

When we are dealing with an action that falls under a practice, the two views are readily confused, because the reason the action no longer works is *because* it no longer exists. But on the Practical Contradiction Interpretation it is the failure of efficacy, not the non-existence, that really matters. (Korsgaard 97)

This, I believe, is a pseudo-distinction. On the practical contradiction interpretation, the failure of efficacy is completely dependent upon, at least in the case of conventional actions, the non-existence of the action. This is a unidirectional relation, the non-existence of the action is what makes the efficacy of the action fail. If the action type in question no longer exists in the world of the universalized maxim, then of course any specific instance of that action type will prove to be ineffective—it cannot be done in the first place. This is not a reversible relationship, the inefficiency of an action cannot be the grounds for its non-existence. The lack of efficacy of a type of action does not necessarily breed its non-existence.<sup>11</sup> Just because an action may not be

---

<sup>10</sup> I am indebted to Husain Sarkar for initially suggesting to me the striking similarities between the logical contradiction and practical contradiction interpretations.

<sup>11</sup> Korsgaard may take exception to this premise, especially in cases involving what she calls “conventional actions”, but it seems *prima facie* true to me. See Korsgaard 84-5.

effective does not mean that one cannot commit it, whereas if an action is non-existent, then no one can effectively accomplish said action. Thus, the type of contradiction appealed to in the practical contradiction interpretation is dependent upon the non-existence of the action appealed to in each case. With this established, it seems apparent that the most important aspect for the practical contradiction interpretation is the non-existence of the action.

On the other hand, the logical contradiction interpretation can also be seen as entailing the practical contradiction interpretation. If an action is non-existent in the world of the universalized maxim, then, *prima facie*, it will not be an effective method for bringing about one's ends. The non-existence of an action entails the inefficacy of that action, and thus the logical contradiction interpretation entails the practical contradiction interpretation. I would, therefore, claim that the logical contradiction interpretation and the practical contradiction interpretation make exactly the same claims regarding the contradiction in conception.

An analogous argument can be constructed for the contradiction in will test as well. The inefficacy of contradiction in will can be seen as dependent upon the non-existence of the action in question in the world of the universalized maxim. Relying on the above proposed theories of contradiction in will,<sup>12</sup> one sees the non-efficacy of an action, on the practical contradiction interpretation, as the proposed action undermining something that a rational agent necessarily wills. Likewise, on the logical contradiction interpretation, the non-existence of an action results from willing two contradictory wills into the world of the universalized maxim. In either case the notion of something necessarily willed by a rational agent is essential. On the logical contradiction interpretation, the non-existence of the willed action springs from its contradiction with an action necessarily willed by a rational agent. For the practical contradiction interpretation, the inefficacy of the action can be seen as springing from this as well. An action

---

<sup>12</sup> See pages 50-1 and 55-7.

producing a contradiction in will would necessarily be inefficacious if it was non-existent in the world of the universalized maxim. If it is a universal law of nature that no one may act in such a manner, then the action cannot bring about its desired end, because it could never be committed in the first place. Conversely, its impossibility necessarily entails its inefficacy, as the inability to commit the action entails the inability to effectively commit the action. So, on the contradiction in will, the practical contradiction interpretation can be seen as disintegrating into the logical contradiction interpretation again.

Using the sketches for theories of contradiction in will sketched above for the logical and practical contradiction interpretations, the contradiction in will of the practical contradiction interpretation can be seen as essentially the same type of contradiction as the contradiction in conception, only contained in a different medium. Let me try and show this. The sketch that Korsgaard provides assumes a set of purposes essential to the will of a rational agent. These purposes would be things that any rational agent would necessarily will. An immoral action is one that contradicts these essential purposes. Thus, in the rusting talents example the agent's action is immoral because it inhibits his "freedom to adopt and pursue new ends" (Korsgaard 96). Thus, the proposed action undermines the efficacy of that will that is necessarily, indeed essentially, willed by a rational agent. The sense of contradiction is exactly the same here as in the contradiction in conception, the contradiction is merely carried in the will. In the world of the universalized maxim, the universalization of the proposed action would universally undermine the essentially willed "freedom to adopt and pursue new ends." The efficacy of those tendencies necessarily in place in the will of a rational agent is undermined in the world of the universalized maxim. Thus, the style of contradiction is similar in cases of both contradiction in conception and contradiction in will.



The next step is to show that this lack of efficacy flows from the nonexistence of the action in question. Using the logical theory of contradiction in will sketched above, we can see that the failed efficacy results from the nonexistence of the action in question. In the rusting talents example, the logical view can be amended to interpret the immorality of the action as springing from the nonexistence in the world of the universalized maxim of allowing talents to lay idle, because the agent must also necessarily will the universal development of those talents as an embodiment of the “freedom to adopt and pursue new ends.” If this argument holds, then even in the case of the contradiction in will, the failure of efficacy that the practical contradiction interpretation relies upon can be seen as dependent upon the nonexistence of the action attempted in the world of the universalized maxim.

#### **IV. The Teleological Contradiction Interpretation**

The third major contradiction interpretation is the teleological view of Paton, partly endorsed by Lewis White Beck. As discussed above, Paton contends that the natural law envisaged in FLN is a teleological law of nature and not a causal one, and this interpretation of FLN affects the way that Paton thinks it is applied by Kant in the four examples of the *Groundwork*. Paton distinguishes two types of natural law in Kant’s writings: causal and teleological.

In its strictest sense Kant usually takes a law of nature to be a causal law, and it is essential to such a law that it should have no exceptions: the same cause must always produce the same effects. Hence it may be thought that if we are to find contradiction in a maxim when it is conceived as a law of nature, this must be because it would assert that the same cause could produce different effects. (Paton 148)

Natural law in its strictest form is causal, in which the same cause always produces the same effect. He attempts to analyze the suicide example in terms of causal law, and determines that

the cause of self-love usually produces the end of the furtherance of life, but it is causing the agent's death in this case (Paton 148).

We are, in short, conceiving a law of nature to admit of arbitrary exceptions, and so we are falling into a contradiction. From this the inference is made that the maxim must be contrary to moral law: it is a breach of a perfect duty to oneself. (Paton 148)

Thus, the suicide is immoral because the causal law of nature is arbitrarily breached, and the maxim violates the conditions of FLN.

Paton, however, remains unsatisfied by this analysis. He says:

Kant is right in saying that there can be no arbitrary exceptions either to a moral law or to a law of nature: but it is manifestly impossible to find, by this method, breaches of a law of nature from which we are entitled to infer breaches of a moral law. We assert no breach of a law of nature if we say that food, which ordinarily causes life, may in circumstances—for example, in certain kinds of illness—cause death. Nor need there be any breach of causal law if self-love, which ordinarily causes life, should in special circumstances cause death. Indeed we may say generally that any attempt to make the causal law of nature a test of moral law is foredoomed to failure. (Paton 149)

The basis of his argument here is that, on some noteworthy exceptions, some natural causes may produce contrary effects than they normally would. There is no violation of a causal law if various circumstances come together to produce effects contrary to the typical effects of a certain natural cause. From this, Paton generalizes that no causal law could be the basis of a test for moral law.

Paton, however, maintains that “in nature causal law is not the only kind of law recognized by Kant” (Paton 149). Kant makes the distinction between the causal laws of nature and the teleological laws of nature, which references “the concept, namely, of purpose or end” (Paton 149). Teleological natural laws consider organisms and their faculties “*as if* they had a purpose” in order to see if “in this way we can understand them better” (Paton 149). He also

maintains that this is the type of law that Kant intends for us to utilize in the tests of FLN. He explains:

When we are asked to conceive a proposed maxim as a law of nature, we must conceive it as a teleological law of nature; for it is a maxim of action, and action as such (quite apart from moral considerations) is essentially purposive. Furthermore, we are asked to conceive it primarily as a law of human nature, even if we are setting it against the background of nature as a whole; and human nature must be regarded as essentially purposive. (Paton 151)

Maxims of action must be universalized into teleological laws of nature, with the purposive element of the maxim taken as the teleological purpose of the action or faculty in question. This is easy to formulate, because all human action is purposive.

In universalizing the maxim of an immoral action, Paton claims that a contradiction will arise among teleological purposes. He describes the universalization test as such:

When we ask whether we can *will* a proposed maxim as if it were to become thereby a law of nature, we are asking whether a will which aimed at a systematic harmony of purposes in human nature could consistently will this particular maxim as a law of human nature. (Paton 151)

Paton conceives of the teleological system as being a harmony of human purposes, not necessarily a harmony of all natural purposes, and an agent evaluating his action under FLN must consider his maxim and its effect upon the teleological system he authors in the maxim's universalization.

Beck accepts and acknowledges the importance of the causal law/teleological law distinction with regard to FLN, and he agrees with Paton, at least in part. Beck, however, accepts that causal laws could be used as a basis of interpretation for FLN. He discusses an example of lying in this context.

The first test of a maxim is the mere universalizability of the maxim, i.e., the interpretation of it as a descriptive universal principle. Of some maxims, Kant says that they destroy themselves if made universal. That is, the effectiveness of such a maxim is dependent upon the fact that it does not correspond to a

universal, or even general, description of human behavior. It is not possible to will rationally that such a maxim should be universal, even though the proposition "All men should lie" is not logically self-contradictory. A maxim like "I should lie" depends for its effectiveness upon the fact that it is not universal, that its theoretical correlate "I lie" is not universalizable into a judgment, "All men always lie"; for, if it were, there would be no such thing as a lie at all. One's lies show mendacity and cleverness only because they are exceptions to a general rule. But general rules which have exceptions are not laws of nature; the latter have no exceptions. (Beck 159-60)

Beck walks us through an example of the application of FLN in a case of lying, and uses the concept of universal causal law of nature as the determining ground of the maxim's immorality. Thus, he disagrees with Paton's claim, saying, first, that causal law can be taken into account as having a bearing on moral considerations. The second point of disagreement between Beck's interpretation and Paton's, involves the style of contradiction Beck picks out in his example. In this passage, Beck is claiming that the universalization of the maxim of lying is prohibited because lies would not, as such, exist in the world of the universalized maxim. This line of argument makes Beck a partial adherent to the practical contradiction interpretation. The efficacy of the action is undermined because everyone else is acting similarly and the action cannot be committed in the world of the universalized maxim as a result.

Beck does, at least partially, commit himself to Paton's view that teleological laws of nature are of importance to FLN. On his account, this systematic harmony of ends is merely "an ideal to be achieved in action; it is a regulative Idea for practice and not for knowledge only" (Beck 160). This view is parallel to Paton's. His discussion reaches to Kant's other writings on teleology and the idea that it is a useful way of looking at the world, but does not represent the way the world really is. "This regulative conception of nature, believed by almost all eighteenth-century scientists and philosophers to describe the actual cosmos, is the model for our thought of the moral realm" (Beck 161).

Beck discusses a method of application of FLN. "I ask...whether I, as creator of the world in which every part should have its natural place and function, would will that certain maxims should have the force of law" (Beck 161). He asks if someone authoring a system of nature would create one in which "the natural ends of things would be systematically thwarted?" (Beck, 161) So that the idea of non-contradiction he endorses is similar to Paton's:

That is, when I will an immoral action, one that would sow discord among rational beings, I will according to the maxim of the act and also will (tacitly) that my maxim *not* be universal. And such a maxim is not then analogous to a law of nature. (Beck 161)

This is at least similar to the contradiction in will test that Paton offers. Beck discusses the contradiction in wills of a maxim that is immoral, and he does so within the context of a discussion of the harmony of purposes that considerations of teleology lead us to seek in the moral sphere. So given Beck's discussion, it seems that he does partially endorse a teleological view similar to Paton's, but only for the contradiction in will test. His views on the contradiction in conception test identify him as an adherent of a theory similar to Korsgaard's.

Paton identifies three ways in which universalized maxims affect a systematic harmony of human purposes:

We can test them so far as they affect others, by considering how far the universal adoption of these maxims would further, or fail to further, or would actually destroy, a systematic harmony of purposes among men. (Paton 156)

Thus an agent's action may be judged permissible, a violation of an imperfect duty, or a violation of a perfect duty, respectively. Thus, these three categories provide the full spectrum of moral evaluation, even with the easily built in ability to distinguish between perfect and imperfect duties. Paton's interpretation of the four examples will be helpful here.

