Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Mathew Sperling April 18, 2018

A Kantian Solution to the Problem of Moral Dilemmas

by

Mathew Sperling

Thomas R. Flynn
Adviser

Department of Philosophy

Thomas Flynn
Adviser

Marshall Duke
Committee Member

Cynthia Willett
Committee Member

2018

A Kantian Solution to the Problem of Moral Dilemmas

By

Mathew Sperling

Thomas Flynn
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Philosophy

2018

Abstract

A Kantian Solution to the Problem of Moral Dilemmas

By Mathew Sperling

Critics of universally normative theories of ethics often proceed by constructing multi-pronged thought experiments wherein multiple morally good actions are incompatible with each other or in which all possible actions are odious. If a moral system requires the impossible – the performance of mutually exclusive actions or the rejection of an exhaustive set of actions – then a moral dilemma results. The theory is not, then, action-guiding and applicable in a morally relevant situation and is therefore no longer universally normative. An ethical program is similarly at fault if it is entirely silent, offering neither guidance nor any explanation for why such a situation falls outside the moral domain. Furthermore, a moral theory that does discharge the dilemma by promoting one prong over the other may face suspicion if its grounds for doing so are entirely arbitrary (without any meta-ethical justification) or if its prescriptions are contradictory or supremely counter-intuitive.

In this paper, I argue that the Kantian system of practical philosophy supplies a meta-ethically robust means of discharging moral dilemmas. I trace Kant's arguments from postulates that are necessary for the possibility of morality through the derivation of a normative principle in the form of the categorical imperative. The logical necessity of this principle, precludes internal contradiction and the subjective manner in which it is actualized through the evaluation of maxims (first order principles of actions) tailors its directives to the set of actions available in a particular scenario. The categorical imperative, then, allows for universally consistent action-guidance without rigidly enforcing broad normative requirements to the exclusion of other equally valid imperatives.

Finally, through an analysis of the set of thought-experiments known as "trolley scenarios," I examine the limited scope of the categorical imperative's practical evaluation. I assess and dismiss normative principles which are excluded from a categorical analysis of moral duty by the meta-ethical postulates of Kantian ethics. These normative principles, not being derived from pure practical reason, unsurprisingly also result in intuitive contradictions. Ultimately the limitations of the categorical imperative provide intuitive consistency, allowing it to properly distinguish practically necessary perfect duties from assertorically valid, but contingent, imperfect duties.

A Kantian Solution to the Problem of Moral Dilemmas

By

Mathew Sperling

Thomas Flynn
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Philosophy

2018

Acknowledgements

Special thanks are owed to professors Flynn, Willett, Duke, and Huseyinzadegan of Emory University for their guidance. Thanks also to Professor Kleingeld of the University of Groningen for corresponding with me about her paper on Kant and the trolley problem. My Parents, my entire family, and all of my friends were endlessly patient as I interrogated them about their moral intuitions – when they would or would not pull a particular lever or push people off bridges. This thesis could not have been completed without their help and support. Finally, my gratitude to Tarrek Shaban for his help in creating the figures in this paper.

Table of Contents

List of Figures

Introduction

The concept of moral dilemmas is a pernicious one in modern moral philosophy. It has long beleaguered advocates of universal moral obligation, suggesting that, perhaps, there exist morally relevant situations in which no moral system is action-guiding. Among the chief virtues of Kantian moral philosophy is that it purports to allow for universal obligation. In the *Groundwork of the Metaphysics of Morals* and, later, in the *Critique of Practical Reason* and *The Metaphysics of Morals*, Kant investigates the metaphysical and epistemological structures from which it is possible to derive a fundamental principle of moral obligation and begins to develop the system of ethics that such a principle would establish. Kant's ethical system is among the oldest and most metaphysically robust to present a justification of universal duty independently of any reliance on divine law or on hypothetical ends which, he argues, are merely contingent and never, therefore, universally binding. Kant, however, wrote very little on the subject of conflicts between moral obligations. In the preface to *The Metaphysics of Morals*, he cryptically claims that – at least with regard to perfect duties – no such conflict can exist. Despite Kant's reticence on the topic, much effort has been spent and ink spilled attempting to understand how Kant's ethical system resolves or fails to resolve moral dilemmas. The following pages will, drawing on similar investigations by Jens Timmermann and Pauline Kleingeld, examine whether moral dilemmas present a possible counterexample to the Kantian ethical project.

I propose that a system of Kantian ethics is able to answer and resolve the problems posed by seeming moral contradiction in a way that is impossible for a simple arrangement of immediately applicable moral principles. The resolution lies in the Kantian commitment not only to normative conditions of discharging moral obligation but also to the meta-ethical structure which determines the conditions of obligation. I first examine the relationship between moral

dilemmas and the deontic structure of moral requirement and introduce the challenges that moral dilemmas pose to moral systems in general (section 1). I then outline Kant's meta-ethical arguments, which proceed from the concept of duty itself to the identification of autonomy with the moral law, and consider the possibility and nature of moral conflict within this conception of morality (section 2). Finally, I describe the manner in which these meta-ethical principles commit a Kantian to the resolution of moral dilemmas and transition into describing the normative ethical commitments which can be derived from – or are excluded by – these principles (section 3). Throughout the paper, I rely extensively on examples and thought experiments, particularly in section 3 where I explicate Judith Jarvis Thomson's "trolley problem" as I compare the intuitive insufficiency of certain normative principles to their theoretical invalidity under the Kantian framework.

I.

A moral dilemma, generally, consists of an actual or imagined situation in which it is not possible for an agent to act entirely morally or in which it is unavoidable to act immorally. More specifically, moral dilemmas are situations in which one moral obligation conflicts with another in a way that cannot be resolved in favor of acting upon either obligation. This may be because neither requirement supersedes the other, in which case the dilemma is an ontological one, or because it is impossible to know which moral requirement takes precedence, in which circumstance the dilemma is epistemological. Of these, only deontological dilemmas can be leveled as counterexamples against the Kantian ethical project. As will be demonstrated in section 2, Kant's practical philosophy is primarily concerned with intent and the development of

the will. Action-guidance necessarily follows from this foundation but consequences in the world only follow possibly or accidentally. As a result of this structure and of the principle that "ought" implies "can" it will be true in any case of moral obligation that, if a resolution is possible, it will also be knowable.

Moral Dilemmas can, furthermore, result from the conflict of either positive or negative obligations. In the former case, a dilemma occurs when there are at least two actions each of which is morally required and all of which cannot be pursued. In the latter instance, a moral dilemma arises when at least two actions are available to the agent, each of which is morally prohibited, but at least one of which must be undertaken. Because the nature of moral dilemmas is that one action or a set of actions must be taken to the exclusion of others, most situations of this type can be viewed as arising equally from either requirement or prohibition.

In his lecture, *Existentialism is a Humanism*, Jean–Paul Sartre provides an example of a moral dilemma that can be alternatively construed as arising from positive or negative obligation. Sartre tells the story of a pupil who, during the Second World War, comes to the philosopher for advice. The young man is confronted with the decision of either remaining in France to care for his ailing mother or departing England to join the Free French Forces. His mother, old and unwell, has no one but himself. She is separated from his father who, in any case, is a collaborator with the Germans. Her elder son, the student's brother, had been killed in the 1940 invasion of France. It is evident to the young man that "his mother lived only for him and that his absence – perhaps his death – would plunge her into utter despair."[1] He also, however, is driven by "primitive but noble feelings"[2] to avenge his brother's death and to contribute to the fight for

[1] Jean-Paul Sartre, *Existentialism Is a Humanism* (New Haven, CT: Yale University Press, 2007), 30.
[2] Ibid.

his country and against tyranny and fascism – even if it means laying down his life. The student is thus torn between apparent duties. Read as arising from positive obligation, the student faces the requirement to care for his mother that conflicts with the requirement to fight for his nation, values, and countrymen. Construed as a conflict of negative obligations, the student faces, on one side, a prohibition against abandoning his doting mother and, on the other, against forswearing his values and country.

A foundational requirement for a moral dilemma (already implied in the discussion above) is that it must be a situation in which moral obligation is implicated. This is to say, a moral obligation must be present and capable of contravention. For some systems of universal ethics, such as utilitarianism, this requirement is unimportant because uniquely-guiding moral obligation is ubiquitous. However, for Kantian ethics and other deontological frameworks, this aspect is critical in order to get to the heart of what is meant by a *moral* dilemma. Kantian ethics is not uniquely action-guiding in all scenarios, nor should it be. There are actions that, as they stand in relation to other possible actions, are neither required nor prohibited, but merely permissible.

It will be useful here to explicate, in some detail, the ways in which the deontic value of a single action can change depending on the presence of other possible actions. In circumstances in which no moral obligation prevails upon an agent, all possible actions are permissible. When choosing from among breakfast cereals to eat, it is of no moral consequence whether I select Weetabix or Special K. Of course, in practice, there are almost always certain possible actions that are prohibited. Potentially, I could eschew my cereal and steal my roommate's much more appealing French toast. Alternatively, I could forgo breakfast altogether and, instead, commit suicide. Either of these actions would be prohibited by Kantian ethics. Critically, the presence of

these forbidden actions does not affect the deontic value of any individual permissible action. They will each remain perfectly permissible in relation to each other. *A prohibition against committing certain acts can be equally construed as a requirement to take some other action.* In the presence of prohibited actions, there is a requirement to undertake an action from among those that are not forbidden. Contrarily, the presence of a required action will dramatically alter the deontic value of other actions. If, while I am considering my breakfast options, it comes to my attention that my roommate is choking, I may be morally required to come to his assistance. In the presence of such a requirement, actions that would otherwise be prohibited remain so but actions that would otherwise be merely permissible become prohibited. The weight of a moral requirement is such that, when a requirement prevails upon an actor, all other possible actions are rendered impermissible.

