The roles of the moral and the political in the philosophies of Kant and Rawls

Saiming Wong

Follow this and additional works at: https://repository.hkbu.edu.hk/etd_oa

Recommended Citation
Wong, Saiming, "The roles of the moral and the political in the philosophies of Kant and Rawls" (2018). Open Access Theses and Dissertations. 593.
https://repository.hkbu.edu.hk/etd_oa/593

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at HKBU Institutional Repository. It has been accepted for inclusion in Open Access Theses and Dissertations by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.
DATE: November 27, 2018

STUDENT'S NAME: WONG Sai Ming

THESIS TITLE: The Roles of the Moral and the Political in the Philosophies of Kant and Rawls

This is to certify that the above student's thesis has been examined by the following panel members and has received full approval for acceptance in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairman: Prof Hjort Mette
Dean, Chair Professor, Faculty of Arts, HKBU

Internal Members: Dr Zhang Ellen Ying
Head, Associate Professor, Department of Religion and Philosophy, HKBU

Dr Chan Shing Bun Benedict
Assistant Professor, Department of Religion and Philosophy, HKBU

External Members: Prof Shell Meld Susan
Professor and Chair
Department of Political Science
Boston College

Prof Lau Chong-Fuk
Professor
Department of Philosophy
The Chinese University of Hong Kong

In-attendance: Prof Palmquist Stephen R
Professor, Department of Religion and Philosophy, HKBU

Issued by Graduate School, HKBU
The Roles of the Moral and the Political in the Philosophies of Kant and Rawls

WONG SAI MING

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Principal Supervisor:
Prof. Stephen PALMQUIST (Hong Kong Baptist University)
December 2018
DECLARATION

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at Hong Kong Baptist University, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

I have read the University’s current research ethics guidelines, and accept responsibility for the conduct of the procedures in accordance with the University’s Research Ethics Committee (REC). I have attempted to identify all the risks related to this research that may arise in conducting this research, obtained the relevant ethical and/or safety approval (where applicable), and acknowledged my obligations and the rights of the participants.

Signature: ________________________

Date: December 2018
Abstract

The primary focus of this dissertation is the problem of the roles of the moral and the political in political philosophy as exemplified in the philosophies of Kant and Rawls. The research question which intrigues me in the subject matter is whether and to what extent morality has a role in political philosophy. As I argue in this dissertation, while Kant’s political philosophy is grounded by his moral philosophy, Rawls holds the opposite view that a political conception of justice should not be derived from any specific moral doctrine. Their contrasting views are further complicated by the fact that Rawls is often regarded as a Kantian due to his partial assimilation of Kant in his theory of justice. A comparative study of their views on the roles of the moral and the political in philosophy is thus particularly instructive in answering the above research question. This dissertation therefore approaches the subject matter from four different angles. In the first chapter, I start with a holistic interpretive account of Kant’s moral and political philosophy that is quite different from those in the current literature. Not only do I argue that Kant’s moral philosophy is unmistakably an indispensable ground of his political philosophy, but I also argue for a positive duty in politics that is moral by nature. In the second chapter, I shift my focus to Rawls and examine his understanding of morality as reflected in his moral conception of the person in his political philosophy. I argue that the conception is the result of an intended reformulation of Kant’s notion of autonomy that is in turn based on an unintended misreading of the same. I go on to relate several weaknesses in Rawls’s theory to his understanding of morality and argue that their resolutions require an accurate understanding of the relationship between the moral and the political. The third
chapter is a Kantian appraisal of the four roles of political philosophy proposed by Rawls. It is relevant to the subject of this dissertation because the four roles are designed with a strict separation between the political and the moral in mind. If the four roles turn out to be defensible, it would amount to an important defense for such a separation. The fourth chapter offers an alternative for those who are more accustomed to the political than to the moral by proposing a new interpretive approach to Kant’s philosophy starting from the political and ending with the moral. By ending this dissertation with this alternative, I hope my research is not merely a comparative study but can also offer a new perspective for a more in-depth understanding of the relationship between the moral and the political in philosophy.
Acknowledgements