Paton deems Kant's example of suicide to be the weakest of his four examples (154). Paton maintains that

When Kant considers duties towards oneself, he does not test maxims by their fitness to produce a systematic harmony of purposes among men if they were to become universal laws of nature. He does, however, test them by reference to harmony of purpose, a harmony between the ends proposed by the maxim, when universalised [sic] as a law of nature, and what he calls 'purposes of nature'. (Paton 154)

The universalization of a maxim must not, with regard to, presumably perfect duties to oneself, undermine the harmony of purposes proposed by the maxim in question. Thus, a maxim will set up a set of purposes, and this set must be consistent and harmonious if the action is to be morally permissible. In the case of a perfect duty to oneself, there will be no harmony if the maxim is immoral. In the suicide case, Kant defines the purpose of self-love as the furtherance of life, and Paton is careful to reiterate "that this is its purpose of function and not merely its effect" (Paton 154). Paton goes on to explain Kant's conclusion and derivation of the contradiction in the universalization of the suicide's maxim:

If I conceive myself as having created man and given him self-love with this end in view, can I will, or even conceive, it to be a law of nature that this self-love should in certain circumstances aim at producing death? Kant's answer is 'No.' (Paton 154)

His analysis of Kant's explanation is strictly in line with the text of the *Groundwork*, as he analyzes the contradiction as arising between the purpose of self-love, continuance of life, and the end of the agent's action, his own death.

Paton argues that Kant's explanation is not entirely adequate, however. Paton is willing to conjecture that Kant has already assumed suicide to be an immoral action, because "Why should it not be a merciful dispensation of Providence that the same instinct which ordinarily leads to life might lead to death when life offered nothing but continuous pain?" (Paton 154). Paton's view reaches back to criticisms of Kant's concept of self-love. His view is that self-love may in fact be concerned with the furtherance of pleasure over pain. On such an account, Kant

may be taken as referring to the baser instinct of self-preservation and not self-love in his discussion of the suicide example.

It is unclear to me as to why Paton would go on to critique the efficacy of Kant's example. In order to prove his interpretation of the contradiction in conception and contradiction in will test, Paton need not show that his view is the most plausible (though this would be a strong encouragement to adopt the view), he need only show that his interpretation is the most in line with the text as Kant wrote it. It seems that in the case of the suicide example, he has succeeded where the logical and practical contradiction interpretations fail, and he need not worry about the overall plausibility of Kant's analysis, as his explication seems the most accurate. He does, however, go on to cite that "the essential thing for our present purposes is, however, not the plausibility of the particular argument, but rather the general principle on which the argument is based" (Paton 154-5). With this given, it seems that Paton's analysis of the contradiction in conception is the only one of the three that can readily accommodate Kant's explication of the first example, regardless of the flaws inherent to it.

In the case of the lying promise, the man is contemplating promising to repay a loan in order to secure it in the first place, even though he knows that there is no way he will be able to repay it as promised. Paton's analysis runs:

If [the] maxim were to be a universal law of nature, so that everyone in financial difficulties made similar false promises, this would defeat the very purpose of such false promises. The law would cancel itself out and could not even be conceived, let alone willed, as a law of nature. Here again, it must be remembered, a teleological law of nature, on Kant's view, must assert the adequacy of every organ for its purpose and could not admit of purposes which defeated themselves. (Paton 153)

Paton's analysis is similar to Korsgaard's on the practical contradiction interpretation, but he emphasizes the defeat of the purpose of the false promise as the source of the contradiction in conception. He claims that Kant's analysis

does not offer a satisfactory basis for moral judgment unless we make the further assumption that the keeping of such promises and the mutual confidence thereby aroused are essential factors in the systematic harmony of human purposes. (Paton 153)

Under Paton's view, the contradiction in conception cannot be seen unless promise keeping and mutual confidence are included in a harmony of human purposes. If this is not the case then Paton's view breaks down on this example. According to him, the lying promise is prohibited, because the harmony of purposes, of which mutual confidence is a part, is disrupted by the intended deception of the man in this example. If this assumption regarding the harmony of purposes is absent, it seems that Paton's interpretation could not adequately explain Kant's analysis of the example. His first description of the example seems to favor the practical contradiction interpretation's view of the example, as the purpose and the efficacy of the action are closely tied in this case. Paton claims that in this case, Kant is "[taking] the principle of our action and [asking] whether if universalised [sic] it is compatible with a systematic harmony of purposes in society" (Paton 153).

I find Paton's analysis of this perfect duty to be a bit more dubious than his analysis of the suicide example. In the suicide case, Paton's interpretation added little to what Kant actually says, and he showed that Kant's analysis was in line with his interpretation. In this case, however, he clearly has to include ideas not overtly discussed by Kant in connection with this example. While Paton's interpretation of the contradiction in conception arising in this case is plausible, it seems that the practical contradiction interpretation offers an alternative more closely following the explicit reasoning of Kant.



A man allowing his talents to rust and giving himself over to idleness contemplates the morality of his maxim in Kant's third example. Kant claims that a rational being has a duty to develop his talents because he has been given them for many different purposes. Paton's interpretation is that "if I conceive myself as having created me with all sorts of talents, I should certainly feel myself to be willing inconsistently if I willed it to be a law of nature that these talents should never be developed or used" (Paton 155). In the world of the universalized maxim, since the universal law resulting from the man's maxim can be conceived, it is clear that if I authored a world in which men were given talents, but also willed that they neglected these talents, then my will would conflict with itself. By willing that men possess talents, I also will that men develop those talents, at least a rational author would. I contradict myself by willing that men do not develop those talents. The harmony of purposes is not destroyed, because such a world can be conceived of, but it is not fostered by such a universal law, therefore the man giving himself over to idleness would be acting immorally on such a maxim. This is the nature of Paton's contradiction in conception and contradiction in will test. With regard to perfect duties, immoral maxims *destroy* a systematic harmony of ends, while with regard to imperfect duties, immoral maxims fail to *foster* such a systematic harmony of ends (Paton 150).

In his discussion of this example, Paton claims that

Here we cannot start, as in the case of duty to others, merely with the purposes which men naturally have and proceed to ask how they can be combined in a systematic harmony. We have to argue that man has a duty to himself as a man to use his powers for the purpose inherent in themselves, and above all to develop and use the powers in virtue of which alone he is a man. (Paton 155)

There is a significant defect in Paton's reasoning in this passage: it is circular. FLN is a negative test of duty, meaning that it can readily determine those actions that are immoral, but cannot, at least easily, aid in finding positive duties. Thus, the idea that man has a duty to

develop the faculties within him is an idea that has already been derived from another formulation of the categorical imperative. It is different to say, as Kant does in this example that a rational agent must necessarily will the development of his faculties, but tie this necessity of development to the concept of rationality, not to duty initially. That such development is a duty stems from this more primary fact. The objection here is essentially an accusation of circularity in the argument. Neglecting one's talents is immoral, but only because developing one's talents is the morally right action. The pre-existing assumption that the development of talents is the moral thing to do is an assumption that is made before the action in question is evaluated. Thus this example relies on a moral evaluation that has not yet been made for its conclusion to stick.

The fourth example of the man who refuses to aid any who need it is an example of an imperfect duty to others, so on Paton's teleological contradiction interpretation the man's universalized maxim will fail to promote a systematic harmony of human ends or purposes. Paton acknowledges that the universal law of nature resulting from the man's maxim could subsist as a universal law of nature, but he finds, as does Kant, that such a will would contradict itself, "for since each of us at some time is bound to seek help for himself, he would thereby will an exception to this law, and consequently he could not will it to be a law" (Paton 152). The analysis here, of course, is not empirical, though it relies on empirical assumptions regarding human nature. "The argument turns rather on the fact that human beings are in need of mutual help, and that only by means of mutual help can the systematic harmony of their purposes be attained" (Paton 152). Paton's analysis seems to go one step further than Kant's in this example, and he is drawing on discussions from *The Metaphysics of Morals* as well as the *Groundwork* example. Paton does not emphasize the need for mutual assistance naturally present in the human condition, but instead focuses on the teleological framework thereof. He fits this need

within the overall harmony of human purposes and the agent's maxim is declared immoral, because he necessarily wills, as a rational being, mutual human aid and assistance, but is also willing the neglect of his fellow man—a contradiction in will. It is of interest that this example alone makes use of a discussion of human nature and not just rational nature. If it is mutually aid and assistance tied only to the idea of humanity and not rationality, it may have interesting consequences for Kant's moral theory on this point.

### **V. An Objection of Herman's and the Seeds of a New Contradiction Interpretation**

In her "Mutual Aid and Respect for Persons," Barbara Herman develops and addresses several casuistical questions surrounding the fourth example of beneficence. She shows that a maxim of non-beneficence produces a contradiction in will regardless of the attitude toward risk taking of the agent, as well as the agent's stoicism, resources, or wantonness. She emphasizes the importance of two empirical aspects of humans: "their dependency and their true needs" (Herman 590). She does, however, admit that, "Nonetheless, one may feel, for moral reasons, that the role of dependency in the argument introduces a disturbing element" (Herman 590). She explains the "disturbing element" thus:

It was an important result of the argument for beneficence that among dependent, vulnerable rational beings capable of mutual aid, variations in such things as risk tolerance, or resources, do not affect the application or stringency of the duty of beneficence. But suppose there are rational beings who are not vulnerable and dependent (call them angels); the argument for beneficence could not require them to reject a maxim of nonbeneficence toward human beings. (We suppose that they are in a position to intervene in human affairs.) Angels could will a world in which no one is able to help since they cannot need help. Would they then have no duty of beneficence toward human beings? Should not all rational beings have the same duties? (Herman 590)

The question Herman raises is whether or not another type of rational being, which was wholly or entirely self-subsistent should have a duty to assist humans when it could, since it

shares with humans its rational nature. Humans are bound to a duty of beneficence regardless of their risk tolerance and other such considerations, because it is a common feature of humanity, an aspect of human nature, that as rational beings we are vulnerable and dependent and require, some more frequently than others, the assistance of those surrounding us in order to accomplish our ends. In this thought experiment, Herman conceives of angels that are rational beings, but that are also self-subsistent. These angels are not dependent, rational beings; they are rational beings that are non-dependent on anyone for anything. Thus, they have no need for others to fulfill a duty of beneficence towards them.

The explicit reasoning of Kant is that all rational beings are bound to the same moral law, regardless of any contingent, empirical elements of their nature:

If we add further that, unless we want to deny to the concept of morality any truth and any relation to some possible object, we cannot dispute that its law is so extensive in its import that it must hold not only for human beings but for all *rational beings as such*, not merely under contingent conditions and with exceptions but with *absolute necessity*, then it is clear that no experience could give occasion to infer even the possibility of such apodictic laws. For, by what right could we bring into unlimited respect, as a universal precept for every rational nature, what is perhaps valid only under the contingent conditions of humanity? And how should laws of the determination of *our* will be taken as laws of the determination of the will of rational beings as such, and for ours only as rational beings, if they were merely empirical and did not have their origin completely a priori in pure but practical reason? (G 4:408)

To deny, says Kant, that the moral law holds for all rational beings and not merely human beings is to deny the very truth of that moral law. The moral law cannot hold only under empirical circumstances that would admit of exceptions, and no a posteriori principle that allows such considerations could hold as an apodictic moral precept. Those characteristics that comprise human nature are not necessarily present in the contingent natures of *other* rational beings, so that a moral law based on this nature would be a different moral law from that dictated to a

rational being with a non-human nature. The moral law must hold for all rational beings, however, because it is founded upon principles of rational nature and pure practical reason.

Herman seems to accept this doctrine of Kant:

All rational beings are subject to the same fundamental practical principle—the Categorical Imperative. This is all that follows merely from the fact of their rationality. Not all rational beings will have the same duties. The duties they have (that is, what follows if they apply the CI to their maxims) vary as their natures vary...So the mere fact that angels might not have the same duty of beneficence that we do should not in itself pose a problem. (Herman 591)

But it does pose a problem, a serious one at that. Herman's line of argument relies on the idea that all rational beings are bound *only* under the CI, and that the duties the CI prescribes changes from rational being to rational being. "Not all rational beings will have the same duties," she says (Herman 591). Only the CI flows from the idea of a rational nature; specific duties draw from empirical, contingent facts about the nature of each type of rational being. Therefore, each type or class of rational being would be bound to a different set of duties. Each rational being with a distinct rational nature is bound not only to the CI, but also to a set of specific duties that are binding only to that rational being and other rational beings with the same contingent nature.