As a consequence of this deontic structure, it is already possible to dismiss certain circumstances as possible sources for moral dilemmas. Kantian ethics is not intended to be uniquely action-guiding in all circumstances and the absence of an imperative in a situation is not enough to constitute a moral dilemma. The choice of cereals, while not resolvable by Kantian practical philosophy, does not meet the threshold demands for a moral dilemma because no overriding moral obligation is implicated. The situations which will be relevant are those already outlined above wherein it is not possible both to avoid all prohibited acts and to discharge all moral requirements. A critical question which must be addressed is the precise status of typically permissible actions which are prohibited to the actor by the prevailing presence of some moral requirement. I have claimed that such actions are rendered impermissible. The question persists whether, in a Kantian ethical system, such actions remain of greater deontic value than actions which are themselves prohibited, independent of the presence of any moral requirement. If my

roommate is choking, in the above example, and I have resolved to ignore his plight, is it still worse to steal his food? Regardless of whether I eat my cereal or his French toast, I will be committing a prohibited action.

This question is even more critical in cases of moral dilemmas. In Sartre's example, it seems apparent that, even if neither joining the French forces in England nor staying with his ailing mother will fully discharge the student's duties, he is at least required to do one of the two. He could not, for example, flee to Switzerland or abandon his mother to become a collaborator. This intuitive truth will be difficult to justify in Kant's deontological ethics. As Sartre notes, the guidance Kant provides to "Act [in such a way] 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"[3] is not sufficiently action-guiding in this instance. The student worries "… if I stay with my mother, I will treat her as an end, not as a means. But by the same token I will be treating those who are fighting on my behalf as a means; and the converse is also true, that if I go to the aid of the combatants I shall be treating them as the end at the risk of treating my mother as the means."[4] Regardless of his actions, it seems the student will be violating a moral requirement and that his behavior will therefore be prohibited. If all prohibitions are equally forceful, then it may seem that no deontic distinction can be drawn between treating his mother as a means to serve his country or, alternatively, as a means to betraying it. In order to explain fully how or even whether Kantian ethics avoids such a conclusion, it will be necessary to devote some space

---

[3] Kant, *Groundwork of the Metaphysics of Morals*, 4:429.
Kant's writings will be cited by the title of the work followed by the standard pagination of the Berlin Academy edition of Kant's work. The translation which I use is from *The Cambridge Edition of the Works of Immanuel Kant*, the full citation for which can be found in the bibliography.

[4] Sartre, *Existentialism is a Humanism*, 31.

to the premises, structure, and rationale of Kant's practical philosophy. This discussion can be found in section 2 below and will provide the basis for a full exploration of the relationship between the Kantian system and the problem posed by moral dilemmas.

Before undergoing such an extensive examination, however, it will be beneficial to understand what is at stake. What exactly are the ways in which moral dilemmas are used to jeopardize systems of universal ethics? The first, already discussed above, occurs when – within the structure of the disputed system of ethics – the moral dilemma is what I will call a *true* or *normative* dilemma. This is to say that it is a situation in which moral requirements or prohibitions are implicated and for which no resolution can be given. As such, the ethical system fails to be singularly and universally action-guiding. However, even if it can be shown – as I hope to demonstrate in regard to the Kantian practical project – that a system of ethics resolves all morally relevant circumstances, two other types of argument may still be marshaled against the system.

The possibility of a normative moral dilemma is precluded if a system of ethics resolves all morally relevant circumstances. Another form of dilemma, however, may cause equal difficulties. It is evident, even to those of us who would deny normative dilemmas, that situations exist in which most people will feel conflicted about what is morally required or – even worse – in which the majority of people will feel strongly that one action is required and our preferred moral system will contradict this intuition.[5] These types of conflicts do not arise from any rigorous standard of ethics but, instead, from empirical studies of what people believe about obligation. As such, I will refer to these types of problems as *apparent* or *descriptive* dilemmas.

---

[5] In his speculative philosophy, Kant uses the word "intuition" as a precise term of art. In this essay, I use the term "intuition" in the vernacular sense of a *prima facie* belief, not in Kant's technical sense.

Descriptive dilemmas raise problems for universal moral systems in two ways. The first of these difficulties involves "practical residue" and presents a challenge to the idea that it is possible or even desirable to resolve moral dilemmas. It is most apparent when considering situations in which it is normal to feel conflicted about moral obligation. Even if a system of universal ethics claims that Sartre's pupil does not face a true moral dilemma – that, in fact, his moral obligations can be fully discharged by a single set of compatible actions – the student is unlikely to feel as if he has fulfilled his duty. Instead, he is likely to feel what Sartre describes as "anguish." Even if it could be demonstrated to the student with absolute logical certainty that the upright and moral course of action is to join his free countrymen in London, he will feel that he has done wrong to his mother. Bernard Williams, who has written extensively on this subject, identifies, in particular, the emotions of regret and relief [6] but to these we could add emotions such as remorse, guilt, shame, and a feeling of blame-worthiness for any harms resulting from the action at issue. If these emotions are both appropriate and indicative of some moral failing, as at least feelings of guilt and blame-worthiness seem to be, then they could indicate that the moral dilemma has not been resolved. However, the presence of such emotions also lends credence to the notion that such dilemmas can never be resolved (because emotional residue will always remain) or that resolution might not even be desirable (because doing so would entirely divorce actual moral responsibility from the more intuitive responsibility entailed by shame and guilt).

The second manner in which descriptive dilemmas are problematic for universal systems of ethics arises when it is accepted as desirable for an ethical system to endorse those actions which are popularly or intuitively considered moral and to prohibit actions which are generally

---

[6] Williams, Bernard, and Atkinson, W. F., *Ethical Consistency* (Proceedings of the Aristotelian Society, Supplementary Volumes 39, 1965), 112 & 113.

thought to be wrong. If this position is accepted, then a conflict arises when an action-guiding system allows or requires an action that is intuitively wrong or prohibits an action that is generally thought to be permissible. In such a situation, it can be said that a descriptive dilemma exists for the ethical theory. This is, largely, the approach taken by Philippa Foot and Judith Jarvis Thomson in their examinations of moral dilemmas. To take one of Foot's examples, the majority of people agree that if a judge were "faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community"[7] it would be inappropriate and impermissible for the judge to frame and have executed some innocent person. If an ethic theory such as utilitarianism permits such an action or even (other methods of avoiding larger loss of life not being available) requires it, then the situation and its intuitive force can be used as a *reductio ad absurdum* against the theory.

Similarly, if an ethical system refuses to recognize a distinction that is intuitively valid or distinguishes between two cases that are generally thought to be the same, there could be reason to regard the system with suspicion. Foot and Thomson use this as part of a method of outlining and testing moral systems, modifying a moral hypothesis until no contradictions or descriptive moral dilemmas remain. The transcendental method by which Kant develops his ethical system is incompatible with this but some Kant scholars, such as Pauline Kleingeld,[8] have adopted the proposition that it would be preferable for Kantian ethics to align with popular intuitions and both to preclude true dilemmas and to resolve apparent dilemmas.

---

[7] Foot, Philippa, *The Problem of Abortion and the Doctrine of the Double Effect* (Oxford Review 5, 1967) 9.

[8] Kleingeld, Pauline, *The Kantian Solution to the Trolley Problem Puzzle* (presented at The American Philosophical Association Eastern Division One Hundred Tenth Annual Meeting, December 28, 2013).

II.

Whereas Foot and Thomson's method of ethical investigation requires a generally accepted principle to modify, the Kantian transcendental project is an entirely different animal. It begins by inquiring into the boundary requirements of morality.[9] The pivotal realization of Kant's practical philosophy is that human freedom is required for morality. In the Third Antinomy of the *Critique of Pure Reason*, Kant had demonstrated that freedom is at least problematically possible as a transcendental idea; i.e. it is possible as a limiting concept on the scope of pure reason but cannot necessarily be applied qualitatively to any object.[10] The central development which marks the shift from pure speculative reason to practical philosophy is the elucidation of the way in which freedom manifests itself in the rational will.

Freedom, as a transcendental idea is a form of causality other than that in accordance with the mechanistic laws of nature. It is a causality "through which something happens without its cause being further determined by another previous cause, i.e., an absolute causal spontaneity

---

[9] My reconstruction of Kant's arguments states the most defensible reading of the practical philosophy that he first established and that many later commentators have developed. In Kant's own writings, he is prone to self-contradiction and to the introduction of dubious ideas which seem not to follow from the main body of his reasoning. If accepted, these dicta would jeopardize the integrity of his transcendental system of ethics. An example: In the *Groundwork of the Metaphysics of Morals* and in the *Critique of Practical Reason*, Kant insists that the only possible moral obligation is derived from pure practical reason and includes nothing contingent. However, in *The Metaphysics of Morals* – to say nothing of his *Anthropology from a Pragmatic Point of View*, Kant assumes that the "natural superiority" of men (6:279) will be reflected in the actualization of practical reason. A comprehensive and reasonable system of ethics can be extracted from the premises of Kant's practical writings, but only by departing from Kant's own political and pre-philosophical attitudes. I maintain a presumption of gender equality even when Kant's language is explicitly gendered. As such, the system I describe is Kantian, if not precisely Kant's. My reading of Kant has particularly been influenced by the commentaries of Henry Allison, Thomas Hill, Pauline Kleingeld, and Jens Timmerman.