First and foremost, I must thank Hong Kong PhD Fellowship Scheme for keeping me alive when I spent all my time in the last three years philosophizing. I must thank my supervisor Professor Stephen Palmquist for all the invaluable advice and support. I cannot express my gratitude enough to him. Without him, the dissertation in its current form is not imaginable. Professor Lo Suet-kwan is not directly involved in this project, but she had been my teacher for a decade before my PhD study. Her passion and her knowledge equipped me both emotionally and intellectually in the years before the start of this project. A huge thank you goes to my wife, Samantha Yung, for her unfailing support throughout the years. I must thank her for putting up with me when I philosophize at the wrong time and when I do not philosophize at the right time. I do not understand why she loves a would-be philosopher (hopefully) like me, but it means everything to me. I must also thank my parents and my sisters for their support, although I did not take the time to explain to them what I have been writing all these years. Lastly, I must thank myself for bearing with me when I procrastinated and failed to perform. An important finding of this project turned out to be a deeper understanding of myself and this means a lot to me. Thanks for completing this project on time!
# Table of Contents

**Introduction**  
1

**I  Re-defining Political Philosophy: Identifying the Moral and the Political in Kant’s Philosophy**  
5  
1. *Introduction*  
5  
2. *A Brief Review of the Debate*  
7  
3. *The Basics of the New Interpretive Approach*  
14  
4. *The Moral Law as the Limiting Condition of the Freedom of Action*  
18  
5. *From the Moral Law to the Idea of a Public Will*  
27  
6. *The Concept of Public Will and the Two Kinds of Lawgiving*  
39  
7. *Positive and Negative Duties in Kant’s Political Philosophy*  
46  
8. *Closing Remarks*  
53

**II The Covert Subversion of Kant’s Account of Practical Reason in Rawls’s Understanding of Morality**  
56  
1. *Introduction*  
56  
2. *The Foundational Role of Rawls’s Understanding of Morality*  
58  
3. *Rawls’s Misinterpretation of Kant’s Autonomy of the Will*  
63  
4. *Lawgiving or Choosing Principles?*  
71  
5. *Determination or Expression?*  
74  
6. *Three Criticisms Against Rawls*  
78  
7. *Dualism between the Right and the Good*  
83  
8. *Indefensibility of the Starting Point for Rawls’s Conception of Justice*  
88  
9. *Closing Remarks*  
94

**III When Political Philosophy Becomes Too Applied: A Kantian Appraisal of**
### the Roles of Political Philosophy Proposed by Rawls

1. Introduction
2. Rawls's Practical Role of Political Philosophy
3. Rawls's Orientation Role of Political Philosophy
4. Rawls's Reconciliation Role of Political Philosophy
5. The Role of Probing the Limits of Practicable Political Possibility
6. Closing Remarks

### IV A Political Conception of the Human Being in Kant: An Alternative Interpretive Starting Point of Kant’s Political Philosophy from a Political Perspective

1. Introduction
2. The Third Way: To Approach the Moral from the Political
3. Controversies Surrounding Constructivist Normative Starting Point
4. Rawls’s Verdict against Kant’s Philosophy
5. A Political Conception of the Human Being in Kant
6. Lawless Freedom, the Public Will and the Principle of Right
7. From the Limiting Condition of Politics to Morality
8. Closing Remarks

Conclusion

Bibliography
Abbreviations

The abbreviations of Kant’s works cited in this paper are:

A/B  *Critique of Pure Reason* (1781/1787)
Gr  *Groundwork of the Metaphysics of Morals* (1785)
IUH  *Idea for a Universal History with a Cosmopolitan Aim* (1784)
CPrR  *Critique of Practical Reason* (1788)
CPJ  *Critique of the Power of Judgment* (1790)
TP  *On the Common Saying: That May Be Correct in Theory, but it is of No Use in Practice* (1793)
Rel  *Religion within the Boundaries of Mere Reason* (1793)
PP  *Toward Perpetual Peace* (1795)
MM  *The Metaphysics of Morals* (1797)
SRLP  *On a Supposed Right to Lie from Philanthropy* (1797)
CF  *The Conflict of the Faculties* (1798)
KGS  *Kants gesammelte Schriften*

Volume numbers and pagination in KGS are indicated after the abbreviations.

The abbreviations of Rawls’s works cited in this paper are:

IMT  *The Independence of Moral Theory* (1975)
KC  *Kantian Constructivism in Moral Theory* (1980)
JFPM  *Justice as Fairness: Political Not Metaphysical* (1985)
PL  *Political Liberalism* (1996)
JFR  *Justice as Fairness: A Restatement* (2001)
LHPP  *Lectures on the History of Political Philosophy* (2007)
Introduction

The title of this dissertation by itself already gives rise at least to the following question: *What do you mean by “the moral” and “the political”?* Intriguingly, this is a question that is strikingly similar to the one I had in mind when I started having the idea of doing this research.