Herman is not quite persuasive on this point, and my discontent is fueled by Kant's own words:

[I]t is of the greatest practical importance not to make its principle dependent upon the special nature of human reason—as speculative philosophy permits and even at times finds necessary—but instead, just because moral laws are to hold for every rational being as such, to derive them from the universal concept of a rational being as such, and in this way to set forth completely the whole of morals, which needs anthropology for its *application* to human beings. (G 4:411-2)

This passage claims that "special" human reason cannot be the basis for the moral law, because the moral law is intended to hold for all rational beings and must be derived from the concept of

a rational being in order to hold for all such beings. All of morals must be constructed in this manner, flowing from the principle governing all rational beings. Human nature and a practical anthropology are needed only when applying the moral law to human beings.

Expanding on Herman's own example, suppose that there is a community, not just of angels, but a mixed community of angels and humans.<sup>13</sup> As a part of their contingent, empirical natures, humans sometimes require the help and assistance of others, whereas the angels are independent of anyone for anything. On Herman's line of argument, the humans have a duty of beneficence to one another, but the angels have no duty to assist humans as they can when needed, even though they are members of the same community to which humans also belong.

But this seems contradictory to Kant's own words on the subject:

[I]t is of the utmost importance to take warning that we must not let ourselves think of wanting to derive the reality of [the CI] from the *special property of human nature*. For, duty is to be practical unconditional necessity of action and it must therefore hold for all rational beings (to which alone an imperative can apply at all) and *only because of this* be also a law for all human wills. On the other hand, what is derived from the special natural constitution of humanity—what is derived from certain feelings and propensities and even, it possible, from a special tendency that would be peculiar to human reason and would not have to hold necessarily for the will of every rational being—that can indeed yield a maxim for us but not a law. (G 4:425)

Kant explicitly states that even specific instances of duty must be derived from the concept of a rational being as such and not from an contingent, empirical nature of the being in question. If Kant were making an appeal to human nature, and not solely to rational nature, to ground not only the CI but also our individual duties, then he would find himself lost in relativism.

In our example, the angels and the humans, as members of the same community, would be bound to different duties of beneficence (namely, humans would have that duty to one another, while the angels would have none), and thus different members of *one* community

---

<sup>13</sup> I am indebted to Husain Sarkar for proposing this important objection to me.

would be bound to different duties based solely on their contingently true characteristics. This is relativism, and it is decidedly anti-Kantian. Along these same lines, one may ask the question: what if human nature changes over time? If this is the case, then each successive generation of humanity may be bound to increasingly different duties. If either of these situations is plausible, Kant's moral philosophy runs the risk of disintegrating into relativism, something to which Kant would, no doubt, have been opposed.

The only common ground among the members of the angel/human community is their rational natures, not their contingent, empirical natures, and this must be the basis of the moral duties in such a community. The threat of relativism is avoided by deriving not just the CI, but also individual instances of duty from the concept of a rational being as such, as Kant prescribes in the *Groundwork*. If the duty of beneficence is derived from the concept of rationality, and not human nature, then relativism can be avoided, and Kant's moral theory stands as he intended it, and the fourth example is not as problematic as Herman would have it be. Indeed, the duty of beneficence cannot be derived from merely human nature, or else Kant has left a massive gap in the consistency of his ethical theory. Therefore, both the CI and individual duties must originate in the common ground of rational nature.

The reason I discuss this issue here, while dealing with Paton's teleological contradiction interpretation, is that his theory seems most prone to this line of criticism. Paton wants to define the type of contradiction that Kant intends for the contradiction in conception and contradiction in will, and he makes use of the idea of a harmony of human purposes in his teleological contradiction interpretation. The appeal here is explicitly to human nature through a harmony of human purposes. For example, in his discussion of the fourth example, he says, "The argument turns rather on the fact that human beings are in need of mutual help, and that only by means of

mutual help can the systematic harmony of their purposes be attained” (Paton 152). The appeal is to human nature, and not rational nature. Beings that are not dependent on anyone for anything would have no duty of beneficence. Thus, Paton’s view, like Herman’s view, seems susceptible to the criticism that it leads to relativism. Kant’s moral philosophy is not relativistic, nor should any contradiction interpretation allow it to become so. A revision of Paton’s theory must be made in order to prevent this.

From here, a new contradiction interpretation can be posited. Herman goes on to produce a sort of proto-theory of contradiction in will in the rest of her article. A full exposition and development of this hypothesis will not be developed immediately, but the main points are of extreme interest here. In a longer passage, Herman attempts to justify and ground the duty of mutual aid:

As we are rational agents, we set ends. We are able to formulate and act from a conception of the good. If to set ends is to put oneself to the realization of more or less complex goals and projects, one respects one’s humanity in oneself by developing those capacities needed to realize a wide variety of ends (*DV* 51). Thus an imperfect rational being must acknowledge the obligatoriness of developing his powers and talents: they are necessary conditions of the possible expression of his rationality. As a person’s true needs are those which must be met if he is to function (or continue to function) as a rational, end-setting agent, respecting the humanity of others involves acknowledging the duty of mutual aid: one must be prepared to support the conditions of the rationality of others (their capacity to set and act for ends) when they are unable to do so without help. The duty to develop (not neglect) one’s talents and the duty of mutual aid are thus duties of respect for persons. The ground of the duty of mutual aid then reveals its moral point. The good it looks to is the preservation and support of persons in their activities as rational agents. (Herman 597)

The manifestation of our rationality is our ability to set ends, and part of respecting the humanity in one’s person is the development of the ability to set ends. The true needs of an agent are those that she requires in order to function as a rational being and set ends for herself. Herman grounds the duty of mutual aid in the importance of the rational activity of moral agents. Mutual



aid is required of us because each of us as a rational being will require help to ensure the preservation of our rational natures and our end-setting ability. We must ensure that the true needs of others are met, because we must respect the rationality in their persons. The duty of beneficence is grounded on rational nature, not human nature.

But still, what's in it for the angels? What can motivate the angels to act with beneficence towards the humans, even though both species know that the humans will never have the opportunity to return the favor to the angels?<sup>14</sup> In the *Groundwork*, Kant seems to ground the duty of beneficence and mutual aid in reciprocity of that duty, even if that reciprocity is merely potential.<sup>15</sup> But the angels will never need the assistance of the humans, ever. The duty of beneficence cannot possibly be reciprocated in this case. The angels are being used. It seems that the Formula of Humanity as an End in Itself (FH) cannot require the angels to assist the humans because in such a setting, the angels are mere means to the human's well-being. The humans can never treat them as ends in themselves in this relationship because it is not reciprocal. The angels become the slaves of the humans. So, what's in it for the angels?

The angels and the humans are, without doubt, two empirically distinct species. Let us say that the angels' lack of physical needs is the only empirical difference between the two groups. This still is significant enough to ground dividing their natures into two distinct empirical ones and thus dividing the groups into two species. Nevertheless, in moral discourse, the two empirically distinct species must be seen as the same species. In the moral sphere, the empirical difference between the two species is a moot point, it has no claim on the moral status of either group. What matters when discussing moral obligation is the moral standing of both groups. Angels and humans are rational beings, capable of setting ends for themselves and

---

<sup>14</sup> Again, I must thank Husain Sarkar for raising this further line of questioning against my analysis of and answer to Herman's objection.

<sup>15</sup> See pages 40-3 for a full explication of this example.

acting morally, motivated by their respect for the CI. As they are moral beings, the two distinct empirical species are actually one moral species. As they are the same moral species, rational moral agents, they have the same duties toward one another derived from the same moral law, the CI.

This can be seen as, at least in part, a motivating force for the angels, dictating, essentially, what's in it for them to help the humans and assist in the maintenance of their physical needs. The angels must realize that the humans are part of the same moral species as themselves, the group of rational creatures capable of moral action. In their dealings with such creatures, the angels must disregard all other considerations if they are to act morally towards them. The only relevant feature of the humans' nature in this case is that they too are moral agents. The angels must act to further the humans' standing as moral agents, just as the humans must do the same towards the angels. If the angels do not perform their moral duties towards other rational moral agents, then they could be accused of what is only negligibly different from racism. Such racism takes the form of emphasizing certain contingent, physical characteristics of the other species that are irrelevant in treating them as moral agents. In this case, it is the fact that humans have physical needs. That the angels lack this physical need is only contingently true of their empirical natures, and they must see through this to the heart of the matter, that the two species are morally one, in order to understand the binding obligation of their duty of beneficence towards humans, even though humans will never have the opportunity to repay their kindness. Every moral being is bound to further the moral standing of every other moral being, regardless of their empirical natures.

Herman's thought allows for a way out of the relativist trap that Paton's view entails. By not grounding the duty of beneficence on human nature alone, we can ensure that not just

humans would be bound to such a duty. In the angel/human community, both humans and angels would be bound to help other humans, because their true needs must be met in order to preserve their rational natures, their abilities to set ends. Since the angels are rational agents, they too are bound to assist humans, because the duty was derived from the concept of rational nature as such. Furthermore, there is no difference between the angels and the humans with regard to moral standing. Both are the same moral species, as they are both rational agents, capable of moral action. Since the two empirically distinct species are the same morally, we can see why the angels should be motivated to help the humans, though the humans can never return the favor. The trap of relativism in the imaginary community is avoided: all rational beings are bound to the same duty.

Herman's exegesis here is in line with Kant's thinking in other texts. Indeed, I think that using an argument similar to the one that Herman articulates here and amending Paton's teleological contradiction interpretation, one can produce what is, hopefully, a convincing and novel contradiction interpretation that not only correlates with the four examples of FLN in the *Groundwork*, but also falls in line with other Kantian texts.

### **Towards a New Contradiction Interpretation**

Each of the three major contradiction interpretations has its own set of weaknesses. The logical contradiction interpretation can offer no truly satisfactory theory of contradiction in the will, and also has problems explaining the suicide example. Likewise, the practical contradiction interpretation encounters obstacles in analyzing the first example, but it is able to provide a theory of contradiction in will. Paton's teleological contradiction interpretation can successfully address the four examples, but it does so in such a way, as demonstrated by Herman's objection, that leads to relativism, a decidedly un-Kantian notion.

Since Paton's interpretation can address each of the four examples, it seems the most plausible view. Where Paton's theory fails is its appeal to a harmony of human ends. For him, a maxim contains a contradiction in conception if it undermines an overall harmony of human ends, whereas a contradiction in will would show that a maxim would fail to foster such a harmony. The argument against Herman's objection shows how this reliance on an overall harmony of human ends, with its explicit appeal to human nature, disintegrates into relativism.

If Paton's theory is to be refined, the reliance on human ends must be uprooted and something more in line with Kant's own views must be sowed in its place. I think that if the Kantian notion of teleology, as expounded in the third *Critique*, is examined, we can find a sound and strong way to fix Paton's theory and avoid the descent into relativism.

## I. Teleology and the *Critique of the Power of Judgment*

Kant explicitly lays out his teleology in the Second Part of *The Critique of the Power of Judgment*, the third and last of his major Critiques. Teleology, in general, concerns the perception of the world as though it were laid out in an organized fashion; everything in nature is purposive and constituted towards a certain end. As Kant says in the *Groundwork*, “Teleology considers nature as a kingdom of ends,” and here “the kingdom of ends is a theoretical idea for explaining what exists” (G 4:436n). This is a traditional teleological conception of nature. Kant maintains that the ultimate end of nature is the human being, but the human being is conceived of as a natural end as well, and thus this title of ultimate end belongs to the phenomenal realm. In the phenomenal realm, the human being is the natural end to which all other ends are subjugated, thus making him the ultimate end of nature. Kant claims:

[W]e have shown that we have sufficient cause to judge the human being not merely, like any organized being, as a natural end, but also as the **ultimate end** of nature here on earth, in relation to which all other natural things constitute a system of ends in accordance with fundamental principles of reason, not, to be sure, for the determining power of judgment, yet for the reflecting power of judgment.<sup>16</sup> Now if that which is to be promoted as an end through the human being’s connection to nature is to be found within the human being himself, then it must be either the kind of end that can be satisfied by the beneficence of nature itself, or it is the aptitude and skill for all sorts of ends for which he can use nature (external and internal). The first end of nature would be the **happiness**, the second the **culture** of the human being. (KU 5:429-30)

---

<sup>16</sup> The distinction between determining power of judgment and reflecting power of judgment is central to the arguments of the *Critique of the Power of Judgment*. In his first exposition of the terms, Kant explains:

The power of judgment in general is the faculty for thinking of the particular as contained under the universal. If the universal (the rule, the principle, the law) is given, then the power of judgment, which subsumes the particular under it (even when, as a transcendental power of judgment, it provides the conditions *a priori* in accordance with which alone anything can be subsumed under that universal), is **determining**. If, however, only the particular is given, for which the universal is to be found, then the power of judgment is merely **reflecting**. (KU 5:180)

So, determining judgment fits a particular instance or occurrence of something into the framework provided by a corresponding universal, in which case the universal is already provided for the power of judgment. Reflecting judgment takes a particular and attempts to generalize it into a universal. In the case illustrated above, reflecting judgment is used, because we generalize a teleology from the particulars of the natural world as it appears.