[10] Kant, *Critique of Pure Reason*, A444/B472 – A450/B478 & A553/B561

beginning from itself a series of appearances that runs according to natural laws."[11] This is all to say that freedom is a faculty by which the universe ceases to progress like a chain of dominoes wherein the fact of future events is already entailed by present facts. However, this idea of freedom cannot be applied to any object of experience. It can only be encapsulated by the empirical thought that an event *ought* to occur.

This *ought* is inescapable. It is a constituent of our empirical consciousnesses to which we are led again and again. Even the most ardent determinist, when cut off in traffic, will be irked by the thought that his fellow motorist ought to have done differently. The transcendental idea of freedom assures us that this notion is no mere folly or contradiction; it is, at least, conceivable. And this persistent *ought* allows us to consider freedom as more than a problematic limit on reason – which, as Kant notes, "is far from constituting the whole content of the psychological concept of that name."[12] Instead, as we shift from speculative to practical philosophy, we are allowed to posit freedom as a postulate of practical reason and to apply it to an object: our intelligible character.[13]

It is possible to examine a person from the standpoint of speculative philosophy, using neurology, sociology, or any of many other anthropological sciences. From this vantage point, we could explain all human action many times over as the product of brain chemistry, upbringing, or circumstance. In such an exercise, our subject is studied as an empirical character and there is neither a need nor room to encompass freedom. Humanity is analyzed as one might appraise a particularly complicated machine. However, human action can also be explained in another way: as the product of willful deliberation. Whenever we apply an imperative to say that

---

[11] Ibid., A446/B474.
[12] Ibid., A448/B476.
[13] Ibid., A539/B567.

a person ought to act in a certain way, we are engaged in this type of practical appraisal. The subject of our investigation cannot be the person's empirical character but is, instead, their intelligible character. In the *Groundwork of the Metaphysics of Morals*, Kant describes this bifurcated nature as though people are at once the citizens of two worlds: the world of sense, in which our actions are determined by natural causality, and the world of understanding, wherein human actions are freely determined and in which practical philosophy is exclusively possible.[14] The form that freedom takes as a postulate of practical philosophy applied to the intelligible character will, along with the universal demands of reason, determine the structure and contours of Kantian ethics.

Will, for Kant, is the volitional capability of rational beings. In order for the understanding to grasp morality, it is not sufficient that the will, and therefore human actions, should be undetermined or even spontaneously determined. Mere chance or caprice is not a freedom whereby it could be said that events ought to unfold according to a moral precept. If a person ought to act in a certain way, it cannot be that their action is determined by outside causes, nor that it is accidental. Instead, the freedom which is applied to the intelligible character is particularly a faculty through which the will is its own cause. Freedom is, in fact nothing other than autonomy: the will acting as a law to itself.[15]

Because freedom and reason each apply in the same way to all rational beings, we would also expect morality to apply universally. This, in fact makes a good working definition for understanding the Kantian conception of morality: It is a system of guiding deliberation which applies to all people in all situations. It will not be altered by incidental facts about a rational

[14] Kant, *Groundwork of the Metaphysics of Morals*, 4:453.
[15] Kant, *Critique of Pure Reason*, A534/B562.

being such as nationality, occupation, or even desires or inclinations. Morality will, instead, be universally and uniquely action-guiding. This has an intuitive ring of truth to it. If a person applies moral principles selectively, adopting or rejecting them depending on circumstances, or applying different principles to herself than to others, it suggests insincerity and hypocrisy. Moreover, Kant cautions that basing morality on accidental features such as what satisfies our pleasures necessarily results in a heteronomy of the will in which human actions can only have been determined by outside causes.[16] For moral obligations, then, it must be that the deliberative will could always freely determine the course of human action and therefore be autonomous. Recalling the deontic structure from section 1, moral prohibition amounts to the contravention of an obligation and is therefore consistent with heteronomy of the will. When a person acts immorally, it can only be said that they could have self-determined but instead allowed their actions to be dictated by external causes. Accordingly, it becomes clear that freedom is not simply necessary for morality; to act freely is precisely to act morally.

Kant is able to state, therefore, that "freedom and unconditional practical law reciprocally imply each other."[17] Henry Allison writes eloquently and at much greater length about this "reciprocity thesis"[18] which is found in a variety of forms both in the *Groundwork of the*

---

[16] Kant, *Groundwork of the Metaphysics of Morals*, 4:414 & 415.
[17] Kant, *Critique of Practical Reason*, 5:29.
[18] Allison, Henry E. *Morality and Freedom: Kant's Reciprocity Thesis* (The Philosophical Review 95, no. 3. July 1986) 394.
In this article, Allison draws a distinction between a practical, compatibilist freedom which is simply the deliberative property of the empirical will and autonomous freedom that is not determined by exterior forces and which, he insists is freedom in a "transcendental sense". To incorporate this division into my paper would obscure the distinction I am trying to make between transcendental freedom as a problematic idea of reason and practical freedom as a postulate of moral reasoning. Additionally, Kant's terminology supports both Allison's distinction and mine in similar words. In the *Critique of Pure Reason*, Kant writes that "**Freedom in the practical sense** is the independence of the power of choice from necessitation by impulses and sensibility" – exactly the definition of practical freedom as I use it. A534/B562

Metaphysics of Morals and in the *Critique of Practical Reason*. Here, it will be sufficient to reproduce the arguments in the third section of the *Groundwork* in which Kant derives the structure of moral law from the autonomous form of the free will:

> Natural necessity was a heteronomy of efficient causes, since every effect was possible only in accordance with the law that something else determines the efficient cause to causality; what, then, can freedom of the will be other than autonomy, that is the will's property of being a law to itself? But the proposition, the will is in all actions a law to itself, indicates only the principle, to act on no other maxim that that which can also have as object itself as a universal law.[19]

Here at last, is the famous categorical imperative – the underlying law which structures all moral obligation in Kantian ethics. This law, derived analytically from the practical postulate of human freedom, is uniquely suited to serve as the fundamental principle of universal morality. It is purely formal, being derived from pure practical reason and exists apart from the accidental components of practical, deliberative reasoning (such as inclinations and desires) which can only result in heteronomy. In this way, it can be distinguished from hypothetical imperatives which require the existence of a possible or actual externally conditioned end in order to take force whereas the categorical imperative has practical force apodictically.[20] It is action guiding and, if followed, is coextensive with autonomous volition. To follow the dictates of morality, for Kant, is straightforwardly to "act only in accordance with that maxim through which you can at the same time will that it become a universal law."[21] In order to determine whether moral dilemmas present a problem for the Kantian ethical project, what must be assessed is whether there is any

---

[19] Kant, *Groundwork of the Metaphysics of Morals*, 4:446 & 447.
[20] Ibid., 4:414 & 420.
[21] Ibid., 4:421.

situation in which the categorical imperative requires multiple incompatible actions or prohibits all possible actions (including taking no action at all).

This question, however, is difficult to settle. The categorical imperative, firstly, does not directly evaluate actions. Instead, it assesses maxims, the subjective principles of actions. Jens Timmermann has an excellent article on the technical although not entirely precise use of the term "maxim" in Kant's writings. Generally speaking, a maxim expresses the reason or motivation underlying an action and is comprised of an action to be undertaken, the conditions under which the action will be performed, and the end to which the action is directed. Maxims can be distinguished from either hypothetical or categorical imperatives which are prescriptive instead of subjective. (An imperative takes the form "In situation X, *one should* take action P," whereas a maxim expresses the attitude that "in situation X, *I will* take action P.") However, the term "maxim" incorporates a wide assortment of subjective principles of varying degrees of relation to actual action.

Often, when Kant writes about maxims, he refers to first-order principles of volition which directly condition actions. Timmermann notes that "Whenever we act in the full sense of the word – that is, when we act on more than a mere reflex or impulse – we act on some maxim of this sort."[22] So, if I see a dollar on the ground and stop to pick it up, my maxim might be something along the lines of, "If I see a small sum of money on the ground and no owner is apparent, I will pocket the money in order to enrich myself." This type of direct volitional principle is what I will be largely concerned with in applying the Kantian moral framework to the problems suggested by possible moral dilemmas. A different class of maxim, is a more broad

[22] Timmermann, Jens, *Kant's Puzzling Ethics of Maxims* (The Harvard Review of Philosophy 8, 2000), 40.

and generally applicable principle or "life-rule"[23] such as, "when I have the opportunity to enrich myself at no significant inconvenience, I will take advantage of it." Inasmuch as it is necessary for people to adopt such overarching maxims, they are still required to do so in accordance with the categorical imperative. However, in order for such maxims to effect themselves as action, there must be, as a mediate step, a maxim of the first kind which is a primary volitional principle. Because moral dilemmas concern action, it will be sufficient to examine first-order maxims. Lastly, a principle of volition by which other maxims are chosen can also be called a maxim.[24] These are, in some regards the most important to correctly cultivate because they determine whether the categorical imperative, and as such the demands of morality will be adhered to. Thus, it would be morally good to adopt a maxim of putting duty before pleasure and not the reverse. However, like life-rules, these fundamental volitional principles are brought into action by their instantiation in first-order maxims.

In section 1, I analyzed the deontological categories of requirement, permissiveness, and prohibition. A moral dilemma arises either when two or more required actions are mutually exclusive or when all possible actions are prohibited. However, the categorical imperative serves as a means of regulating the pre-active selection of maxims, not of judging actions themselves. In order to determine whether moral dilemmas can arise in the Kantian framework, it should be explained how the evaluation of maxims will determine the deontic status of actions. First, it must be clear that, although any first-order maxim can lead to only one action, any single action can be potentially motivated by a variety of different maxims. In the *Groundwork of the Metaphysics of Morals*, Kant explains that, while actions in accordance with duty can be the

[23] Ibid., 41.
[24] Ibid., 40.

product of an autonomous and moral will, they can also result from heteronomous determination. He provides the example of a shopkeeper, presented with the opportunity to cheat an unsuspecting customer. If the shopkeeper acts honestly, he could be motivated either by the imperative of pure practical reason or by externally conditioned selfishness and fear that he will be caught.[25] We cannot know if his maxim is one that could be willed to be a universal law ("I will, in all dealings, treat my customers with dignity and not take advantage of their trust") or if it is a maxim in violation of the categorical imperative ("I will, in dealing with my customers, think first about my reputation").