I had already spent a considerable amount of time on Kant’s philosophy when I started reading Rawls. Being quite immersed in Kant’s philosophy, especially his practical philosophy, I have to confess that I was indeed quite confused by the way Rawls uses the word “moral”. In Kant, it is unambiguously clear that the word “moral” is used only in relation to *morality*, which in turn refers to the fact that *reason legislates unconditionally in the will of human beings*. In Rawls, however, we are invited to accept a host of meanings for “moral” in his theory (one can see, at least, “moral conception”, “moral powers”, “moral sensibilities”, “moral psychology”, “moral theory”, “moral doctrine” and “moral philosophy” in his writings) in the absence of a common ground of *morality*. It therefore seemed to me that the use of the word “moral” in Rawls does not refer to a *fact*, as in Kant, but is solely determined by how Rawls himself understands it in different contexts.¹

Rawls’s use of the word “moral” may not bother me as much if his use were consistent within his own *empirical* theory and made no references to any *philosophy of morality*. In other words, I would readily accept an *empirical* use of the word in an *empirical* theory that does not really deal with the *philosophical* problem of *morality*. This, I believed, and still believe, is the position Rawls is

¹ Please see more in Chapter II and Chapter IV.
trying to take in his theory of justice and political liberalism. In accordance with
his method of avoidance to leave disputed philosophical questions aside, I was
therefore prepared to read Rawls without Kant in mind. Unfortunately, as will
become evident in this dissertation, Rawls’s political philosophy is mingled with a
host of philosophical claims, which, in my view, go beyond its proper ambit as an
empirical theory. I would leave the exact identification and the detailed
discussions of all these problematic claims to the main chapters of this dissertation.
For this introduction it suffices to note that these claims can be identified in his
understanding of the moral, his understanding of the political, and his
understanding of the connection between the moral and the political respectively.
The examination of these three types of philosophical claims will be dealt with in Chapter II, Chapter III and Chapter IV respectively.

This explains the rationale behind the title of this dissertation. The term the moral is chosen deliberately to leave open the possibility of leaving aside the
philosophical problem of morality in political philosophy. While the moral in
Kant’s political philosophy has to do with the morality in his full-fledged moral
philosophy, the moral in Rawls’s political philosophy represents nothing more
than his own understanding of the word in the absence of an explicit philosophical
account of morality. Their contrasting approaches squarely represent the two

---

2 Please see more in Chapter II.
3 For instance, Rawls’s understanding of the moral can be seen as a philosophical claim in opposition to Kant’s moral philosophy. Please see more in Chapter II.
4 Please see more in Chapter II.
5 Please see more in Chapter III.
6 Please see more in Chapter IV.
7 In this dissertation, I will argue that Rawls has been very consistent in his understanding of the moral, of the political and of the connection between the two throughout his writings both before and after his political turn. Please see more in Chapters II, III and IV.
opposite views in the long-standing debate of whether political philosophy should be done without moral philosophy. Yet the subject matter is further complicated by the fact that Rawls is often regarded as a Kantian due to his partial assimilation of Kant in his theory of justice. A comparative study of their views on the roles of the *moral* and the *political* in philosophy is thus particularly instructive in answering the following research questions: *To what extent does the moral have a role in political philosophy? And how does its role compare with that of the political?*

This dissertation will consist of four chapters. In Chapter I, I will offer a holistic interpretive account of Kant’s *moral* and *political* philosophy. This chapter does not deal with Rawls but will provide a point of reference for the rest of this dissertation. In Chapter II, as I noted above, I will shift my focus to Rawls and examine his understanding of the *moral* as reflected in his *moral* conception of the person. His view will be contrasted with Kant’s and it will be shown that the *moral* conception is the result of an intended reformulation and an unintended misreading of Kant’s notion of autonomy. I will go on to relate several weaknesses of Rawls’s theory to his (mis)understanding of the *moral* and argue that their resolutions require an accurate understanding of *morality.*

Chapter III is a Kantian appraisal of the four roles of political philosophy proposed by Rawls. It is relevant to the subject of this dissertation because the four roles are designed with a strict separation between the *political* and the *moral* in mind. If the four roles turn out to be defensible, it would amount to an important defense for such a separation. In the chapter, I will argue that these roles are not defensible from a Kantian perspective. Chapter IV offers an alternative interpretive approach to Kant’s moral and political philosophy by starting with a *political conception of the human being* and ending with his *moral philosophy.*
This alternative interpretive approach serves as a response to Rawls’s verdict against Kant by introducing a new perspective for the evaluation of the connection between the moral and the political in political philosophy.