So the end that is promoted in human beings as parts of nature can be either happiness or culture; happiness being the kindness shown the human being by nature, and culture being efficacy for manipulating nature to bring into line with the human's own ends.

The happiness of the human being cannot be the ultimate end of nature, because the concept of happiness which humans possess is different for every individual and changes almost constantly for each individual (KU 5:430). "[I]t is impossible for the most insightful and at the same time most powerful but still finite being to frame for himself a determinate concept of what he really wills here," when he wills happiness (G 4:418). After explaining why happiness fails as the ultimate end of nature, Kant says:

As the sole being on earth who has reason, and thus a capacity to set voluntary ends for himself, he is certainly the titular lord of nature, and, if nature is regarded as a teleological system, then it is his vocation to be the ultimate end of nature; but always only conditionally, that is, subject to the condition that he has the understanding and the will to give to nature and to himself a relation to an end that can be sufficient for itself independently of nature, which can thus be a final end, which, however, must not be sought in nature at all. (KU 5:431)

Man is the ultimate end of nature only in that he can set up a relationship between himself and nature through an end that does not exist in nature and is a final end. Kant continues:

In order, however, to discover where in the human being we are at least to posit that ultimate end of nature, we must seek out that which nature is capable of doing in order to prepare him for what he must himself do in order to be a final end, and separate this from all those ends the possibility of which depends upon conditions which can be expected only from nature. Of the latter sort is earthly happiness. (KU 5:431)

To discover the location of the ultimate end of nature within the human being, we must first discover those things that prepare a human for her place as a final end of nature. Happiness cannot be the source of this, because this end is possible only through nature, not outside of it. The argument here has a corollary in the *Groundwork*, in Kant's teleological argument against happiness being the grounding of a moral system:

[I]n a being that has reason and a will, if the proper end of nature were its *preservation*, its *welfare*, in a word its *happiness*, then nature would have hit upon a very bad arrangement in selecting the reason of the creature to carry out this purpose. For all the actions that the creature has to perform for this purpose, and the whole rule of its conduct, would be marked out for it far more accurately by instinct, and that end would have thereby been attained much more surely than it ever can be by reason. (G 4:395)

There is, however, something in nature that qualifies man and helps him to become “the titular lord of nature”:

Thus among all his ends in nature there remains only the formal, subjective condition, namely the aptitude for setting himself ends at all and (independent from nature in his determination of ends) using nature as a means appropriate to the maxims of his free ends in general, as that which nature can accomplish with a view to the final end that lies outside of it and which can therefore be regarded as its ultimate end. The production of the aptitude of a rational being for any ends in general (thus those of his freedom) is **culture**. Thus only culture can be the ultimate end that one has cause to ascribe to nature in regard to the human species (not its own earthly happiness or even merely being the foremost instrument for establishing order and consensus in irrational nature outside him). (KU 5:431)

The ultimate end of nature is that which facilitates the final end of nature, as this passage shows. The only end of nature that could qualify as such an ultimate end is a human being’s ability to set maxims for herself and the efficacy with which she is able to set such maxims. The ability to set ends must be free, and the ends must be set of humanity’s free nature, not dictated to humans by nature. Efficacy in accomplishing one’s ends is the ability to use and manipulate nature to bring it in line with the ends such a free agent sets. This ultimate end is established within nature with a view toward the final end which lies outside of it. Culture is the efficacy and ability of a human to set ends for herself. This is the ultimate end of nature, found in the human being, and neither the human being’s happiness nor its ability to establish order in nature qualify as the ultimate end of nature.

Kant continues, “A **final end** is that end which needs no other as the condition of its possibility” (KU 5:434). If we assume that the teleologically posited ends of the world are real

and assume that there is an “**intentionally acting** cause” in the world, “then we must also raise the question of the objective ground that could have determined this productive understanding to an effect of this sort, which is then the final end for which such things exist” (KU 5:434-5). Consequently, if the teleology we assign to the world is assumed to be real and an “intentionally acting cause” is also assumed, then the final end of nature is the real grounds by which our teleological conception of the world gains its credence. This final end of nature cannot be so constituted that it is also an end to something else, as the ultimate end of nature is the condition of the final end of nature. All other ends are subjugated to it, while it is subjugated to none.

Kant begins his analysis of the final end of nature:

Now we have in the world only a single sort of beings whose causality is teleological, i.e., aimed at ends and yet at the same time so constituted that the law in accordance with which they have to determine ends is represented by themselves as unconditioned and independent of natural conditions but yet as necessary in itself. The being of this sort is the human being, though considered as noumenon: the only natural being in which we can nevertheless cognize, on the basis of its own constitution, a supersensible faculty (**freedom**) and even the law of the causality together with the object that it can set of itself as the highest end (the highest good in the world). (KU 5:435)

Human beings possess a teleological causality, a causality intent upon ends, and they feel that this causality is necessarily bound to a restriction, the moral law, which is perceived as a necessity and independent of natural conditions. This process of the self-imposition of the moral law occurs in the human being considered only as a noumenon. Only the human being considered as noumenon can be considered as a free agent. The freedom of a human lies outside of nature and is, thus, not a part of nature.

This idea of freedom being outside of the world of appearances—transcendental freedom—is argued for in the *Critique of Pure Reason*. Because it lies in the realm of the noumena, practical reason is not bound by time:



It, reason, is present to all the actions of human beings in all conditions of time, and is one and the same, but it is not itself in time, and never enters into any new state in which it previously was not; in regard to a new state, reason is **determining** but not **determinable**. Therefore one cannot ask: Why has reason not determined **itself** otherwise: But only: Why has it not determined **appearances** otherwise through its causality? (KrV A556/B584)

The solution to the Third Antinomy of Pure Reason establishes the possibility of this freedom (KrV A444/B472—A451/B479), but it is not until later that Kant reconciles the notion of freedom with the notion of natural causality and shows that they are not mutually exclusive:

Now even if one believes [a malicious lie] to be determined by [empirical] causes, one nonetheless blames the agent, and not on account of his unhappy natural temper, not on account of the circumstances influencing him, not even on account of the life he has led previously; for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the previous state, as though with that act the agent had started a series of consequences entirely from himself. (KrV A555/B583)

An action may be causally determined in the phenomena by various things, such as an agent's upbringing, but in the noumena, he is completely free, and we treat him as such.

After establishing the centrality of the notion of the human being as a noumenon, a rational and free agent, Kant progresses to his formulation of the final end of nature:

Now of the human being (and thus of every rational being in the world), as a moral being, it cannot be further asked why (*quem in finem*) it exists. His existence contains the highest end itself, to which, as far as he is capable, he can subject the whole of nature, or against which at least he need not hold himself to be subjected by any influence from nature.—Now if things in the world, as dependent beings as far as their existence is concerned, need a supreme cause acting in accordance with ends, then the human being is the final end of creation; for without him the chain of ends subordinated to one another would not be completely grounded; and only in the human being, although in him only as a subject of morality, is unconditioned legislation with regard to ends to be found, which therefore makes him alone capable of being a final end, to which the whole of nature is teleologically subordinated. (KU 5:435-6)

A human's existence contains the highest end of nature, to which the rest of nature can be subjugated, and nature may hold no influence over the noumenal part of the human, her freedom.

The human being is the final end of nature, because without her, the teleological system of nature would find no final end to rest upon, and thus all ends would be subordinate to other ends, and the system would have no grounding. Nature is “teleologically subordinated” to human kind as moral agents, because human contain the ability to legislate over all of nature, without nature legislating back in turn. The final end of nature is the freedom of human beings as rational beings.

Thus, under Kant’s teleological guidelines, the human being as a rational being, as an agent of morality, is the final end of nature. The human being constitutes the final end of nature only in that she is also a rational being, and a rational being is a moral agent. *Moral agency* is what determines the human being’s ability to be a *final end*. *It seems that this is what needs to be refined in Paton’s argument.* Paton appeals to the harmony of human ends in his teleological contradiction interpretation. Paton’s moral teleology puts human ends first as the final, ultimate end of nature. It is not, however, *human* ends that are the final ends of nature, it is *rational* ends. It is the rational ability of a human being to make a moral choice that is the final end of nature. Nature is oriented toward the freedom of humans as noumena. With this considered, a key alteration to Paton’s work can be made, and a, hopefully, more plausible contradiction interpretation produced.

## II. A New Teleological Contradiction Interpretation

Taking Barbara Herman’s objection to Paton’s explication of the fourth example, the efficacy of revising Paton’s teleological conception to bring it more in line with that which Kant demonstrates in the third Critique can be shown. Applying the problem of the community containing both angels and humans shows that Paton’s view leads to relativism because it relies

on human nature as a teleological end of nature. The harmony of human purposes seems to suggest this, because it entails that a rational agent necessarily wills the harmony of all of her purposes.<sup>17</sup> If this dependence can be removed, then Kant's fourth example could not be seen as leading to relativism.

I believe that the teleology outlined above can be shown to solve this problem and is explicitly connected to certain passages of the *Groundwork* which refute Herman's argument. The only common characteristic between angels and humans that could demonstrate that both have a duty of beneficence to other humans is their rational natures. Certain passages from the *Groundwork* quoted above, I feel, definitively demonstrate that the moral law and specific instances of duty derived from it must be based entirely on rational nature and not merely on human nature or other contingent, empirical natures possessed by other rational beings. Paton's teleological contradiction interpretation fails because it is based on a system where only human ends are considered. Kant's system focuses on the rational freedom of humans as the final end of nature. This rational freedom is not merely human, but must be in place for all rational beings as such, because it is derived from the concept of the rational being as such. If the ultimate end of nature is rational freedom, then, revising Paton's view, the furtherance of rational freedom is the end of teleology in the moral sphere. Angels, as rational beings, have a duty of beneficence to humans, because humans are free, rational beings as well. The common duty of beneficence is created from the conception of their common rational natures.

Let me explain further. Kant explicitly appeals to teleological considerations in the four examples of FLN in the *Groundwork*. It seems that Paton is dead right in picking up on this and attempting to posit an interpretation of contradiction in conception and contradiction in will that

---

<sup>17</sup> There is more on this notion of what a rational agent wills as a function of different contradiction interpretations to come. See Section IV of this chapter.

utilizes this notion of teleology. I believe he goes wrong in what he thinks makes these teleological contradictions manifest. Contradictions in conception undermine an overall harmony of human purposes, whereas contradictions in will simply fail to foster such a harmony.

I would revise this theory, instead positing that while the contradictions are essentially teleological, they appeal to the idea of the final end of nature as described in the *Critique of the Power of Judgment*. Contradictions in conception and contradictions in will arise because contradictory maxims contain or entail some element that works against the final end of nature, the furtherance of rational freedom. Thus, taking Paton's view and altering it: maxims that produce contradictions in conception undermine the final end of nature, the freedom of rational beings, whereas maxims leading to contradictions in the will fail to foster that freedom. Another analysis of the four examples can help explicate this interpretation.

### III. The Four Examples Revisited

In the first example of the suicide, the man makes his maxim: "from self-love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness" (G 4:422). Kant's statement of the immorality of the proposed action relies upon the conflicting teleological purposes of self-love and suicide. The purpose of self-love, Kant claims is the furtherance of life, whereas the purpose of suicide is the end of life. I conjecture that this proposed simple teleological contradiction is enough to show a contradiction in conception.<sup>18</sup> If all of nature is teleologically oriented to the final end of rational freedom, then if a contradiction in teleological purposes emerges, then the freedom of at least one rational agent is being undermined. If all teleological purposes are harmoniously oriented toward the final end

---

<sup>18</sup> There is a fourth previously written contradiction interpretation that I have not discussed here, and it is the simple teleological contradiction interpretation. The view is mentioned, in little more than passing, by Korsgaard, but she cites no bibliographic references for the theory. See Korsgaard 87-8.

of rational freedom, then the existence of a contradiction between two teleological purposes shows that a free agent has created an end which is not harmoniously oriented toward rational freedom. The same can be said for the second example of the lying promise.