Pauline Kleingeld supplies an idea of how, by examining the maxims that could justify an action, we can determine its deontic status: If "all maxims which could possibly underlie the act in question […] fail to meet the demands of the Categorical Imperative"[26] then the action is contrary to duty and therefore morally prohibited. It follows that for an action to be permissible, at least one maxim which could possibly underlie the act must satisfy the categorical imperative. A required action, then, is one for which at least one maxim which could underlie the act fulfills the categorical imperative and for which all maxims which could possibly justify acting otherwise fail to satisfy the demands of the categorical imperative. Required and prohibited actions are, for Kant, "perfect duties."[27] The maxims which motivate them can, given the specifics of any situation, be derived directly from the form of the categorical imperative. These maxims have, as their end, the fulfillment of the moral law itself. When an action is merely permissible, maxims of this kind will be less instructive; universalizable maxims will be consistent with taking or forgoing the action. Our autonomous will is not, in these situations,

---

[25] Kant, *Groundwork of the Metaphysics of Morals*, 4:397.
[26] Kleingeld, *The Kantian Solution to the Trolley Problem Puzzle*
[27] Kant, *The Metaphysics of Morals*, 6:240

trapped like Buridan's mule. Instead, it must adopt for itself an externally conditioned end which can be consistently but only assertorically willed as universal.[28] Thus, in choosing my breakfast cereal, I might adopt the end of preserving my health. I could will a maxim of choosing healthily to be universal without contradiction but could equally accept maxims related to other ends such as frugality of sustainability. The pursuit of such values constitutes an "imperfect duty" and is consistent with autonomy but subordinate to the demands of strict perfect duty.[29] Because moral dilemmas claim to concern requirements or prohibitions, not permissible actions, they are mostly analyzed as pertaining to perfect duties.

On its face, this all does very little to curtail the worry that there might exist, consistent with the Kantian system, (1) a situation in which either an agent must choose[30] between actions each of which, considered in isolation from the other cannot be justified by any moral maxim or (2) a situation in which the principle of universality seems to require an agent to perform multiple incompatible actions. As Sartre noted in reference to his student's dilemma, there seems to be no universal rule favoring either familial or patriotic duty. The student certainly would not will that all French fighters in exile return home to tend to their loved ones. Nor would he blame any of them if they individually decided to do so. Kant would have us believe that no maxim can be willed as a universal law if it does not treat humanity as an end in itself rather than merely a

---

[28] Ibid., 6:224 & 452.
[29] Ibid., 6:411.
[30] In Kant's later practical philosophy (especially *The Metaphysics of Morals*), he develops a distinction between *Wille*, the autonomous volitional faculty of rational beings, and *Willkür*, an externally conditioned capacity to select from among various grounds and inclinations. *Willkür* is often translated as a 'capacity for choice.' (See, for example, Allison, *Morality and Freedom*). For my purposes however, it will be sufficient to understand an agent as choosing an action when, more precisely, the autonomous will determines that undertaking the action is a practical necessity.

means.[31] This clarification is of no help, however, when humanity is implicated on each side of a dilemma.

The strict architectonic of Kant's practical system does, at least, proscribe the advancement of certain objections. Kantian morality is tied entirely to human reason and epistemology. The possibility of a moral requirement of which we could not even be aware (an epistemological dilemma) is only compatible with a theory of morality which would base obligation on some outside heteronomous feature such as conformity to the will of a divine being or the ultimate, unknowable consequences of action. Because the Kantian system centers on rational evaluation of principles of volition, and because both reason and volition are autonomous features[32] within a person's intelligible character, epistemological dilemmas are impossible.

Additionally, because moral obligation is rationally derived from the practical postulate of a free will, the emotional residue of an action will have no bearing on its deontic status. Kant could contend, to Williams, that feelings of dissatisfaction which persist even if the categorical imperative is obeyed must necessarily fall into one of two categories. Emotions such as guilt or feelings of blame-worthiness are more correctly considered expressions of moral culpability. If a person has followed the dictates of moral reason and if they agree that others would have been

---

[31] Kant, *Groundwork of the Metaphysics of Morals*, 4:429.
Kant presents the law to treat humanity always as an end in itself as an alternative iteration of the categorical imperative. There seems to be a sense in which Kant believed that these two formulations of moral law were interderivable (*Groundwork* 4:431). There at least seems to be some *prima facie* validity to the notion that because humanity is, for Kant, the conjunction of rationality and autonomy, those very same structures around which the moral law is formulated, respect for the moral law and regard for humanity are equivalent. However, elucidating a proof of this would constitute a lengthy paper in itself. Instead, it is sufficient to say that to universalize any maxim that treats humanity as a mere contingent means would fail to regard even the actor as an autonomous will and, therefore, result in a practical contradiction.
[32] In fact, it might even make more sense to say that they are the features of autonomy.

right to do as they did, then such feelings cannot be appropriate. Alternatively, feelings such as regret or shame, which may very well be appropriate, have no relation to moral law or a person's intelligible character. They are externally conditioned. A moral action could result in negative emotions. If I realize that fulfilling a promise necessitates missing a much-anticipated event, I may be sad to miss the event but, nevertheless, fulfill my moral duty. Alternatively, I may find that the event was less important to me than I anticipated. The contrast can be easily made by the following distinction: If the categorical imperative is action-guiding in all morally relevant scenarios, it would follow that, because feelings of guilt implicate morality, I could, on principle, make decisions in such a way that I never appropriately feel guilty again. However, I could not conduct myself in such a way that it would never be appropriate for me to be regretful or sad. Situations wherein grief is called for, such as the death of a close friend, are not within my control. (There may, however, exist an imperfect duty to cultivate a feeling of regret when I have unavoidably harmed someone; doing so ensures that I will continue to be careful and deliberate when the happiness of others is at stake.) For a more comprehensive consideration of a Kantian approach to emotional residue, see Thomas E. Hill, whose treatment of the subject I take to be largely compatible with the outline of Kantian meta-ethics above and with arguments I develop in section 3.[33]

This response to the objection raised by appeals to emotional residue is relatively uncontroversial as an interpretation of Kant's practical project. It follows fairly straightforwardly from the condition that the moral law should be necessary and apodictic in form as opposed to contingent and assertoric. Additionally, there is ample support for it in the *Critique of Practical*

[33] Hill, Thomas E., Jr. "Moral Dilemmas, Gaps, Residues: A Kantian Perspective" (In *Moral Dilemmas and Moral Theory*, edited by H. E. Mason, 167-98. New York, NY: Oxford University Press, 1996).

Reason and in the *Groundwork*, where Kant asserts that "The moral worth of an action does not lie in the effect expected by it"[34] among which, he counts the condition of happiness. Nor does this worth, "lie in any principle of action that needs to borrow its motive"[35] from such heteronomous effects. However, a more general proof that Kantian ethics precludes normative dilemmas and that it conforms to typical moral intuitions will be more disputed. Kant seems to suggest both, at varying points but never provides an affirmative demonstration of either

III.

In the introduction to *The Metaphysics of Morals*, Kant provides, in a short paragraph, his least oblique statements about the possibility of conflicting moral obligations:

> A *conflict of duties* would be a relation between them in which one of them would cancel the other (wholly or in part). – But since duty and obligation are concepts that express the objective practical *necessity* of certain actions and two rules opposed to each other cannot be necessary at the same time, if it is a duty to act in accordance with one rule, to act in accordance with the opposite rule is not a duty but even contrary to duty; so a *collision of duties* and obligations is inconceivable. However, a subject may have, in a rule he prescribes to himself, two *grounds* of obligation, one or the other of which is not sufficient to put him under obligation, so that one of them is not a duty. – When two such grounds conflict with each

[34] Kant, *Groundwork of the Metaphysics of Morals*, 4:401.
[35] Ibid.

> other, practical philosophy says, not that the stronger obligation takes precedence but that the stronger *ground of obligation* prevails.[1]

What seems to emerge from this is that Kant believes that a normative moral dilemma – a situation in which a person is obligated to undertake multiple, mutually exclusive actions – could not possibly arise within the constraints of his ethical framework. Unfortunately, Kant fails to follow up on this assertion anywhere else in his practical philosophy. It is left to his readers and commentators to reconstruct the logic that could have led Kant to dismiss the possibility of a conflict of duties.