The following four main chapters are therefore connected by the overarching research question presented above. Each chapter will have its own introductory section, I will therefore say no more about their contents here, or else I will be repeating myself later on. A few points are nonetheless worth noting before the first chapter. First, since the subject matter of this dissertation is the roles of the moral and the political in the philosophies of Kant and Rawls, I will not compare the substantial contents of their political philosophies. Second, in order not to further complicate the issues at hand, I will minimize the number of references to works on Kant’s moral and political philosophy prior to Rawls, and will largely focus on relevant works from Rawls’s times onward. Third, I have not in this dissertation made any special effort to conceal my preference for Kant over Rawls. I hope I can be excused for this inclination because Kant’s philosophy serves as the invaluable point of reference which renders this research possible in the first place. However, I do not intend this dissertation to be taken simply as a Kantian critique against Rawls, for this was never meant to be the purpose behind the research. The ultimate purpose is nothing less than clarifying the role of the moral in political philosophy per se.
I Re-defining Political Philosophy: Identifying the Moral and the Political in Kant’s Philosophy

1. Introduction

The current literature on Kant’s political philosophy, to a considerable extent, is filled with dissatisfaction, suspicion and confusion. The disappointments and misgivings arise on two fronts, both from within and without Kant’s philosophy. From within, interpreters still fall short of giving a convincing interpretive account for a seamless accommodation of Kant’s political writings in his critical philosophy. In particular, most interpreters find it difficult to reconcile Kant’s Doctrine of Right with his moral writings. From without, the assimilation of Kant in modern political philosophy dominated by the liberal tradition is partial at best and misleading at worst, for Kant’s critical philosophy seems to be at odds with certain fundamental liberal beliefs according to most liberal thinkers. Although these two debates go on in the literature more or less independently of each other, they arguably stem from the same philosophical problem: the lack of a convincing account of the connection between the moral and the political in Kant’s philosophy. In the former debate, Kant scholars cannot reach a unanimous agreement among themselves as to how Kant’s political concepts and principles in his political writings can be “derived” from the moral concepts and principles in his critical philosophy. In the latter debate, liberal thinkers tend to avoid Kant’s moral writings too prematurely in their assimilations of his political thought,

---

8 I will review the debate around the difficulty in the next section.
9 Please see Chapter II and Chapter IV for details.
evidently because they read into Kant’s philosophy their own understanding of the connection between the moral and the political.\textsuperscript{10} If the difficulties encountered in these two debates are no coincidence, they strongly suggest either that Kant’s political philosophy is not adequately supported by his moral doctrine, or that the interpretive approaches adopted in the literature so far are fundamentally flawed.

This first chapter serves to vindicate the latter claim by a reexamination of Kant’s account of the moral law and the law of right. As will become evident in the reexamination, the difficulty encountered in the reconciliation of the moral and the political in Kant’s philosophy arises not from Kant himself but from the misreading of both the moral law and the law of right on the part of his interpreters. This chapter will therefore start with a brief review of the debate on the problematic unity of the moral law and the law of right in Kant. It will be followed by a detailed explanation of an interpretive approach that starts with a reexamination of Kant’s account of the moral law. I will then explain how and why the moral law leads to the idea of a public will and the law of right. Towards the end of this chapter, I will further argue that the concept of public will in Kant has the law of right as its negative principle and the moral law as its positive principle. While the former leads to a negative duty of right in politics, the latter determines a positive duty in politics that corresponds to a positive agreement with the moral imperative of humanity as an end in itself. Kant’s political philosophy, therefore, should be seen as consisting of both duties, rather than of the negative duty of right alone. The connection between the moral and the political in Kant’s philosophy thus manifests itself in at least two related respects. First, the moral law not only serves as the foundation of the law of right but also implies it.

\textsuperscript{10} A typical example can be found in Rawls’s theory of justice and his partial assimilation of Kant. Please see Chapter II for details.
Second, the moral law is an integral part of Kant’s political philosophy, for it determines a positive duty in politics. I believe these two conclusions, which this chapter serves to vindicate, will shed light on the debate reviewed in the following section.