In the lying promise case, a man is short of money and needs it fast, so he contemplates making a false promise to return borrowed money, which he knows he can never repay. His maxim is: “when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen” (G 4:422). Kant seems to appeal, in this example, to the inefficacy of the action in question in the world of the universalized maxim. I propose that this explanation results from the contradictory teleological purposes present in the maxim. The purpose of promising is to ensure truthful communication, whereas the teleological purpose of lying is to deceive. This simple teleological contradiction results because the agent is undermining the freedom of the person from whom he wishes to borrow money. The freedom of a rational agent is undermined if someone is intentionally keeping information from them that they need to make a fully free and rational choice. The withholding of information makes the agent uninformed, and thus removes, insofar as it lies in the informant, the lender’s freedom. The simple contradiction emerges, because the rational freedom of at least one rational being is undermined.

Thus, a clear theory of the contradiction in conception can be illuminated. A contradiction in conception emerges when the rational freedom of some rational being, be it the agent or some other, is undermined in the maxim provided by the agent in question. This undermining of freedom is manifest in a simple teleological contradiction within the maxim; the maxim cannot even be thought of as a universal law of nature, because nature is teleological, and a teleological system of nature is a harmonious system of ends. This harmonious system of ends

must be oriented towards its final end, and a contradiction at any point implies a failure of at least one end to be in line with nature's final end.

In the case of the example of rusting talents, an imperfect duty to oneself,

[A man] finds himself in comfortable circumstances and prefers to give himself up to pleasure than to trouble himself with enlarging and improving his fortunate natural predispositions. But he still asks himself whether his maxim of neglecting his natural gifts, besides being consistent with his propensity to amusement, is also consistent with what one calls duty. (G 4:423)

Kant finds no contradiction in conception in the case, so that a nature could subsist with the man's maxim, but he could not will a nature as such, because

For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes. (G 4:423)

Kant's rejection of the system of nature legislated by this man's maxim relies upon the maxim undermining the possible purposes that could be fulfilled by the man's talents. The implication is that the man's neglect of his talents is immoral, because it limits the possible ends that he could set for himself and fulfill. The nature legislated by his universalized maxim cannot be willed because it fails to foster his rational freedom as the teleological final end of nature. The limiting of his ability to set and fulfill ends for himself is a limiting of his freedom, and it is on this ground that Kant dismisses the maxim.

The fourth example of beneficence has already been partially illustrated in response to Herman's objection, but a clearer view can be given from this new position. Kant's rejection of the man's maxim of neither giving nor receiving help goes:

For, a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (G 4:423)

The contradiction arises here because the man is denying to himself something that he may one day need. This reasoning is moral because we may, on occasion, need the assistance of others in realizing our ends. Thus, there are instances where our freedom cannot be fully realized without the help of others. The man fails to foster freedom in his own person and in others, by denying to himself and others resources that are, on occasion, necessary for the realization of ends. He fails to foster the teleological system of nature, because he fails to promote the final end of nature—his rational freedom.

Thus, a developed interpretation of contradiction in will can be expounded. Unlike the contradiction in conception, the universalized maxim does not contain an outright contradiction, instead it can be universalized to a law of nature, but a rational agent could not will it as such. In such a system of nature, the ends prescribed by the maxim would fail to foster the final end of nature, the freedom of rational beings. The ends of the maxim do not necessarily contradict any in the teleological system, though they are out of place in that they do not contribute to the furtherance of the entire system. Immoral maxims in cases of imperfect duties are those that do not undermine the rational freedom of some rational being, but simply fail to foster it. When a maxim undermines the freedom of a rational being, one of the teleological purposes in the maxim directly contradicts another, and this entails that the systematic harmony oriented toward the final end of nature is thrown off by what the agent wills. Thus, the undermining of the system's orientation toward its final end undermines the agent's or another's rational freedom. A maxim that fails to foster this includes some purpose that does not directly contradict another purpose oriented toward the final end of nature, but merely does not mesh with the system as a whole. The maxim is not contradictory, so much as it just wants to be outside the teleological system directed toward the final end of rational freedom.

One extremely interesting implication of this interpretation of the contradiction in conception and the contradiction in will is that FLN can be seen as prescribing at least one positive duty, and this positive duty represents the core of Kant's moral philosophy: *For any action I consider to be of moral weight, I will commit it with the furtherance of the rational freedom in myself and others as its end.* FUL and FLN are usually seen as negative tests for duty—i.e. they clearly delineate what maxims are immoral, but do not differentiate between moral maxims and amoral maxims. This new teleological contradiction interpretation opens up a way to analyze maxims, under FLN, to determine whether they are amoral or moral. Immoral maxims fail to foster or undermine the freedom of a rational being, be it that of the agent or someone else. Moral and amoral maxims neither undermine nor fail to foster the final end of nature.

It seems that two dichotomies emerge from neither undermining nor failing to foster the final end of nature. One is simply taking neither of the immoral courses, while the other is furthering and promoting the final end of nature and being in line with the teleological system oriented thus. Amoral actions are those that follow the first course and simply avoid undermining or failing to foster the rational freedom in some agent. Such a maxim, while not immoral, carries with it no real moral weight, because it has no genuinely moral content. So it seems that moral action must be included in the second branch of this dichotomy. It appears, *prima facie*, to be the case that moral action is the opposite of immoral action, so that for what makes an action immoral, its opposite must be what makes an action moral. If this is indeed the case, then a moral maxim is one that adopts the final end of nature as its end and promotes and furthers this final end. A positive maxim of duty can thus be derived from FLN: For any action



(of moral weight) that I consider, I will commit it in order to further the rational freedom in myself and/or others if applicable.

Amoral action can be seen as neither furthering nor undermining the final end of nature. Such an action could be moral, except that its maxim contains no moral content. It is in conformity with duty, but not committed *from* duty. Kant's example of the shopkeeper from Section I is helpful here:

For example, it certainly conforms with duty that a shopkeeper not overcharge an inexperienced customer, and where there is a good deal of trade a prudent merchant does not overcharge but keeps a fixed general price for everyone, so that a child can buy from him as well as everyone else. People are thus served *honestly*; but this is not nearly enough for us to believe that the merchant acted in this way from duty and basic principles of honesty; his advantage required it; it cannot be assumed here that he had, besides, an immediate inclination toward his customers, so as from love, as it were, to give no one preference over another in the matter of price. Thus the action was done neither from duty nor from immediate inclination but merely for purposes of self-interest. (G 4:397)

In this case the shopkeeper acted either from immediate inclination or from self-interest, and his action was in conformity with what morality dictates. His action was not moral because the maxim upon which he acted was not moral. So, amoral cases emerge when the maxim of the agent neither hinders nor furthers the rational freedom of another, though his action would have moral character if his maxim did too. The ability to derive a prescriptive duty plays into the notion of what a rational agent wills.

#### **IV. The Differences in the Analyzed Contradiction Interpretations**

Korsgaard analyzes each contradiction interpretation as dictating a different notion of what a rational agent wills. She discusses these briefly:

Each interpretation must presuppose some notion of rationality in determining whether a rational being can will the universalization of a maxim at the same time as that maxim without contradiction. The Logical Contradiction view works with a notion of contradiction indistinguishable from that of theoretical rationality and this is a great advantage. But this advantage is lost when we turn to

contradictions in the will, which then require another interpretation. The Teleological Contradiction view works with a rather rich notion of rationality as aiming at a harmony of purposes. I think on Kant's view pure reason does aim at a harmony of purposes, but that only morality tells us how that is to be achieved. We cannot reason morally from that idea. The Practical Contradiction view uses a specifically practical notion of rationality and of a contradiction which springs from the notion of the will as a causality. This is not a morality-laden notion of rationality, for on Kant's view this notion is needed to explain *instrumental* rationality. (Korsgaard 101-102)

Each contradiction interpretation seems to entail a different notion of a rational will. By positing what type of contradiction is to be avoided in questions of moral maxims, each interpretation posits the basis of the contradiction as being something that a rational will necessarily wills. For the logical contradiction interpretation, its theory of contradiction is such that it relies on the exact same style of contradiction that is applied in theoretical reasoning—i.e. it utilizes a sense of contradiction that is the same as that defined by classical logic. In this regard, the logical contradiction interpretation posits an uncontroversial thesis, and practical rationality follows theoretical rationality.

On Paton's teleological view, "rationality aims" at a harmony of human purposes, such that a rational agent wills the harmonious co-existence of all his ends. The rational agent wills none of his ends to contradict any other, so that practical rationality is the complete harmony of human ends within a teleological system. There is a difference between theoretical rationality and practical rationality on this view. While both aim at non-contradiction, the styles of contradiction they espouse are different. This alone is not damning to Paton's view, because Kant himself seems to treat practical rationality and theoretical rationality differently. For example, Kant offers proofs for both rational freedom, using practical rationality, and transcendental freedom, using theoretical rationality.

The practical contradiction interpretation posits a view that entails a decidedly practical sense of contradiction, and what it entails about what a rational agent wills is extremely plausible and tenable. In the case of Korsgaard's view, a rational agent must will the efficacy of her actions. A practically rational agent will commit only those actions that are efficacious and can be brought about in the first place. It seems uncontroversial that a rational agent would will the efficacy of her actions, for otherwise, she would have no need to will any actions in the first place. One virtue of this view, is that it incorporates what Kant says in the *Groundwork* about hypothetical imperatives being analytic. Working with a decidedly practical view of reason, Kant maintains that hypothetical imperatives are analytic, because the means is contained in the end, and the addition of means to such an imperative adds nothing to our conception of the ends. If analytic, on the practically rational definition, means that the means are contained in an end, then one can see that Korsgaard's view coincides nicely. Because the efficacy of an end is determined by its means, an analytic proposition for the practical rationality of effectiveness is a hypothetical imperative in which the means are already contained in the end. Like Paton's view, Korsgaard's gives a definition of practical rationality which is different from Kant's theoretical rationality. Both, again, are aimed at non-contradiction.

Each of these contradiction interpretations offers a defensible position on what a rational agent wills. The question, however, is not which is independently the most defensible, but what is most in line with Kant's own writing. Again, each has its own merit. The logical view entails no difference in terms between theoretical and practical reasoning. Paton's teleological view is plausible and generally makes sense as something a rational agent would will. Korsgaard's view allows for a specifically practical reading of contradiction. I claim, however, that none of these can adequately explain, at least a majority of, Kant's examples, and that, furthermore, each

departs from the Kantian texts at some point. I would argue that the key text to understanding Kant's own theory of contradiction in conception and contradiction in will is the *Critique of the Power of Judgment*, and, once Kant's own teleology is closely examined, its extreme importance—nay, its indispensability—for the moral sphere becomes apparent. In this regard, only a refined version of Paton's interpretation can be seen as both adequately answering the problems of consistency posed by the four example of FLN and cohering with some of the key Kantian texts.

These different interpretations of rational willing are just one difference in the proposed contradiction interpretations. Another difference is the type and structure of the maxim required by each theory. Since the contradiction in conception requires that a maxim, in universalized form, demonstrate a contradiction only by virtue of its being universalized, it seems that the proper type and form of maxim is most essential here. The contradiction in will may appeal to elements outside of the maxim to find the contradiction, as Kant clearly does in the last two examples of FLN.

For the logical contradiction interpretation, it seems crucial to include purposive content, as a contradiction resulting in the impossibility of the universal law must be uncovered. This search may require as much information included in the maxim as is possible, and Wood demonstrates this when he formulates the universal law resulting from the maxim of the suicide case: "It is a universal law of nature that from self-love, all make it their principle to shorten life when by a longer duration it threatens more ill than agreeableness" (Wood 85). Because the logical contradiction interpretation is so pressed to find a rigorous style of contradiction, it seems that a complete maxim of ends is necessary in order to find enough content upon which to base a contradiction.

For the practical contradiction interpretation, it seems that only an incomplete maxim of ends, a maxim without purposive content, is necessary. Since proponents of this view seek out a contradiction resulting from the undermined efficacy of the proposed action with regard to its intended goal, it seems that only the action and its goal are to be included in this type of maxim. The efficacy of an action when universalized in a given maxim depends on what the intended outcome of that action is to be. Therefore, the maxim should only require the action type and the agent's goal in that action for a contradiction in the practical sense to emerge.

The teleological view of Paton, like the logical contradiction interpretation, seems to require a purposive element in maxims for a contradiction to be seen as arising. The undermining of human purposes occurs when a proposed action's teleological purpose undercuts the purposive content of the maxim.

In this regard, the revised version of Paton's view that I propose requires the same thing. It is self-love that teleologically contradicts suicide, not the goal of ending one's suffering. This simple teleological contradiction indicates that one of his teleological ends is contradictory to his rational freedom. The contradiction arises because the agent's maxim undermines his own freedom as a rational agent. If teleological purposes are being considered, then having purposive content is imperative for the maxim.