Jens Timmermann provides a superlative explication of this passage which elucidates its origins in the structure of Kantian ethics.[2] The categorical imperative can confer obligation primarily through the construction of perfect duties which are strict and juridical. Perfect duties are derived directly from the structure of the autonomous will and must hold as objective practical necessities. They have, as their grounds, the fulfillment of the moral law itself. In *The Metaphysics of Morals*, these perfect duties are equated with "duties of right."[3] One such duty cannot be in conflict with another because they are not independently derived. The categorical imperative is not like a list of virtues or duties wherein one item might conflict with another; it is, instead, a system of practical logic proceeding from the practical postulate of autonomous freedom and terminating in moral obligation. Whereas in a system of independently derived virtues, it might be necessary to ask whether one virtue takes precedence over another, the categorical imperative precludes this type of balancing. The categorical imperative simply cannot

---

[1] Kant, *The Metaphysics of Morals*, 6:224.
[2] Timmermann, Jens, *Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory* (Archiv für Geschichte der Philosophie 95, no. 1, 2013), 40-44.
[3] Kant, *The Metaphysics of Morals*, 6:229–232.

require two perfect and contradictory duties in the same situation in the same way that a mathematical function cannot produce two outputs from the same data. Two actions cannot each be universal and necessary in the way perfect duties must be for all rational actors.[4]

The compelling strength of a perfect duty is such that, in any case where one is implicated, it takes precedence over any other practical considerations. However, in the absence of these juridical duties, the categorical imperative to select a maxim which could be universalized continues to oblige. In such a situation, the categorical imperative confers obligation in the form of imperfect duties, identified in *The Metaphysics of Morals* with "duties of virtue."[5] Timmermann points us to a passage in the *Critique of Practical Reason* in which, "such law as provides merely a *ground* of obligation is distinguished from that which is in fact *obligatory*."[6] The ground of obligation for an imperfect duty cannot be, as it was for perfect duties, the simple fulfillment of the moral law. Instead the ground is exactly that end which the autonomous will is allowed to adopt for itself in cases in which the categorical imperative is not formally action-guiding. Inasmuch as these ends are taken to be subjectively universal and do not, therefore, violate the form of moral law, they are virtues. The will could adopt, as a conditional end, such endeavors as pursuing health, education, beneficence, or simply happiness. These grounds of obligation may conflict with each other[7] or two mutually exclusive actions may share a common ground but, as Timmermann makes clear, this collision must resolve into a

---

[4] Timmermann, *Kantian Dilemmas*, 42.
[5] Kant, *The Metaphysics of Morals*, 6:379–384.
[6] Kant, *Critique of Practical Reason*,
Here Kant should be understood as using the term "law" to refer to the statement of a duty such as "do not steal" or "develop one's talents" as opposed to, as he often does, using it to designate the universally applicable law of the categorical imperative.
[7] They may conflict with each other in practice. At the theoretical level, Timmermann suggests, all virtues must be taken to be consistent. (Timmerman, *Kantian Dilemmas*, 47.)

single logically consistent and morally necessary obligation. Anything less would be no obligation at all. Timmermann explicitly links this to the principle that "ought" implies "can." If it is not possible to realize all the grounds of obligation in our actions, then there can be no obligation to do so.[8]

This is a perfectly adequate logical proof that obligations cannot conflict in the Kantian moral system. Unfortunately, this formal explanation offers very little workable guidance regarding what course of action to take when grounds of obligation collide. Kant tells us that the "stronger ground of obligation prevails."[9] However, because such grounds are chosen conditionally, there does not seem to be any practically objective criteria by which to determine which is stronger. Indeed, notes taken by students of Kant indicate that he was very much a man of his time in assigning precedence to some conditional ends over others.[10] No argument can be drawn from the categorical imperative or from the strictly formal structure of Kantian morality to suggest that any ground of obligation is stronger than another or even, as Kant himself seems to have believed, that multiple grounds of obligation are stronger than any single ground.[11] Although it is clear that only one imperfect duty can emerge, it is continuously unclear in examples which it should be.

More concerning still, although it seems to be a logical truth that perfect duties – each of which may be motivated purely by respect to the moral law – cannot conflict, it is far from plain how this could be the case in actuality. Timmerman, following thinkers such as Onora O'Neill and Alan Donagan, argues that perfect duties (of right) are negative duties and that an infinite

---

[8] Ibid., 61.
[9] Kant, *The Metaphysics of Morals*, 6:224.
[10] Timmermann, *Kantian Dilemmas*, 52, 53&55.
[11] Ibid., 52.

number of negative duties are compatible because they can be fulfilled by non-action.[12] This explanation is insufficient for more than one reason. First, it must be apparent that duties of right can be either positive (the duty to fulfill a promise) or negative (the duty not to steal) as can duties of virtue (giving to charity is a positive imperfect duty while refraining from cruelty to animals is negative). Additionally, while it may be true at the theoretical level that negative rights do not conflict, it is certainly not true at the empirical level. If my brakes fail and I must choose between my car stopping in the bushes of one neighbor or another, the right of one of my neighbors not to have their property destroyed will be infringed upon. It seems at least intuitively true that perfect duties can clash or even that the same duty can contradict itself. If I have two debts but only enough money to repay one, what am I to make of the Kantian admonishment that there is a perfect duty to square one's accounts? Even if Timmerman can contend that these examples are only apparent dilemmas, he would need to demonstrate their normative resolution. Logical proofs such as the one Timmermann supplies are of scarce help to Sartre's student as he agonizes over whether or not to abscond to London. They instruct him simply that, of his options, only one could possibly be required.

As we struggle to make these abstract arguments concrete, it will be useful to have a collection of examples upon which to apply the Kantian principles that materialize. Fortuitously, Judith Jarvis Thomson, building on the work of Philippa Foot, has provided a versatile set of thought experiments that can be adapted to illustrate the intricacies and shortcomings of these principles. These "trolley cases," besides serving as examples to demonstrate how Kantian practical philosophy can resolve purported instances of normative dilemmas, are also archetypal descriptive dilemmas. It has been well documented in Thomson's work that for almost any

[12] Ibid., 45.

normative principle of action, a trolley case can be imagined for which the principle will recommend a counterintuitive action. The "trolley problem"[13] is to devise a system of ethics in which each trolley case is resolved to the satisfaction of most people's moral intuition. In order to acquaint ourselves with the difficulties of the trolley problem, it will be worthwhile to examine briefly some normative principles and describe their shortcomings both in resolving the trolley problem and in terms of Kantian practical philosophy.

Foot, in the earliest use of a trolley problem, proposes that a principle that condemns *doing* harm more stringently than *allowing* harm might resolve a variety of moral dilemmas. She compares the situation of a judge who can only stop a riot and the death of five people by condemning a single innocent man to death with the plight of a trolley driver who must choose between steering his vehicle either down a track on which five men are working and will be killed by the vehicle or another on which only one worker is present.[14] Foot argues that, the magistrate must choose between doing harm to one or allowing harm to five and, as per her principle, ought to allow harm rather than initiating it. In contrast, the driver will do harm regardless of who his trolley strikes and it is permissible (or even possibly required) for him to harm a lesser number of people rather than a greater.[15] The advice given by Foot to both the judge and the driver is at least intuitively valid. However, her normative principle is neither supported by Kantian practical reason nor descriptively sufficient in other thought experiments.

This distinction between doing and allowing harm is closely related to the differentiation made by Timmermann between negative and positive duties and is insufficient for similar

---

[13] Thomson, Judith Jarvis, *The Trolley Problem*, (The Yale Law Journal 94, no. 6, May 1985), 1396.
[14] Foot, *The Problem of Abortion*, 9.
[15] Ibid.

reasons. I have already noted that Kant assigns no normative priority to negative duties. Foot's principle, however, rests on even more tenuous grounds: It relies on the commonsense difference between action and inaction. There is no reason to draw such a distinction in Kantian philosophy. Volitional inaction (as opposed to the inaction that occurs when a person is asleep, unaware, or frozen by shock or indecision) must, like volitional action (not reflexive action), be motivated by a first-order maxim. There is no evidence and no strong argument to be made that a maxim justifying inaction would be subject to a separate analysis under the categorical imperative than would a maxim resulting in action. Furthermore, it is unclear that action and inaction are discrete practical designations. If I am lowering a heavy chest of drawers out of a third story window and see my neighbor walking underneath just as I feel my grip beginning to give, do I commit an action or an inaction by failing to reassert my grip and allowing the furniture to fall? There are compelling arguments asserting either conclusion but, as long as I choose to conduct myself in such a manner that the chest of drawers falls onto my neighbor, the division is largely semantic. It cannot be that the determinations of the fully formal and internally conditioned Kantian practical system would turn on a principle as indistinct as this. And, indeed, it does not.

These flaws with the principle of allowing rather than causing harm become evident in descriptive ethics with the introduction of another trolley case. In her seminal essay on the trolley problem, Thomson adjusts Foot's case in two critical ways. First, the person with the power to steer the trolley is, instead of a driver, a bystander at a switch. Second, the tracks are arranged so that if the switch is not flipped, the trolley will kill the five workmen whereas, if the switch is thrown, it will kill the one.[16] These modifications make it clear that to pull the switch would cause the death of the one whereas to do nothing would simply be to allow the deaths of the five.

[16] Thomson, *The Trolley Problem*, 1397.

However, intuition insists that this bystander iteration [Figure 1] is scarcely removed from the original case involving the driver and that, as before, it is at least permissible to drive over one person to save five. Foot's principle, then cannot match the intuitive differentiation between trolley cases and the case of the judge.

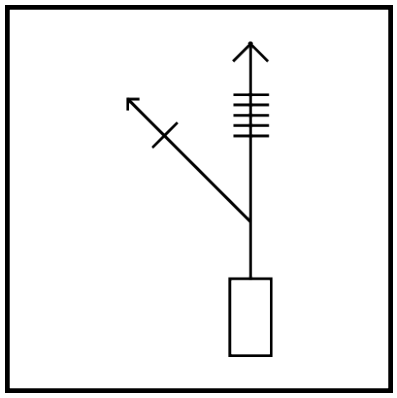

Figure 1: Bystander Iteration[17]

Thomson points to a different normative principle that, at first, seems more promising. In a reference to Kantian ethics, she notes that the judge, it might be argued, impermissibly treats the single innocent as a means to the end of stopping the riot.[18] The bystander, alternatively, only foresees that the one workman will die and does not *use* his death as a necessary development in saving five lives. Foot makes the similar point that while the judge "needs the death of the innocent man for his (good) purposes,"[19] the subject of the trolley case would not, if the one worker somehow survived, "brain him with a crowbar."[20] This principle is supported by another iteration of the trolley problem, the footbridge iteration, [Figure 2] wherein the agent is observing the oncoming trolley from a footbridge over the tracks. The trolley is rushing towards five workers and cannot be diverted but can only be stopped by throwing a heavy man

[17] Based on an image created by Judith Jarvis Thomson. (Thomson, *The Trolley Problem*, 1402.)
[18] Ibid., 1401.
[19] Foot, *The Problem of Abortion*, 10.
[20] Ibid.