2. A Brief Review of the Debate

Before I introduce the new interpretive approach, it is helpful to review what other interpreters have said with regard to the connection between the moral and the political in Kant’s philosophy. To simplify matters, I will only consider the debate on the relationship between Kant’s moral writings and his *Doctrine of Right*, for this is where the contrast between the moral and the political manifest itself most conspicuously. The most prominent issue in the debate is arguably the reconciliation between the concept of freedom in Kant’s critical philosophy and the concept of external freedom in his *Doctrine of Right*. While Kant maintains that the law of the former concept of freedom, i.e. the moral law, also serves as the incentive to action, he holds in the latter concept of freedom that its corresponding law of right need not and should not be represented as the incentive to action when one’s aim is not to teach virtue but only to set forth what is right (see MM 6:231). In this section, I will review a spectrum of positions on the subject matter, starting with those who do not find a reconciliation between the two particularly problematic.

First, we have a number of Kant scholars who do not particularly take issue with the differences between the moral law in Kant’s moral writings and the law of right in the *Doctrine of Right*. Gregor, who published the first complete English translation of the *Metaphysics of Morals*, acknowledges at first that “the sort of constraint involved in juridical legislation [in the *Doctrine of Right*] is
problematic” (1993, 50). Yet she also readily accepts that since right is, as Kant holds, a capacity to put another under obligation, the Doctrine of Right is dependent upon his analysis of obligation, i.e. his moral philosophy (1993, 52-53). In order to account for the differences between the moral law and the law of right, she proposes in her own interpretation that “[any] constraint exercised upon [the capacity for choice] in accordance with a moral law will be obligation, whether such constraint is exercised only by one’s own will, through the thought of the lawfulness of one’s maxim, or also by another” (Gregor 1993, 70). Later on in this chapter, I will show that Gregor’s explanation is incomplete, for the concept of obligation alone is not sufficient to account for the differences between the moral law and the law of right.

Gregor’s position is shared by many others who also find the concept of unconditional obligation a sufficient connecting concept between Kant’s moral writings and the Doctrine of Right. Kersting, for instance, insists that the law of right “has the status of a synthetic a priori practical proposition; and on account of its practical necessity it must presuppose the validity of Kant’s doctrine of the fact of reason and the ensuing thesis of the reality of transcendental freedom” (Kersting 1992, 347). “If the concepts of pure practical reason and transcendental freedom should prove to be conceptual chimeras and ethical ghosts”, claims Kersting, “then the whole theory of unconditional practical obligation would also collapse… [a]nd the crash of the categorical imperative would then bring down with it the universal law of right…” (Kersting 1992, 347). At its best, however, Kersting’s argument can merely establish that Kant’s universal law of right requires his moral doctrine. It cannot explain why the moral doctrine requires the universal law of right or how the dependence is tenable. Mulholland (1990) also advances an argument similar to Gregor’s and Kersting’s. Mulholland starts by
asserting that “in the analysis of the concept of moral law that Kant has developed, it has been shown to be analytic that persons can be morally subject only to laws that are universal” (Mulholland 1990, 173). He then goes on to point out that “in all relations of rights, we have a condition in which one person has a duty, another a title to compel performance of that duty, in a relation in which duties and titles are determined by universal laws” (Mulholland 1990, 173). The connecting concept for Mulholland is therefore the notion of universal laws, the plausibility of which is not particularly distinctive from Gregor’s and Kersting’s. Their position is echoed by some other scholars who take the argument more lightly. Westphal, for instance, asserts that “[o]ur obligation to act morally entails an obligation to do whatever is necessary to act morally” (Westphal 1992, 410). He then goes on to argue that since “membership in civil society helps to bring about an improvement in moral character that is otherwise not generally possible”, our obligation includes membership in a state and its coercive laws (see Westphal 1992, 410). Reiss, in the introduction of his edition of Kant’s Political Writings, maintains that since “[t]he principle of universality demands that our social and political relations should be governed and our political conflicts settled in a universal manner”, “[t]his requires the existence of law” and therefore also a metaphysics of law (see Reiss 1991, 20).