For contradiction in will, it seems that only a maxim of action is required for both Korsgaard's view and my own, as each relies on some element not contained in the agent's maxim for seeking out a contradiction. It seems that Paton's view would also entail this, because in his explication of the examples of rusting talents and beneficence he introduces concepts not inherent in the agent's intent to find a contradiction in the will. Each makes some judgment about what a rational agent must necessarily will, and each theory introduces this as something

that is contradicted in contradictions in will. I feel that Kant's own exegesis of the four examples justifies this inclusion of outside material, because he does just this in both examples of imperfect duties. I do not address the logical contradiction interpretation's needs with regard to maxims for imperfect duties, because it has such a hard time accommodating the contradiction in will in the first place.

## V. Some "Casuistical Questions"

In *The Metaphysics of Morals* Kant poses serious "casuistical questions" which amount to a series of serious moral riddles. I will attempt to answer some of these riddles in an attempt to show the efficacy of my revision of Paton's view in handling morally sticky situations. I will address three casuistical questions from *The Metaphysics of Morals*, one from the discussion of suicide, a second from Kant's arguments concerning lust, and the third from the case against lying.

Among the casuistical questions Kant poses concerning suicide is the following:

Can a great king who died recently be charged with a criminal intention for carrying a fast acting poison with him, presumably so that if he were captured when he led his troops into battle he could not be forced to agree to conditions of ransom harmful to his state?—For one can ascribe this purpose to him without having to presume that mere pride lay behind it. (MS 6:423)

The question is one of noble suicide—the morality of martyrdom. In this case, a king (Frederick the Great) would take his own life before allowing a forced treaty to weigh upon his homeland and harm his people. There are two immoral acts that Kant seems to take very decisive and firm stances on in *The Metaphysics of Morals*, and suicide is one of his hard line, uncompromising points. Early in this section he says:

To annihilate the subject of morality in one's own person is to root out the existence of morality itself from the world, as far as one can, even though

morality is an end in itself. Consequently, disposing of oneself as a mere means to some discretionary end is debasing humanity in one's person (*homo noumenon*), to which man (*homo phaenomenon*) was nevertheless entrusted for preservation. (MS 6:423)

This passage is in line with the teleology laid out in the third *Critique*, and here Kant argues that humanity as bound by natural law in the world of appearances is meant to protect and further humanity as noumena, as free moral agents. Suicide is wrong, because it destroys rational freedom—the source of morality—in the agent.

Our new contradiction interpretation is easily set to handle this moral puzzle as Kant explains it. If the king had committed suicide after being taken hostage, even with the maxim of protecting his homeland, his action would still have been immoral, because he would have undermined the status of his rational freedom as the final end of nature. To start with, he cannot be “forced” to agree to a treaty degrading to his homeland. As *homo noumenon*, he is not bound to the laws of nature, for he is outside them, and he cannot be coerced into anything as a result of what occurs in the phenomenal realm. Thus, suicide from fear of coercion is almost a type of bad faith—it demonstrates a lack of respect and faith in one's powers of freedom. Suicide is a complete destruction of such freedom.

The example from lust concerns the proper orientation of the union of the sexes in marriage. The casuistry runs:

Nature's end in the cohabitation of the sexes is procreation, that is, the preservation of the species. Hence one may not, at least, act contrary to that end. But is it permitted to engage in this practice (even within marriage) *without taking this end into consideration?* (MS 6:425)

The concern here is with the morality of sexual behavior conducted even within marriage the aim of which is not reproduction. Kant is asking whether any sexual conduct except for the furtherance of the species is morally permissible. Not defiling oneself by lust is considered by

Kant to be a perfect duty to oneself, so that a contradiction in conception must be found in the maxim if the action is immoral.

In this case, the agent proposes to engage in sexual behavior, but takes no heed of the end of procreation being the furtherance of the species. A contradiction in conception would indeed arise in the maxim of such an agent. There would be a simple teleological contradiction between his purpose and the purpose of procreation. The agent is undermining rational freedom in this case. The question, it seems, is whose rational freedom is being undermined? According to Kant, it is the agent's own freedom that suffers from carnal defilement because a person gives himself over to his animal instincts in indulging in lustful actions. He forgoes his rational freedom in order to entertain the base interests of his instincts.

The casuistical questions that Kant associates with lying in *The Metaphysics of Morals* are all, in my opinion, some variation on the idea of the philanthropic lie, so that this is the problem that I will deal with here. The problem of the philanthropic lie is most famously stated in a case that could have arisen (and no doubt did arise) in Hitler's Germany. In this moral puzzle, the agent is hiding Jews, who will be sent to concentration camps if discovered, in her attic. A Nazi officer knocks on the woman's door, and asks her if she is hiding Jews in her attic. The woman may either protect the Jews she is hiding and lie to the officer, or she may forsake the Jews in her attic and tell the truth to the officer.

Lying is the second action that Kant took a very hard and definitive stance on in *The Metaphysics of Morals*. He says that "the greatest violation of a human being's duty to himself regarded merely as a moral being (the humanity in his own person) is the contrary of truthfulness, *lying*" (MS 6:429), and later he says:

By a lie a human being throws away and, as it were, annihilates his dignity as a human being. A human being who does not himself believe what he tells another



(even if the other is a merely ideal person) has even less worth than if her were a mere thing; for a thing, because it is something real and given, has the property of being serviceable so that another can put it to some use. But communication of one's thoughts to someone through words that yet (intentionally) contain the contrary of what the speaker thinks on the subject is an end that is directly opposed to the natural purposiveness of the speaker's capacity to communicate his thoughts, and is thus a renunciation by the speaker of his personality, and such a speaker is a mere deceptive appearance of a human being, not a human being himself. (MS 6:429)

Kant seems pretty clear in his stance on lying. It is a deplorable act that destroys the humanity of the liar, and reduces him, in some ways, to less than a mere thing. Kant claims that mere things are useful in that they can be useful, whereas one who lies is of no use to another, and since he cannot be believed is of less worth than that mere thing. His appeal to teleology here is also apparent. Implicitly, the natural purpose of the power of communication is the truthful communication of one's thoughts. When one lies, he denies an essential element of his human nature, "renounces his personality," and is no longer a human being, because he has violated a purpose inherent in the human being.

With this in mind, one may see that Kant's ethics forces the woman to surrender the Jews hiding in her attic in order to tell the truth to the Nazi officer. Any other maxim would produce a contradiction in conception, because the act of lying is inherently contradictory according to Kant. One may immediately object that what is at stake here is a conflict of duties: the woman must lie in order to save the Jews in her attic because she is bound to a greater duty to save the Jews than to tell the truth to the Nazi officer. It seems to me, however, that there is no conflict of duties here. The woman must respect the rational freedom of the officer by telling him the truth—he may decide to walk away, the Jews in the attic may run, the officer may become distracted by someone vandalizing a neighbor's house. Regardless, the woman has fulfilled her

imperfect duty to the Jews by hiding and sheltering them, she is not turning them in because she wants to. She is turning them in because she is bound by the CI to do so.

But this is, nonetheless, a sticky case. To say that such an argument as above would, or even should, be uncontroversial, is to lie outright. It may be that the maxim of lying is always contradictory, but the act of lying may not be what lies at the heart of this case. It may be that the act of turning over the Jews takes precedence in the maxim over the lie. Christine Korsgaard deals extensively with this problem of the philanthropic lie in her book *Creating the Kingdom of Ends*. An exegesis and analysis of her discussion is worthwhile in attempting to understand this casuistical concern.

## 6

**The Supposed Right to Lie**

In Chapter Five of her *Creating the Kingdom of Ends*, Korsgaard deals with the casuistical question of the philanthropic lie at great length. Her discussion of this example moves in three waves. The first attempts to show that if FUL is used to judge the maxim of the action, then the philanthropic lie is permissible. Korsgaard's discussion of FUL is grounded in her interpretation of the contradiction in conception test—the Practical Contradiction Interpretation. “Thus the test question will be: could this action be the universal method of achieving this purpose?” (Korsgaard 135) It also discusses the problem in terms of its formulation in the *Metaphysics of Morals*—the servant who lies on order of his master:

If I say something untrue in more serious matters, having to do with what is mine or yours, must I answer for all the consequences it might have? For example, a householder has ordered his servant to say “not at home” if a certain human being asks for him. The servant does this and, as a result, the master slips away and commits a serious crime, which would otherwise have been prevented by the guard sent to arrest him. Who (in accordance with ethical principles) is guilty in this case? Surely the servant, too, who violated a duty to himself by his lie, the results of which his own conscience imputes to him. (MS 6:431)

In this example, a master orders his servant to tell a visitor he is not at home so that he may slip out and commit some dastardly crime without interference. As a result of this, the guard sent to apprehend the master leaves without detaining him, and the master gets away with his crime. Kant maintains that responsibility for the crime of the master lies as much in the servant because the servant is responsible for all the consequences of his lie. She also grounds her discussion on the example of the murderer at the door from Kant's essay “On a Supposed Right to Lie from Philanthropy” in which a friend knocks on your door and asks you to hide him because he is being chased by a murderer. After he has hid himself away in your home, the murderer then knocks on your door asking whether or not your friend is there. Do you lie to the murderer or

reveal your friend's location? Both options seem to have some serious moral consequences: one involves assisting the murderer in finding your friend and, thus, possibly killing him, while the other involves lying, which Kant seems to think is never a permissible act.

Secondly, Korsgaard hopes to show that the lie is not permitted under either the Formula of Humanity as an End in Itself or the Formula of the Kingdom of Ends. She then uses an idea of John Rawls to show that this does not demonstrate an inconsistency within the ethics; instead, it demonstrates the emphasis Kant's theory places on an ideal system of human relations. Finally, she hopes to show how an ethical system can effectively accommodate principles of such an ideal system of human relations.

### **I. The Philanthropic Lie and Universal Law**

Korsgaard analyzes this problematic example under FUL by first asking of lying in general if it is universalizable, if it could ever be the universal method of achieving any purpose (Korsgaard 135-6). Her argument advances by use of examples. The first of these runs:

For lies are usually efficacious in achieving their purposes because they deceive, but if they were universally practiced they would not deceive. We believe what is said to us in a given context because most of the time people in that context say what they really think or intend. In contexts in which people usually say false things—for example, when telling stories that are jokes—we are not deceived. If a story that is a joke and is false counts as a lie, we can say that a lie in this case in [sic] not wrong, because the universal practice of lying in the context of jokes does not interfere with the *purpose* of jokes, which is to amuse and does not depend on deception. But in most cases lying falls squarely into the category of the sort of action Kant considers wrong: actions whose efficacy depends upon the fact that most people do not engage in them, and which therefore can only be performed by someone who makes an exception of himself (G [4:]424). (Korsgaard 136)

Lying is immoral, under Korsgaard's reading because in most realms of discourse, we anticipate that other participants in our conversations are communicating truthfully. But there are

plausible, indeed factual, situations in which the discourse does not assume that one is communicating truthfully. Korsgaard latches on to a situation in which someone is telling a joke as a primary example of this. In such a situation, the participants, particularly the observers, do not expect to hear the truth, but instead anticipate a non-malicious falsehood to be told for the sake of humor. In this case, the purpose of jokes is not interrupted by the universal practice of lying, for if everyone told lies in this situation, the world could indeed go on functioning as normal and the efficacy of the humorous lie (with no malicious intent) is not undermined by the fact that it is a lie. Without explicitly saying so, Korsgaard implicitly maintains that such a lie is in fact permitted under Kant's FUL and is not immoral. She clearly says that most lies fall into the impermissible category, and that in most instances, people who are lying are attempting to make exceptions of themselves (Korsgaard 136). This, I think is a unique consequence of her Practical Contradiction Interpretation.