(conveniently standing on the bridge) onto the tracks. The trolley will strike this man and, while he will be killed, his mass will stop its progress.[21] People generally agree that we should not push the man off the footbridge.

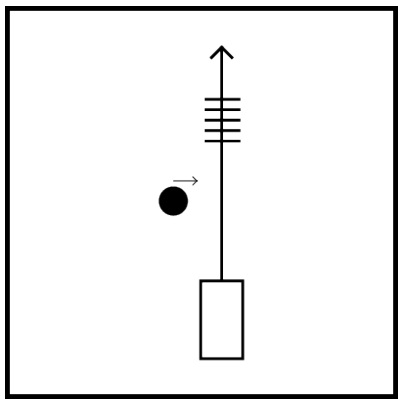

Figure 2: Footbridge Iteration

In this principle, Thomson weds the categorical imperative with the Catholic "doctrine of double effect" whereby it is sometimes permissible to take an action which has a foreseeable harm as a *consequence* but never an action in which the harm foreseen is *intended* to attain some desired result.[22] In order to differentiate between various thought experiments, it must be assumed that we only treat a person as a means if we use them as an instrument to achieve some goal. It is far from obvious that this is what Kant intended by his dictum "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."[23] Kant's principal requirement, we can see, is always to treat people as ends in themselves; treating a person as a mere means is presented primarily as occurring whenever we fail to treat people as an end. To treat a person as a means, for Kant, is not (or, at least, is not exclusively) to use them instrumentally. There is no indication in Kant's writings that it might be

[21] Thomson, *The Trolley Problem*, 1409.
[22] Foot, *The Problem of Abortion*, 6.
[23] Kant, *Groundwork of the Metaphysics of Morals*, 4:429.

possible to treat a person as an end while killing them. Additionally, there seems to be no reason that a maxim which precipitates killing as a necessary consequence should be appraised differently than one which prompts it instrumentally.

This normative principle is further belied by a trolley case known as the loop variant [Figure 3]. In this thought experiment, rather than continuing to diverge from the main track, the fork of the trolley reconnects to form a loop. If the trolley continues along its predetermined path, it will, as before, strike five workmen, killing them. If diverted, the trolley would, if unimpeded, follow the curvature of the loop until finally hitting the workers as before. Blocking its path, however, is a single heavy man.[24] If we accept the identification of "treating as a means" with "using instrumentally," then the heavy man would be impermissibly used in this example. However, intuition maintains that it is as permissible to pull the lever in this situation as it was in the bystander iteration and, so, this principle cannot match intuition. Even worse, in another example, this principle could lead to an intuitive absurdity. Suppose that the trolley, while advancing towards fire workers, could be turned onto either of two adjoining tracks [Figure 4]. On the first, there are three people, all of whom will be killed if the trolley is diverted onto their track. On the second, is a heavy man whose bulk will halt the trolley and, further down the track, two other people. The doctrine of double effect, if applied, would seem to insist that, while it is permissible to turn the trolley onto the first sidetrack, killing three to save five, it is impermissible to direct it onto the second where only the heavy man will be harmed.

[24] Thomson, *The Trolley Problem*, 1402.

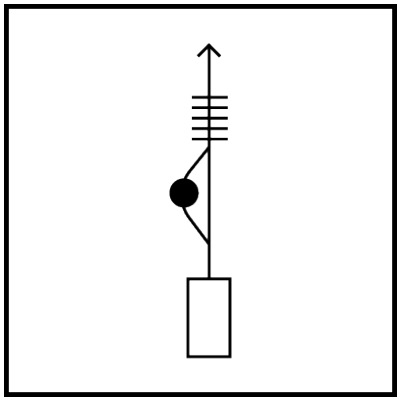

Figure 3: Loop Iteration[25]

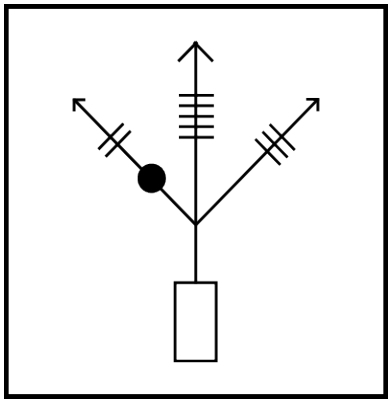

Figure 4: An Absurdity Caused by the Doctrine of Double Effect

After being inundated with these thought experiments, it is entirely possible that a person's moral intuition would begin to shift. It might be tempting to reevaluate the circumstances of the judge faced with an impending riot and adopt the consequentialist principle that the correct course of action is simply whichever saves the most lives. Such a proposition immediately encounters objections both normatively, from practical theory, and descriptively, from the articulation of yet more thought experiments. No normative principle based in consequentialism could ever comply with the Kantian structure of morality. While Kant declares that a moral action must be autonomously willed, consequentialism would determine the principle of volition by analyzing assertoric facts about the world. Nothing could result except heteronomy. Descriptively, consequentialism recommends actions even less appropriate than

[25] Based on an image created by Judith Jarvis Thomson (Thomson, *The Trolley Problem*, 1402.) In Thomson's original figure, the track formed a closed loop so that, if the heavy man were not present when the switch was flipped, the five workers would eventually be struck by the trolley coming in the opposite direction from that in which it originally approached them. Alternatively, if the five men were not present (or if their combined weight proved insufficient to stop the trolley), then the heavy man would be in danger of being struck even if the switch was not touched. The concern could then arise that, even if no action were taken, there might be a danger of using the five workers as instrumental means to protect the heavy man. This concern is not discussed by Thomson and I, considering it an unneeded complication, constructed my figure to avoid it.

other normative principles. In Thomson's transplant scenario, a doctor has five patients each in dire need of organ transplants. If they do not receive them, they will surely die. However, no organs are available. On the final day that operations would be possible, a healthy man comes into the doctor's office. In the course of a routine examination, the doctor realizes that this person is a perfect match for his dying patients.[26] The transplant scenario is meant to resemble the footbridge and loop iterations of the trolley problem wherein taking the life of one person is a necessary component of a process that saves the lives of five others. Consequentialism, if applied to this situation, would advocate that the doctor give the healthy patient a lethal dose of morphine, killing him painlessly, so that his organs may save the lives of the terminally ill five. Moral intuition, it is hardly necessary to say, disagrees strongly.

Pauline Kleingeld suggests that no normative principle that directly constrains action can resolve the trolley problem.[27] Kantian ethics, however, might be able to. The categorical imperative is not a first-order normative principle. Instead, it is a "moral meta-principle"[28] that moderates principles of action. The normative principles discussed above can most easily be understood in the Kantian framework as attempting to enunciate broad obligations (perfect duties, in Kant's terminology). As already explained in section 2, an obligation arises in Kantian practical philosophy either, negatively, when all maxims that could justify an action fail to satisfy the demands of the categorical imperative or, positively, when no maxim that could underlie taking a different course of action fulfills the categorical imperative. The categorical imperative, therefore, could lay bare a distinction between the loop iteration, in which action intuitively seems permissible, and the transplant scenario, in which inaction seems to be

[26] Ibid., 1396.
[27] Kleingeld, *The Kantian Solution to the Trolley Problem Puzzle*.
[28] Ibid.

required. A critical insight of Kleingeld's paper is that, if we can show that killing the healthy patient in the transplant scenario is prohibited, it will be a sufficient contrast to satisfy intuition if we can determine that pulling the lever in the bystander iteration is at least permissible.[29] This soft solution to descriptive dilemmas requires, in most cases, only that the action which intuition prescribes be *permissible*. Only in some few cases will an action be so outrageous that intuition will demand its unqualified prohibition.

Unfortunately, Kleingeld's demonstration of this distinction is not entirely satisfying. She begins by imagining a person, "Tili," who adopts as her maxim the principle that, "When forced to choose, [she will] save more lives rather than fewer, even if this requires using another person as a mere means to that end."[30] Kleingeld invites us to agree that this maxim, so similar to the normative principle of a consequentialist, would be "reprehensible"[31] to a Kantian; the maxim explicitly contains the forbidden estimation of a person as a mere means. In each of the trolley cases as well as in the transplant scenario, this maxim would lead Tili to sacrifice one person in order to save five. It follows from this only that Tili acts from immoral *grounds*, not that her *actions* themselves are prohibited. Enter Manuel,[32] who adopts as his maxim that he will, "When forced to choose, save more lives rather than fewer, other things being equal."[33] In examining this "other things being equal," we might begin to worry that this proof will be unpersuasive. Other things being equal by what criteria? If the criterion is a moral one, then the inclusion of this phrase in a maxim begs the question of the maxim's universalizability. In Kantian ethics, it

---

[29] Ibid.
[30] Ibid.
[31] Ibid.
[32] Manuel's name seems clearly to be a send-up to Immanuel Kant. I have long suspected, but have been unable to confirm, that the name "Tili" may be a reference to teleological ethics – the Greek appellation for consequentialist morality.
[33] Kleingeld, *The Kantian Solution to the Trolley Problem Puzzle*.

is moral to act upon a maxim only if the maxim could, at the same time, be willed to be a universal law. The provision that a maxim will only be acted upon if it is moral to do so is, therefore, only a restatement of the practical necessity of the moral law and is of no help at all in evaluating the maxim in question. When Kleingeld places Manuel at a trolley switch, the question is not whether his maxim is acceptable – this has already been established. Instead, we are left to ask whether *other things are equal*. The criteria for morality, then, can be seen to lie not in Manuel's maxim but in the particularities of his situation. Kleingeld has fallen back on a first-order normative principle, abandoning the uniquely Kantian nature of her proof.