There are nonetheless also some Kant scholars who attempt to address the differences between the moral law and the law of right in the problematic unity of Kant’s moral writings and the Doctrine of Right. In his attempt to argue for a categorical imperative of law in Kant, Höffe examines the peculiar features of Kant’s principle of justice (or the law of right) in Doctrine of Right and devises a hierarchical interpretive framework to accommodate the various forms of categorical imperative he identifies in Kant’s philosophy, including the law of
right (see Höffe 1989, 158). In the framework, Höffe offers an account for the differences between the highest categorical imperative in *Groundwork* and the law of right (or the categorical imperative of law, as Höffe calls it) in *Doctrine of Right* by distinguishing between the formal and the material implications of the highest categorical imperative (see Höffe 1989, 154). In Höffe’s view, *Groundwork* is the theory of ethics in the widest sense of the term and it thus furnishes the *material* basis for the whole *Metaphysics of Morals*, for it belongs to the nature of ethical virtue to satisfy duties, including the legal duties in *Doctrine of Right* (see Höffe 1989, 154). Höffe, however, argues at the same time that *Groundwork* serves as the *formal* basis for *Doctrine of Virtue* only, for “in a formal sense, ethical virtue consists of ‘self-coercion’ of the ‘inner freedom’” and is thus in opposition to the principle in *Doctrine of Right* (Höffe 1989, 154). Höffe’s argument therefore acknowledges and admits an opposition, at least in the formal sense as he calls it, between the categorical imperative (or the moral law) in Kant’s moral philosophy and the law of right in his political philosophy. Yet, on the other hand, since Kant’s moral philosophy serves as the *material* basis for the *Doctrine of Right*, Höffe insists on a moral concept of Law and claims that “however evidently Kant separates Law from Ethics, he does not separate it from the moral point of view” (Höffe 1989, 165).

Guyer also acknowledges the apparent opposition between the categorical imperative and the law of right when he holds that Kant’s universal principle of right is not derived from the categorical imperative, “because the principle of right concerns only the compatibility of our actions with the freedom of others, and does not concern our maxims at all, a fortiori their universality” (Guyer 2002, 25). In view of this difficulty, Guyer also proposes an interpretive framework to reconcile the two. Instead of designating Kant’s categorical imperative as the
highest principle as Höffe does in his interpretation, Guyer claims that he identifies a more fundamental concept of morality in Kant which can serve as the connecting concept between the categorical imperative and the law of right:

The foundational assumption of Kantian morality is that human freedom has unconditional value, and both the Categorical Imperative and the universal principle of right flow directly from this fundamental normative claim: the Categorical Imperative tells us what form our maxims must take if they are always to be compatible with the fundamental value of freedom, and the universal principle of right tells us what form our actions must take if they are to be compatible with the universal value of freedom, regardless of our maxims and motivations. (Guyer 2002, 26)

In all the interpretations discussed above, the interpreters purportedly identify, or clarify (when they think Kant is not clear enough), the connecting concept between Kant’s moral writings and the *Doctrine of Right*. In the following sections when I introduce the new interpretive approach, I will discuss the deficiencies of their interpretations. In the meantime, one thing we should note in their interpretations is that they are all claiming a dependence of Kant’s *Doctrine of Right* on his moral doctrine, while a dependence in the opposite direction might not be sufficiently developed. On the other side of the dispute, however, we have other commentators who hold exactly the opposite view: Kant’s moral doctrine entails his *Doctrine of Right*, but this entailment does not imply that the *Doctrine of Right* is dependent upon (or cannot stand without) his moral doctrine (see Pogge 2002, 151). In other words, they hold that Kant establishes merely a
one-sided dependence of his moral philosophy on his *Doctrine of Right* (see Pogge 2002, 151). This unusual view originates from the German literature\(^{11}\) and is called the *independence thesis* by one of its principal opponents, Kersting (1984).

In the English literature, Pogge (2002), for instance, holds essentially the same view as the *independence thesis* when he stipulates that the *Doctrine of Right* “can stand on its own, independently from [Kant’s] moral philosophy and transcendental idealism” (153). Many other interpreters, however, do not find the idea of a one-sided dependence convincing but insist on the mutual independence of Kant’s moral doctrine and the *Doctrine of Right*. In his monograph on Kant’s political philosophy, Ripstein (2009) notes near the beginning that “Kant denies that political philosophy is an application of the Categorical Imperative to a specific situation” (2), and ends the book by claiming that “[a]ttempts to derive the Principle of Right from the Categorical Imperative on its own fail, and end up rejecting the Principle of Right, or both the Categorical Imperative and the Principle of Right” (388). Ripstein thereby attributes the mutual independence of the two to Kant himself and rejects the applied ethics reading (see Ripstein 2009, 11). Ripstein also criticizes the views of Hermann Cohen and Rawls, who “suppose that the autonomous life is the best one, and political institutions must be designed to promote autonomy”, for this position “places it squarely in the instrumentalist camp” and cannot be Kant’s view (see Ripstein 2009, 11). Wood (2002) shares Ripstein’s view but he adopts a different line of reasoning. Wood alleges that “Kant very explicitly discredits the whole idea that the principle of right could be derived from the fundamental principle of morality by declaring