After attempting to prove that some lies are in fact permissible under FUL, Korsgaard begins her discussion of the murderer at the door example from "On a Supposed Right to Lie from Philanthropy":

A murderer who expects to conduct his business by asking questions must suppose that you do not know who he and what he has in mind. If these are the circumstances, and we try to ascertain whether there could be a universal practice of lying in these circumstance, the answer appears to be yes. The lie will be efficacious even if universally practiced. But the reason it will be efficacious is rather odd: it is because the murderer supposes that you do not know what circumstances you are in—that is, that you do not know you are addressing a murderer—and so does not conclude from the fact that people in those circumstances always lie that *you* will lie. (Korsgaard 136)

She analyzes the puzzle in terms of the murderer's hope at misrepresenting the events to you. The murderer is hoping that you do not anticipate that he is a murderer. Therefore, he is hoping that you are being deceived (in some loose sense) into thinking that he is not a murderer, and that

you will, because you do not suspect him of evil intent, tell him the truth. The lie, even if universalized is still efficacious, because the murderer, even in the world of the universalized maxim, will not expect you to lie. In the world of the universalized maxim, whenever one is harboring a friend and confronted by the person attempting to murder the friend, the person lies to the murderer to protect the friend. In the world of the universalized maxim, the lie is still efficacious in this situation, because the murderer is trying to come across as otherwise and thus does not expect the person harboring the friend to be deceptive, the lie will work and the murderer will pass on to the next house.

After analysis of these two examples, Korsgaard concludes:

These reflections might lead us to believe, then, that Kant was wrong in thinking that it is never all right to lie. It is permissible to lie to deceivers in order to counteract the intended results of their deceptions, for the maxim of lying to a deceiver is universalizable. The deceiver has, so to speak, placed himself in a morally unprotected position by his own deception. He has created a situation which universalization cannot reach. (Korsgaard 137)

The inferential leap that Korsgaard makes is to declare what amounts to a *quid pro quo* relationship between the deceiver and those he intends to deceive. The deceiver's hopes of deception make him someone to whom it is permissible to tell a lie. Lying is permissible here only in so far as it is necessary to counteract the deceiver's ends in deceiving you.

## **II. The Formulas of Humanity as an End in Itself and Kingdom of Ends**

Korsgaard maintains, however, that "when we apply the Formula of Humanity, however, the argument against lying that results applies to any lie whatever" (Korsgaard 137). She initially offers a brief explanation of what counts as treating humanity as an end in itself, and claims that:

“Humanity” is used by Kant specifically to refer to the capacity to determine ends through rational choice (G [4:]437; [MS] [4:]392). Imperfect duties arise from the obligation to make the exercise, preservation, and development of this capacity itself an end. The perfect duties—that is, the duties of justice, and, in the realm of ethics, the duties of respect—arise from the obligation to make each human being’s capacity for autonomous choice the condition of the value of every other end. (Korsgaard 137)

Humanity is interpreted by Korsgaard as being the human faculty for rational choice and the determination of ends through rational choice. This idea can be explicitly connected to the teleology of the Third Critique. Imperfect duties are related to the furtherance of the capacity for rational choice as an end, and perfect duties are manifest the practical necessity of determining the value of every other end in accordance with the power of rational choice in each human being. Generalizing from the lying promise example of FH, she concludes that a test for perfect duties to others can be generalized from the example: “An action is contrary to perfect duty if it is not possible for the other to assent to it or to hold its end” (Korsgaard 138).

She discusses at length what it means for it to be impossible for another to assent to or hold the end of one’s proposed action (138-9), and she concludes: “People cannot assent to a way of acting when they are given no chance to do so” (138). Coercion and deception are the two principle instances whereby one cannot give assent to an action (Korsgaard 138). She again illustrates this in terms of the false promise example, and says that the person loaning the money has no idea what is going on, and so cannot give consent to the action, because he or she does not know what it is (Korsgaard 138). She takes this even further and says that even if the person lending the money understands what is actually going on, but gives the money to the promiser with anticipation of its being return, so long as the lender does not call-out the promiser, he is not properly assenting to the action (Korsgaard 138-9). The lender is assenting to giving a hand-out, not to the lying promise proposed by the promiser.

Similarly, one cannot hold the same end as an agent if something in the agent's action "prevents her from *choosing* whether to contribute to the realization of that end or not" (Korsgaard 139). Again, coercion and deception are the primary instantiations of this failure of a perfect duty to others.

Korsgaard makes a strong claim:

According to the Formula of Humanity, coercion and deception are the most fundamental forms of wrongdoing to others—the roots of all evil. Coercion and deception violate the conditions of possible assent, and all actions which depend for their nature and efficacy on their coercive or deceptive character are ones that others cannot assent to. Coercion and deception also make it impossible for others to choose to contribute to our ends. (Korsgaard 140)

Coercion and deception are the primary means by which humans fail morally under FH. With this strong stance on deception being taken under FH, it is clear that in no way is even the philanthropic lie permissible. For example, in the case of the person harboring his friend, the murderer cannot assent to your end of keeping your friend alive because that end is concealed from him in your lie.

Korsgaard reaches similar conclusion regarding FKE, which is a derivative of FA, saying:

We are not only forbidden to use another as a mere means to our private purposes. We are also forbidden to take attitudes toward her which involve regarding her as not in control of herself, which is to say, as not using her reason. (Korsgaard 141)

She continues:

We owe to others not only a practical generosity toward their plans and projects—a duty of aid—but also a generosity of attitude toward their thoughts and motives. To treat another with respect is to treat him as if he were using his reason and as far as possible as if he were using it well. (Korsgaard 141).

She concludes: "This attitude is something that we *owe* to him, something that is his right. And he cannot forfeit it" (Korsgaard 141). We are bound by the CI, explicitly so under FA, to treat



others as though they are constantly using their reason to its fullest extent. Not only are we bound to do so, Korsgaard concludes, neither can others forfeit the right to this attitude. No one can commit any action that allows us to condemn him as not using his reason; “Even in a case where someone evidently *is* wrong or mistaken, we ought to suppose he must have what he takes to be good reasons for what he believes or what he does” (Korsgaard 141).

As such, lying to any agent is the complete dismissal of this duty to treat others as though they were using their reason at all times. The liar tries to remove the autonomous authority of another when she lies to someone, thus attempting to reduce the person’s reason to a mere means and the person to a mere cause (Korsgaard, 142). So applied to the case of the murderer at the door:

If you make a straightforward appeal to the reason of another person, your responsibility ends there and the other’s responsibility begins. But the liar tries to take the consequences out of the hands of others; he, and not they, will determine what form their contribution to destiny will take. By refusing to share with others the determination of events, the liar takes the world into his own hands, and makes the events his own. The results, good or bad, are imputable to him, at least in his own conscience. It does not follow from *this*, of course, that this is a risk one will never want to take. (Korsgaard 143)

Korsgaard brings her contemplation of the matter into line with Kant’s thought in “On a Supposed Right to Lie from Philanthropy,” where Kant explicitly says that a liar is responsible for the effects of her lie, regardless of whether or not the consequences were foreseeable (SRL 8:427). She goes further, however, and asks whether or not this risk of responsibility is one that we may, on certain occasions, be willing to take.

From her conclusion that the FUL gives different results for this problem than do FH, FA, and FKE, Korsgaard concludes:

This result impugns Kant’s belief that the formulas are equivalent. But it is not necessary to conclude that the formulas flatly say different things, and are unrelated except for a wide range of coincidence in their results. For one thing,

lying to the murderer at the door was not shown to be permissible in a straightforward manner: the maxim did not so much pass as evade universalization. For another, the two formulas can be shown to be expressions of the same basic theory of justification. (Korsgaard 143)

So Korsgaard denies that all the formulations of the categorical imperative are equivalent, but does maintain that their relationship is not one of mere coincidence. The different formulations, particularly FUL and FH are grounded on the same thing:

The Formula of Humanity is stricter than the Formula of Universal Law—but both are expressions of the same basic theory of value: that your rational nature is the source of the justifying power of your reasons, and so of the goodness of your ends.<sup>19</sup> (Korsgaard 144)

So though both formulae are not equivalent, they are grounded on the same principle that the worth of actions springs from the rational nature of the agent performing them.

Citing a passage from Kant's *Lectures on Ethics*, she tries to justify a lie under FH, saying:

The common thought that lying to a liar is a form of self-defense, that you can resist lies with lies as you can resist force with force, is according to this analysis correct. This should not be surprising, for we have seen that deception and coercion are parallel. Lying and the use of force are attempts to undercut the two conditions of possible assent to actions and of autonomous choice of ends, namely, knowledge and power. So, although the Formula of Universal Law and the Formula of Humanity give us different results, this does not show that they simply express different moral outlooks. The relation between them is more complex than that. (Korsgaard 144)

So she saves Kant's ethical theory by maintaining that the different formulae are grounded in the same way, though they do admit different actions as acceptable at times. The possibility of admitting lying as a form of self-defense in some cases under FH is one way of showing the connection between the formulae. FUL prescribes the universalizability of reasons, while FH

---

<sup>19</sup> I do not wish the reader to think that it has escaped me how closely Korsgaard's interpretation here supports my own version of the teleological contradiction interpretation. It is for purely structural reasons that I do not address the relevance of her point here to my version of the contradiction interpretation. However, I hope one can easily see that Korsgaard's interpretation of human rational nature as the source of moral value for FH and FKE coincides with my version of the teleological contradiction interpretation of FLN.

requires that we respect the reasons of others as manifestations of their humanity—their rational agency. Korsgaard maintains that the relationship between FUL and FH, FKE, and FA is a complex one, and she will draw on the ethical theory of Rawls to explicate this relationship fully.

### III. The Ideal and Non-Ideal Theory

Korsgaard appeals to the ethical theory of John Rawls in his *A Theory of Justice* to explain the complex relationship between the formulae of the CI. Her analysis relies primarily on four concepts taken from Rawls' writing: the general conception of justice, the special conception of justice, ideal theory, and non-ideal theory. She describes Rawlsian ideal and non-ideal theories thus:

The task of ideal theory is to determine “what a perfectly just society would be like” while nonideal theory deals with punishment, war, opposition to unjust regimes, and compensatory justice (Sec. 2, pp. 8-9). Since I wish to use this feature of Rawls' theory for a model, I am going to sketch his strategy for what I will call a double-level theory. (Korsgaard 147)

Later, she continues:

Ideal theory is worked out under certain assumptions. One is strict compliance: it is assumed that everyone will act justly. The other, a little harder to specify, is that historical, economic, and natural conditions are such that realization of the ideal is feasible. Our conduct toward those who do not comply, or in circumstances which make the immediate realization of a just state of affairs impossible, is governed by the principles of nonideal theory. (Korsgaard 147)

Ideal theory describes the behavior of people in a perfectly ethical society without injustice and without immoral members. It is also assumed in ideal theory that the ideal is plausible given the potential of societal conditions for its realization. Non-ideal theory, on the other hand, deals with behavior towards those who do not comply with the standards of ideal theory. If corruption is present in the community, non-ideal theory must be used to dictate behavior because the community is now non-ideal.

Rawls' theory also relies on two conceptions of justice: the general conception and the special conception. Korsgaard describes them thus:

The general conception tells us that all goods distributed by society, including liberty and opportunity, are to be distributed equally unless an unequal distribution is to the advantage of everyone, and especially those who fall on the low side of inequality (Sec. 13). Injustice, according to the general conception, occurs whenever there are inequalities that are not to the benefit of everyone (Sec. 11, p. 62). The special conception in its most developed form removes liberty and opportunity from the scope of this principle and says they must be distributed equally, forbidding tradeoffs of these goods for economic gains. It also introduces a number of priority rules, for example, the priority of liberty over all other considerations, and the priority of equal opportunity over economic considerations. (Korsgaard 147)

These conceptions are applied to the two types of theory, because in non-ideal conditions we are to look to the special conception of justice as the goal toward which we are working in our progress toward ideal conditions (Korsgaard 148). Another facet of this distinction is important:

Finally, the general conception of justice commands categorically. In sufficiently bad circumstances none of the characteristic features of the special conception may be realizable. But there is no excuse ever for violation of the general conception. If inequalities are not benefiting those on the lower end of them in some way, they are simply oppression. The general conception, then, represents the point at which justice becomes uncompromising. (Korsgaard 148-9)

So the general conception of justice prescribes that than which no less can be tolerated. The special conception of justice functions properly only under ideal conditions. Under non-ideal conditions, we should work towards the special conception but tolerate no less than the general conception.

Kant's theory, Korsgaard claims, is a single level theory as he sets it out. "The standard of conduct he sets for us is designed for an ideal state of affairs: we are always to act as if we were living in a Kingdom of Ends, regardless of possible disastrous results" (Korsgaard 149). One problem Korsgaard cites as arising from this conception of Kant's ethics as a single level theory is that though my responsibility is clearly limited if my action is moral. "The trouble is

that in such cases as that of the murderer at the door it seems grotesque simply to say that I have done my part by telling the truth and the bad results are not my responsibility” (Korsgaard 150). Another problem with the single level conception of Kant’s theory (she borrows both of these from Bernard Williams) is the phenomenon “of regret for doing a certain kind of action even if in the circumstances it was the ‘right’ thing” (Korsgaard 151). She uses these two concerns to justify incorporating Rawls’ distinction between ideal and non-ideal theory into Kant’s ethics.