When Manuel's maxim is applied to the bystander iteration, Kleingeld finds that other factors are equal and that he may proceed to pull the switch, killing one worker to save five.[34] However, when Manuel stands on the footbridge and contemplates pushing the heavy man onto the tracks, Kleingeld argues that other considerations are not equal at all. Her proof relies on the same equivocation between treating as a mere means and treating as an instrument that has already been untangled above.[35] She states that "When the agent assesses the situation by gauging the man's weight, his impact on the trolley, the agent's own ability to push him over the edge, and so on, the heavy man becomes 'the one' because and only because he could serve as a tool to stop the trolley."[36] This, and not the categorical evaluation of Manuel's maxim, is the underlying principle by which Kleingeld differentiates between permissible and prohibited actions and through which she argues that no universalizable maxim could result in pushing the heavy man off the bridge.

---

[34] Ibid.
[35] Professor Kleingeld, who I have corresponded with, recognizes the insufficiencies of her conception of treating as a mere means and is reworking her paper to include a more precise development of this requirement.
[36] Ibid.

Based upon the divisions made by this normative principle, Kleingeld reasonably identifies the transplant scenario as being of a kind with the footbridge trolley iteration. However, she inexplicably couples the loop iteration with the bystander iteration.[37] Even if we acquiesce to the linking of "treating as a mere means" to "using instrumentally", this does not follow. In the looping version of the trolley scenario, as in the footbridge iteration, the agent must gauge the heavy man's weight and his ability to halt the trolley. In each case, the heavy man becomes "the one" only because he could be utilized as an instrument to save the five workers. It is unclear, then, why Kleingeld insists that, while it is permissible to flip the switch in the loop iteration, it is prohibited to push the heavy man off the footbridge. In scrutinizing the loop iteration, she argues that "whether a person is used as a mere means depends not on the other person's causal role but on the agent's maxim."[38] This is correct but is both inconsistent with her earlier arguments and insufficient to differentiate between pulling the lever in the loop iteration and pushing the heavy man off the footbridge. When applying Manuel's maxim to the footbridge iteration, Kleingeld relied on considerations about the situation other than the maxim. None of these considerations are different in the loop iteration. To argue that in the loop iteration, "Manuel's maxim is applicable, and on the Kantian understanding of the term, Manuel is not guilty of using the one workman as a mere means,"[39] is entirely circular: If Manuel's maxim applies, it must be that considerations other than the number of lives are equal and if all things are equal, then killing the heavy man on the loop is not treating him as a mere means; however, it is only if diverting the trolley towards the heavy man is not treating him instrumentally that Manuel's maxim applies at all.

---

[37] Ibid.
[38] Ibid.
[39] Ibid.

Kleingeld does not provide a Kantian solution to the trolley problem but her key insights will be essential to discovering one. She elucidates that the categorical imperative derives its action-guidance from the evaluation of maxims, not from any normative principle. Additionally, she explains that strong dichotomies are not required to satisfy intuitive distinctions. When it is recognized that no first-order normative principle can resolve the trolley problem, this set of moral dilemmas begins to appear far more intractable. Because it is not possible to incorporate the death of a person as even an intermediate end of a maxim while, at the same time, treating that person as an end and not merely as a means, there appears to be no purely formal method of resolving any of the trolley cases. It resembles the predicament of Sartre's student who realized that to stay with his mother would be to treat his compatriots as a mere means while to leave for London would treat his mother in the same way.

Here, it is finally possible to fill the ambiguity left by Timmermann in the gap between theory and application. Timmermann provides a rigorous proof, based in practical reason, that perfect duties cannot collide. Perfect duties are defined by their necessitation from the pure formalism of the moral law and two mutually exclusive actions cannot be necessary simultaneously. However, perfect duties at least apparently conflict with each other in normative moral dilemmas. The only solution to this incongruity is that if two actions, each of which is a perfect duty when considered in isolation from the other, are mutually exclusive in a situation, then, in that circumstance, neither of them is a perfect duty. If this seems to run counter to common sense, it is only because we make a mistake by imagining that perfect duties are developed from the moral law in abstraction from particular situations. Instead, as Kleingeld shows us, the categorical imperative determines obligation from the appraisal of first-order maxims, which are always directly volitional and particular to a situation. If, given a state of

affairs, no maxim is available which is universalizable or if there are multiple maxims between which the will cannot distinguish on the basis of pure practical formalism, then no perfect duty arises at all.

Instead, the will must adopt a conditional end for itself. If the situation is one which lends itself to the pursuit of multiple virtues, then these assertoric ends are competing grounds of imperfect duties. Trolley cases are, I believe, circumstances that clearly allow multiple grounds of obligation. There is an imperfect duty not to interfere in others' lives in a negative way – not to be maleficent. The grounds of this duty are clearly present in all trolley cases; it would be promoted by not turning the trolley onto the one in order to save the five. Alternatively, there is an imperfect duty of beneficence – to interfere in the lives of others in a helpful way – which would be advanced by saving the lives of the five workers. Similarly, Sartre's student has grounds to fulfill the imperfect duty of caring for one's mother and competing grounds for fulfilling the imperfect duty of fighting against tyranny. In these situations there is no method of differentiation based in pure practical reason. An obligation is present but only a single obligation, not two conflicting ones. The obligation is this: Act according to an end that is at least conditionally universal. Act, in other words, according to a virtue. This obligation to act from universalizability is an overarching perfect duty which is equivalent to and coextensive with the categorical imperative and, like the categorical imperative, is compatible with multiple actions depending on which virtue (or set of virtues) is subjectively deemed strongest.

A result of these considerations that may at first seem troubling is that they would group predicaments such as the trolley cases in a class with decisions about what cereal to eat instead of with considerations such as whether to steal or cheat. This strains the bounds of conventional thinking about morality. In order to elucidate this provocative classification, the pertinent

question is this: On what foundation are obligations constructed? In a perspicacious essay, Foot argues that, even though consequentialism is widely denounced as a normative theory, its premises continue to influence our moral convictions. Foot asserts, and I agree, that in order to think consistently about moral matters, it is necessary to ablate the remnants of teleological thinking from our ethical considerations.[40] This shift is especially critical in Kantian ethics because Kant's practical project and consequentialism follow from antithetical premises. Whereas obligations in Kantian morality are developed from the formal and autonomously conditioned structure of the categorical imperative, consequentialist obligations arise from the consideration of assertoric and purely heteronomous facts about the world. We expect that matters of life and death must entail some moral obligation because we imagine that morality should concern itself with outcomes and that greater stakes entail a greater moral commitment. When we make the Kantian turn to an ethics of autonomy, we see that this cannot be the case and that obligation only follows from the practical necessity established by the will's self-assigned law.

It follows that, in all trolley cases as so far described, it is morally permissible either to sacrifice one life while saving five or to let the five die while sparing the one. Although the virtues promoting these actions, their maxims, and the actions themselves are divergent, there are no morally relevant criteria by which to distinguish them. Timmerman argues, for this reason, that there is no possibility of supererogation in Kantian ethics. "There is no such thing as an action that is '[morally] good but not required.'"[41] It might be more accurate to say that there is no value-neutral, universal standard of supererogation consistent with Kant's practical

[40] Foot, Philippa, *Utilitarianism and the Virtues* (Mind 94 no. 374, 1985), 196–198.
[41] Timmermann, *Kant's Puzzling Ethics of Maxims*, 48.

philosophy. In other words, supererogation is beyond the bounds of the moral structure. This is, of course, exactly what was implied by the term "supererogatory". Thomson provides an example that will be illuminating in this regard.[42] Placing a bystander again at a switch in the rails with a trolley advancing towards five workers, she asks us to imagine that the switch can be flipped in either of two directions. If it is turned in the first direction, the trolley will be diverted and strike a sixth worker, sparing the original five. If the switch is left alone, the five workers will be run over. If, instead, the switch is flipped in the opposite direction, the trolley will turn onto a track where the bystander herself stands and she will be killed [Figure 5]. Thomson's argument is that, if the person would not be willing to sacrifice her own life to save the five, she should not be willing to sacrifice the life of the one worker.[43] A Kantian analysis illustrates a very different point. From the perspective of Kant's ethical project, it is permissible to take any of the three actions and there is no objective practical procedure by which to promote one over

---

[42] Thomson, Judith Jarvis, *Turning the Trolley* (Philosophy and Public Affairs 36, no. 4, 2008), 364.

[43] Thomson's own solution to the trolley problem involves appreciation for two normative principles. The first is a distinction that I have already addressed between violating a negative right by doing harm and merely allowing a harm. The second, which has not yet been considered, is a principle of "'distributive exemption,' which permits arranging that something that will do harm anyway shall be better distributed" so that it does harm to fewer people. (Thomson, *The Trolley Problem*, 1408). Although Thomson advocates for this principle on intuitive grounds, she admits that "I do not find it clear why there should be an exemption for, and only for, making a burden which is descending onto five descend, instead, onto one." (Ibid.) Applying these normative rules, Thomson argues that if the ceiling of a room were about to collapse, killing its five inhabitants, and if the only way to save them was to turn on a mechanical support mechanism which would pump lethal gas into an adjoining room with one occupant, one would be required to allow the ceiling to collapse. Because the harm from the ceiling and the harm from the gas are distinct, her distributive principle does not take effect. Alternatively, if a heating system was pumping poisoned gas into a room with five occupants, Thomson argues that it would be permissible to divert the gas into a room with only one inhabitant. From a Kantian perspective, there cannot be any support for this proposition which would base moral obligation on conditional as opposed to volitional distinctions. The Kantian principles that I outline in my paper suggest that while it would be permissible to allow the ceiling to fall and permissible to divert the gas, neither action is morally required.

the others.[44] However – within a culture, tradition, or personal system of values – sacrificing one's own life for the benefit of others could be seen as admirable or even heroic. This esteem is a form of value that has no foundation in pure practical reason and could therefore be considered supererogatory.