\(^{11}\) According to Pogge (2002, 151), Kersting attributes the view to Ebbinghaus (1988), Reich (1936) and Geismann (1974).
that the principle of right, unlike the principle of morality, is *analytic*” (Wood 2002, 7). This claim is made based on the straightforward argument that “it would be nonsense to think that we need to derive an analytic proposition from a synthetic one” (Wood 2002, 7). In the following sections, I will show that Ripstein’s and Wood’s views are interpretations out of context and should not be attributed to Kant.

Like Wood and Ripstein, Willaschek also argues for the complete independence of the *Doctrine of Right* from Kant’s moral doctrine. Unlike Wood and Ripstein, however, he does not attribute the interpretation entirely to Kant. In his earlier paper on the dispute, Willaschek suggests that the *Doctrine of Right* should not belong in the *Metaphysics of Morals* in the first place, for “Kant (inadvertently) employs two conflicting views about how the realm of right (Recht) is related to morality (Sittlichkeit, Moral) and ethics (Ethik, Tugendlehre) and that the resulting conflicts may account for many of the difficulties one may find with the *Metaphysics of Morals*” (see Willaschek 1997, 205). The two conflicting views, according to Willaschek, are the “official” view, in which the *Doctrine of Right* would belong in the *Metaphysics of Morals*, and the “alternative” view, in which it should not. In his paper, Willaschek argues that the “official” view is “a building plan that had long been ready in his mind”, which “dates back at least to 1784” (Willaschek 1997, 225). Yet, when Kant actually wrote the *Metaphysics of Morals*, he encountered numerous difficulties “which he could solve only by implicitly relying on the alternative view” (Willaschek 1997, 225). “According to the alternative view”, Willaschek claims, “juridical laws are independent expressions of the autonomy of pure practical reason, analogous to, but not derived from, the laws of morality” (Willaschek 1997, 225). Willaschek therefore argues that while Kant’s original building plan was to have the concept of right
derived from the moral law, he failed in the end and relied on a view that implies their mutual independence.\textsuperscript{12}

There are certainly other commentators who have also expressed their views on the subject matter and it is impossible to offer an exhaustive list here. However, the spectrum of positions covered in this section should be broad enough to accommodate the other views that are not reviewed here. In the following sections, I will introduce a new interpretive approach that, in my view, is not only more faithful to Kant’s intention but can also resolve the difficulties mentioned in this section.

3. The Basics of the New Interpretive Approach

To borrow Pippin’s words, the whole debate leaves us in a troubling aporia (see Pippin 2006, 425). As is evident in the last section, neither the derivationists,\textsuperscript{13} who insist that the principle of right be connected in some way with Kant’s overall moral theory, nor the separationists, who want the two strictly separate, are conclusively convincing. In many cases, they cannot even agree among themselves, let alone with their counterparts on the other side of the debate, whether their views can be attributed to Kant himself. Amidst all the dissatisfaction, suspicion and confusion, if we are not thereby led to submit to Schopenhauer’s suggestion that the Doctrine of Right is nothing but a sign of Kant’s senility (see Schopenhauer 2010, Book IV, § 62), it might be necessary to

\textsuperscript{12} In another paper (Willaschek 2002), Willaschek apparently attempts to tame the paradox arising from the two conflicting views by investigating the paradoxical nature of the concept of “juridical imperative” he attributes to the Doctrine of Right. In yet another paper (Willaschek 2009), he elaborates his view and concludes again that “on Kant’s view, the fundamental principles of right are independent expressions of rational autonomy, on a par with, or at least not derivable from, the Categorical Imperative” (67).