#### **IV. The Incorporation of Non-Ideal Elements into Kant’s Ethics**

Korsgaard, at this point, appeals back to what she has said about the differences between FUL and FH in which she calls FH the stricter of the two formulations. She continues:

This comparison gives rise to the idea of using the two formulas and the relation between them to construct a Kantian double-level theory of individual morality, with the advantages of that sort of account. The Formulas of Humanity and the Kingdom of Ends will provide the ideal which governs our daily conduct. When dealing with evil circumstances we may depart from this ideal. (Korsgaard 151)

So FH and FKE can be seen as dictating an ideal system, which can be the ideal theory in a Kantian double level theory. The non-ideal theory here is of course dictated by FUL, which has been shown to be able to make accommodations for the presence of evil in certain situations.

So to apply this to the philanthropic lie, we can see that lying to the murderer at the door is permissible under the Kantian rubric. Ideally, we would be required to tell the murderer the truth, but since the murderer is acting immorally, and thus the Kingdom of Ends is not realized, we are permitted by FUL, which functions here as the general conception of justice does in Rawls’ theory, to lie to the murderer to save our friend’s life. The Kingdom of Ends becomes something toward which we should strive, but FUL provides the bottom-line below which we may not sink.

This incorporation seems to solve the concerns of Bernard Williams. The problem of limited responsibility is solved because the non-ideal state of affairs, dictated by FUL, allows for the lie. Likewise, guilt over what was the right action is explained as guilt over an action that is impermissible in the ideal state of affairs manifest in the Kingdom of Ends but permissible in the non-ideal state ruled over by FUL.

### **V. Some Problems and Concerns**

I have a number of concerns with Korsgaard's interpretation, and I hope that these concerns coupled with a satisfactory analysis of the philanthropic lie example by my revised teleological contradiction interpretation can show its merits over Korsgaard's interpretation.

First, Korsgaard's interpretation works fine for cases such as the murderer at the door only when the person, filling the role of the murderer, lies to you. Its inadequacies show up when similar cases are examined in which the role of the murderer is filled by one who does not lie to the agent in question. A relevant example here is the problem of handing over Jews hidden in your attic to a Nazi soldier inquiring if you have any. All other contingent concerns (such as the problem of whether or not the Nazi apprehending the Jews knows what will in fact happen to them after he takes them into custody or whether or not the agent herself knows that information) aside, the Nazi need not, and most likely will not lie when he shows up at the agent's door inquiring about the Jews—it is a given that he will take them into custody. If Korsgaard wants to escape the handing over of the Jews, then she must make the move of saying that this soldier is surrendering his right to the truth in that he is forcing you to hand over the Jews to him and his coercion justifies your lie. But this is a dangerous road to follow, which it seems can lead us down a slippery slope to some deadly conclusions.

The case I have in mind is the factual case of the potential inability to prosecute a cannibal under the German constitution (Neo-Kantian) that had his victims sign, freely, documents expressing their wish to be eaten. Because the victims had freely surrendered their autonomy, the constitution could not necessarily view them as moral agents. Thus, the prosecution of the cannibal could have been blocked by the idea that since the victims had freely surrendered themselves, they were not moral agents and, thus, were not humans in this case.<sup>20</sup> Korsgaard's view I think, will lead down a slippery slope to just such a case where cannibalism, in which the victim freely surrenders himself, is permissible. If we may lie to someone because she has surrendered, in some way, her right to the truth, then why may we not take action, which would otherwise be immoral, against those who do likewise? Could this lead all the way down to eating someone who willingly surrenders his status as a moral being? I think the frightening potential presents itself. So, one significant failing of Korsgaard's interpretation is that it does not effectively explain the philanthropic lie in the event that the role of the murderer at the door is played by someone who does not lie in the pursuit of his victim.

It is worth noting that Kant intends his discussion of the philanthropic lie in "On a Supposed Right to Lie from Philanthropy" to be a discussion for political philosophy, falling under the realm of the doctrine of right, and not ethics, dictated by the doctrine of virtue. This is the division that organizes *The Metaphysics of Morals* into its two parts. In this essay, Kant states that, "Truthfulness in statements that one cannot avoid is a human being's duty to everyone" (SRL 8:426), which he qualifies with a footnote:

I here prefer not to sharpen this principle to the point of saying: "Untruthfulness is a violation of duty to oneself." For this belongs to ethics, but what is under discussion here is a duty of right. The doctrine of virtue looks, in this

---

<sup>20</sup> The reader is referred to [http://www.cnn.com/2005/WORLD/europe/04/22/germany.cannibal.ap/index.html?section=cnn\\_latest](http://www.cnn.com/2005/WORLD/europe/04/22/germany.cannibal.ap/index.html?section=cnn_latest) and similar articles on CNN.com for further information regarding this case.

transgression, only to *worthlessness*, reproach for which a liar draws upon himself. (SRL 8:426n)

Consequently, what Kant says about the problem here must be attributed to the realm of political philosophy and not to pure ethics. Thus the attribution of all consequences, foreseeable or unforeseeable, of a lie to the liar is a principle only explicit in the doctrine of right, not necessarily in the doctrine of virtue. A similar passage from *The Metaphysics of Morals* anticipates the difference between the lie in each setting, but does not extend its discussion to responsibility for the consequences:

In the doctrine of right an intentional untruth is called a lie only if it violates another's right; but in ethics, where no authorization is derived from harmlessness, it is clear of itself that no intentional untruth in the expression of one's thoughts can refuse this harsh name. (MS 6:429)

But the footnote hints at a much larger issue, a problem that can be seen as stemming from the four examples of FLN in the *Groundwork*. In that work, the lying promise is treated as a violation of a perfect duty to others. But, by the time Kant wrote and published *The Metaphysics of Morals*, twelve years later, the transgression of lying (including the lying promise) is treated as a violation of a perfect duty to oneself. Why the change? The footnote and the passage from *The Metaphysics of Morals* indicate that in the doctrine of right it is necessary to view the transgression of lying as a violation of duty to others, while it is only in the realm of ethics, the doctrine of virtue, to treat it as a violation of a perfect duty to oneself.

I do not wish to speculate here on the reasons why political philosophy and ethics, in Kant's view, do not treat the case in the same way. What I wish to dwell upon is Kant's transition from describing lying as a violation of an ethical duty to others to describing it, twelve years later, as a violation of an ethical duty to oneself. This can have serious consequences for this example of lying, and may show why he takes such a definitive stance on lying in *The*



*Metaphysics of Morals*. It also has the power to undermine Korsgaard's attempt to justify the philanthropic lie under any circumstances, regardless of whether someone forces or coerces the agent.

It is my strong conviction that Kant's theory cannot permit a lie to another, at least, in so far as that person is telling the truth to you. And Kant's move to group truthfulness into perfect duties to oneself seems to justify this conclusion. Again, Kant's stance on lying in *The Metaphysics of Morals* is clear-cut and direct:

By a lie a human being throws away and, as it were, annihilates his dignity as a human being. A human being who does not himself believe what he tells another (even if the other is a merely ideal person) has even less worth than if he were a mere thing; for a thing, because it is something real and given, has the property of being serviceable so that another can put it to some use. (MS 6:429)

Lying is "the greatest violation of a human being's duty to himself regarded merely as a moral being" (MS 6:429). Kant claims that lying renders the liar of less worth than a mere object, which is, at least, useful for some purpose. The liar has completely sacrificed his humanity, his dignity, by his lie; he "is a mere deceptive appearance of a human being, not a human being himself" (MS 6:429).

It is, however, not just the permissibility of a lie that I hold against Korsgaard's double level theory proposal. For one thing, Kant explicitly says in the *Groundwork*, "one does better always to proceed in moral *appraisal* by the strict method and put at its basis the universal formula of the categorical imperative: *act in accordance with a maxim that can at the same time make itself a universal law*" (G 4:436-7). This seems immediately to nullify Korsgaard's claim that FH is a more strict and more rigorous formula of the CI than is FUL or FLN. If this claim is undermined, then the difference between the two formulae is destroyed, and the distinction upon which Korsgaard constructed the double level theory is likewise undermined.

Furthermore, if a contradiction interpretation of FLN can show the impermissibility of a lie in all possible scenarios, then one need not compound and add to the complexity of Kant's ethical theory by introducing the idea of ideal and non-ideal theory. The lack of the discrepancy in duties prescribed by FLN and FH does away with the need for it. It also helps support Kant's own statement that the formulae of the CI are actually just different wordings of the same law.

I believe that my contradiction interpretation can do so, especially when one considers Kant's deliberations on lying from *The Metaphysics of Morals*. Once the lie is viewed as a violation of a perfect duty to oneself, a proper interpretation of FLN should have no problem in accounting for its impermissibility. Kant's revision in the theory seems to indicate a need to consider truthfulness as a duty to oneself, and it may have been problems like the philanthropic lie that caused him to make that move.

The maxim of a liar can be seen as containing a simple teleological contradiction in just the sense that Kant outlines in *The Metaphysics of Morals*:

But communication of one's thoughts to someone through words that yet (intentionally) contain the contrary of what the speaker thinks on the subject is an end directly opposed to the natural purposiveness of the speaker's capacity to communicate his thoughts. (MS 6:429)

The purpose of language is truthful communication. The purpose of lying is deceptive communication. The contradiction is easy to see. What now needs to be established is how this simple teleological contradiction indicates the undermining, by the agent of his own freedom, his status as a final end of nature. In the same passage from which the above is excerpted, Kant indicates that the lying agent is undermining his freedom by renouncing his personality. Later, he speaks of the agent as using his dignity, his trustworthiness as a rational being, as a mere means to accomplish another end (MS 6:430). Either of these seems a plausible explanation of this situation.

## Conclusion

The extreme importance of Kant's teleology for understanding and making sense of Kant's ethics cannot be emphasized enough. Humanity as noumena is at the center of the natural system, and it provides nature with its final end—that end which is outside of nature, but which all ends within nature are subordinated to, while it is subordinated to none. Even humanity as phenomenon is meant to serve this final end of nature. It is with this in mind that we are to understand Kant's contradiction in conception and contradiction in will as used in the *Groundwork of the Metaphysics of Morals*. The three previous contradiction interpretations neglected this idea and, I feel, suffer as a result.

The three contradiction interpretations that stand without an appeal to the final end of nature all demonstrate some fundamental flaws that keep them from successfully answering all four of the examples of FLN in the *Groundwork*. A contradiction interpretation that does not succumb to the pitfalls of the other three be devised from Paton's teleological contradiction interpretation. Refining the view of teleology incorporated in Paton's view can make it immune to Herman's objection, and it can successfully avoid relativism.

The contradiction interpretation that is delineated here maintains that the contradiction in conception is a simple teleological contradiction within a maxim that results from the agent's undermining of his own or another's rational freedom. The agent is undermining the final end of nature. In the case of the contradiction in will, the agent fails to further this final end, but does not directly undermine its status as final end. I maintain that this new interpretation can successfully accommodate all four examples, and answer some key objections to the other three

interpretations while being flexible enough to explain the moral puzzles that Kant laid out as “casuistical questions” in *The Metaphysics of Morals*.

### Works Cited

- Aune, Bruce. *Kant's Theory of Morals*. Princeton: Princeton University Press, 1979.
- Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: The University of Chicago Press, 1960.
- Herman, Barbara. "Mutual Aid and Respect for Persons." *Ethics* 94 (1984): 577-602.
- Kant, Immanuel. *Critique of Practical Reason*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press, 1997.
- . *Critique of Pure Reason*. Trans. and ed. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1998.
- . *Critique of the Power of Judgment*. Ed. Paul Guyer. Trans. Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press, 2000.
- . *Groundwork of the Metaphysics of Morals*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press, 2002.
- . "On a Supposed Right to Lie from Philanthropy." *Practical Philosophy*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press, 1996. 611-5.
- . *The Metaphysics of Morals*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press, 1996.
- Kemp, J. "Kant's Examples of the Categorical Imperative." *The Philosophical Quarterly* 8 (1958): 63-71.
- Korsgaard, Christine. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

Nell, Onora. *Acting on Principle: An Essay on Kantian Ethics*. New York: Columbia University Press, 1975.

Paton, H. J. *The Categorical Imperative: A Study in Kant's Moral Philosophy*. Philadelphia: University of Pennsylvania Press, 1971.

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

Wood, Allen. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999.