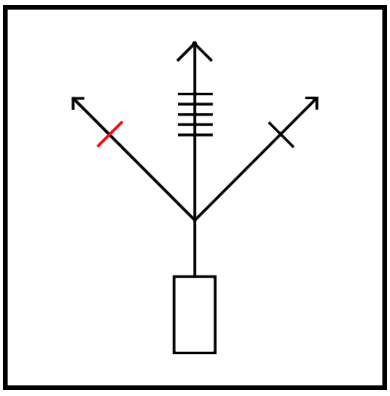

Figure 5: Bystander's Three Options

Must it also be true, however, that in all circumstances that balance life against life (or, more broadly, humanity against humanity) that it is permissible to settle on either side? In the footbridge iteration, if our intuition balks at the idea of pushing the heavy man off the bridge, then this is a reasonable extra-moral criterion upon which to base our action and it is entirely permissible to adhere to this intuitive judgement. In the transplant scenario, however, intuition rejects even the idea that it might be permissible for the doctor to kill his healthy patient in order to save five others. Does the Kantian project insist that, in this situation, there is no maxim that is given necessity by force of the moral law itself? I do not think it does.

---

[44] Although, in the *Introduction to the Doctrine of Right* from *The Metaphysics of Morals*, Kant makes an empirical distinction between duty to oneself and duty to others, there is no indication that, at the theoretical level, the imperative to treat humanity in one's own person as an end is distinct from or stronger than the requirement to treat humanity in other's always as an end. (Kant, *The Metaphysics of Morals*, 6:239.) As such, there is no reason to think that any particular action in this example would be anything more than permissible.

In the transplant scenario, a moral obligation is implicated in the form of a perfect duty not to take the life of the healthy patient. In order to understand why, it will be necessary to dispense with another piece of consequentialist reasoning that persists in our moral considerations: the notion that our volitional deliberations should take into account hypothetical concerns and project our will into future circumstances. The pure formalism of the categorical imperative limits its scope. It cannot deal with any factors that are contingent, hypothetical, or consequential and is confined to a single practical situation – to the selection of an immediate maxim.

The transplant scenario was previously grouped with the footbridge and loop trolley iterations only because various philosophers focused on the instrumentality involved in the situation. It will be far more profitable to focus on the situation's temporality. In each of the trolley problems provided by Thomson, only one decision is involved and only a single action results. Either the switch is flipped, or it is not. Redirecting the trolley is necessary and sufficient for both saving the five and killing the one. However, in the transplant scenario, we should notice that killing the healthy patient, while it is necessary in order to save five lives, is not sufficient. Something else would have to occur in order for the transplant patients to live. Particularly, a separate and future action would have to be taken. The categorical imperative is not equipped to deal with such subjunctives. It evaluates maxims, which are immediate principles of actions, not the future possibilities that are the consequences of proposed acts. So, in the transplant situation, the doctor could not adopt as his principle of action, "I will kill one person, so that I may later save five lives." Instead, with the hypothetical component removed,

his maxim would read, "I will kill one person, while simultaneously saving no one."[45] It is clear that this maxim does not satisfy the demands of the categorical imperative and that the doctor is prohibited from taking action.

By constructing a trolley case that shares this same important quality with the transplant iteration, it will be easy to demonstrate what distinguishes it from the bystander, footbridge, and loop trolley iterations. In this scenario, because the bystander will have to undertake two separate actions in order to save the five workers, we will require two trolleys and two levers. The first trolley will proceed down an empty track, endangering nobody. If the first lever is thrown, this benign trolley will be redirected onto a track where one worker will be killed. The second trolley is on a course to strike and kill five workers. By flipping the second switch, the bystander can move this second trolley onto the track where no one is working. However, due to a safety feature built into the switches, it is not possible to throw the second switch as long as the first trolley remains on the benign track. In order to save the five workers, it is first necessary to flip the first switch, killing the one [Figure 6]. If I am correct in asserting that Kantian ethics prohibits taking an action which causes harm to a person while simultaneously benefiting no one, then neither the doctor in the transplant scenario nor the bystander in this dual action iteration of the trolley problem may permissibly take the life of the one.[46] These situations are therefore adequately distinguished from other iterations of the trolley problem.

---

[45] It is not strictly the simultaneity of the results that matters here but, instead, the uniformity of the actions. This maxim could be equally stated as, "I will kill one person, while, *in the same action*, saving no one."

[46] It is important to note that this distinction applies in all situations in which hypothetical factors might come into play, not exclusively when there is a second subjunctive action on the part of the agent. So, in the case of the magistrate, it becomes clear that executing the one innocent man is by no means sufficient to saving the life of the five riot victims. It can only do so *if* the execution changes the volitional principles of the rioters who must, themselves, be taken to be autonomous. Similarly, in Foot's example of a tyrant who threatens to "torture five men if we

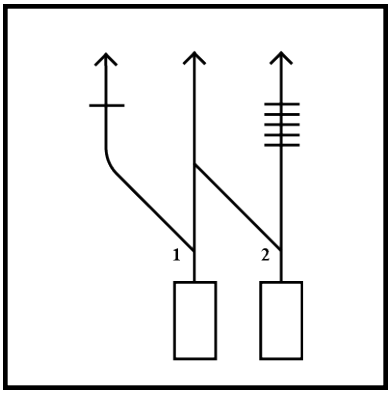

Figure 6: The Transplant Scenario Depicted as a Trolley Case with a Resolution from Practical Necessity

Kant's philosophy, then, is uniquely action-guiding in all circumstances in which moral obligation (perfect duty) is implicated. Inasmuch as there is a ubiquitous requirement to act from virtuous grounds, it is partially action-guiding in all situations. It resolves descriptive dilemmas at least to the extent that it is permissible to undertake those actions towards which moral intuition points. Additionally, although the proof at the end of section 2 demonstrating that emotional residue has no bearing on deontic status still holds, the distinction between obligations, which can never conflict, and grounds of obligations, which can, allows room for a modest connection between sentiment and practical reasoning. It is entirely sensible that a person would experience a feeling of displeasure or regret when they cannot act upon all of the grounds of imperfect duties that are present in a situation.

What counsel, then, is appropriate for Sartre's student? Kantianism requires him to select which virtue he will follow without any guidance or even the possibility of making an objectively correct decision. It is, in this way, compatible with Sartre's own guidance: "invent."[47] But, a Kantian would caution, do not invent without any boundaries. (Do not, for example, flee

---

ourselves would not torture one," it becomes clear that our action is neither necessary nor sufficient to saving the five.

[47] Sartre, *Existentialism Is a Humanism*, 46.

to Switzerland.) Instead, follow a course of action that is conditionally universalizable so that you could consistently lend your approval to others who do the same. As sympathetic observers to the student's plight, this advice is instructive to us as well. We should not, nor would we intuitively, accuse the student of acting out of accord with moral duty regardless of whether he stays in France with his mother or goes to London to fight.

Selected Bibliography

Allison, Henry E. "Morality and Freedom: Kant's Reciprocity Thesis." *The Philosophical Review* 95, no. 3 (July 1986): 393-425.

Donagan, Alan. *The Theory of Morality*. Chicago, IL: University of Chicago Press, 1979.

Foot, Philippa. "The Problem of Abortion and the Doctrine of the Double Effect." *Oxford Review* 5 (1967): 5-15.

Foot, Philippa. "Utilitarianism and the Virtues." *Mind* 94 no. 374 (1985):196-209.

Hill, Thomas E., Jr. "Moral Dilemmas, Gaps, Residues: A Kantian Perspective." In *Moral Dilemmas and Moral Theory*, edited by H. E. Mason, 167-98. New York, NY: Oxford University Press, 1996.

Kant, Immanuel. *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen W. Wood. Cambridge, UK: Cambridge University Press, 1998.

Kant, Immanuel. *Practical Philosophy*. Edited and translated by Mary J. Gregor. Cambridge, UK: Cambridge University Press, 1996.

Kleingeld, Pauline. "The Kantian Solution to the Trolley Problem Puzzle." Paper presented at The American Philosophical Association Eastern Division One Hundred Tenth Annual Meeting, Baltimore, MD, December 28, 2013.

Sartre, Jean-Paul. *Existentialism Is a Humanism*. Translated by Carol Macomber. New Haven, CT: Yale University Press, 2007.

Thomson, Judith Jarvis. "The Trolley Problem." *The Yale Law Journal* 94, no. 6 (May 1985): 1395-415.

Thomson, Judith Jarvis. "Turning the Trolley." *Philosophy and Public Affairs* 36, no. 4 (2008): 359-74.

Timmermann, Jens. "Kantian Dilemmas? Moral Conflict in Kant's Ethical Theory." *Archiv für Geschichte der Philosophie* 95, no. 1 (2013): 36-64.

Timmermann, Jens. "Kant's Puzzling Ethics of Maxims." *The Harvard Review of Philosophy* 8 (2000): 39-52.

Williams, Bernard, and Atkinson, W. F. "Ethical Consistency." *Proceedings of the Aristotelian Society, Supplementary Volumes* 39 (1965): 103-38.