\textsuperscript{13} As Pippin calls them (see Pippin 2006, 420).
reflect upon the dispute and see if there are any fundamental flaws in the interpretive approaches (or the underlying way of thinking) adopted thus far. In this section, I will identify two fundamental flaws in the interpretive approaches of the current literature. Before that, however, I would like to cite a little-noticed reminder from Kant which is very much relevant to these two flaws I am going to identify. In his *Groundwork*, Kant reminds us that:

But in practical matters, it is just when common understanding excludes all sensible incentives from practical laws that its faculty of appraising first begins to show itself to advantage. It then becomes even subtle, whether in quibbling tricks with its own conscience or with other claims regarding what is to be called right, or in sincerely wanting to determine the worth of actions for its own instruction; and, what is most admirable, in the latter case it can even have as good a hope of hitting the mark as any philosopher can promise himself; indeed, it is almost more sure in this matter, because a philosopher, though he cannot have any other principle than that of common understanding, can easily confuse his judgment by a mass of considerations foreign and irrelevant to the matter and deflect it from the straight course. Would it not therefore be more advisable in moral matters to leave the judgment of common reason as it is and, at most, call in philosophy only to present the system of morals all the more completely and apprehensibly and to present its rules in a form more convenient for use (still more for disputation), but not to lead common human understanding, even in practical matters, away from its fortunate simplicity and to put it, by means of philosophy, on a new path of investigation and instruction? (Gr 4:404)
I am quoting the long passage in full because I cannot possibly explain his position in a more concise manner. With respect to the current subject matter, I would like to highlight a few points in the passage that are relevant to the present chapter. First, for Kant, “practical matters” (*praktischen*) include everything related to practical reason, i.e. the will. Practical matters therefore include politics. Second, philosophers should acknowledge that they may be more prone to misjudgment than ordinary people who manage to uphold their faculty of appraising, for a philosopher “can easily confuse his judgment by a mass of considerations foreign and irrelevant to the matter and deflect it from the straight course”. Third, philosophers should attend to their proper task of practical philosophy, which is “to present the system of morals all the more completely and apprehensibly and to present its rules in a form more convenient for use (still more for disputation)”. Philosophers should refrain from leading ordinary people away from their fortunate simplicity and to put their faculty of appraising to “a new path of investigation and instruction” as a result of misjudgments on the part of the philosophers themselves. By highlighting these three points in Kant’s reminder, I do not plan to speculate whether many philosophers are confusing their own judgments and trying to lead ordinary people away from their fortunate simplicity when it comes to a proper understanding of the relationship between morality and politics. They are mentioned here merely because they can provide the necessary support for the identification of the following two flaws in the literature.

First, if Kant’s “practical matters” include everything that is related to practical reason and therefore should include both his moral writings and his *Doctrine of Right*, the common cornerstone of the two should be practical reason. This claim might seem trivial but it is not. As can be observed in the review in the
last section, the interpreters argue for the dependence (or the independence) of the two by identifying and vindicating various shared (or contradicting) concepts in the two doctrines. When a shared concept is identified in both (whether it be the concept of obligation, the concept of universality or the concept of practical necessity), some interpreters too hastily allege that a connecting concept is found and the unity of the two is unproblematic. In view of arguments like this, all their opponents have to do is to identify a contradicting concept in the two (whether it be the concept of lawgiving, the concept of motivation, or the concept of legality), and allege that the two doctrines have to be independent of each other. However, when the common cornerstone of Kant’s moral writings and his *Doctrine of Right* has to be practical reason, interpreters do not really have the liberty to pick up any of the concepts that appear in the two doctrines and enjoy the convenience to argue for or against the unity of the two on so slender a basis. Instead, the proper questions to ask in the subject matter should be as follows: Are the accounts of practical reason in Kant’s moral writings and his *Doctrine of Right* necessarily referring to one and the same faculty of practical reason? Do these accounts of practical reason necessarily require each other? Is this comprehensive account of practical reason based on an indisputable fact? If the answers to these three questions are all yes, the unity of the two doctrines should be declared necessary and unproblematic. This conclusion (or its rejection) cannot be established by appealing to the apparent similarity (or dissimilarity) of any arbitrarily selected concept in the two doctrines. Such comparisons are fragmentary and can contribute nothing to a holistic understanding of practical reason. I will further elaborate on this point in the next section.

While the first flaw leads to a fragmented understanding of practical reason, the second flaw is the unavoidable consequence of an equally fragmented
understanding of Kant’s practical philosophy. According to the above passage (Gr 4:404), in Kant’s view, the proper task of practical philosophy is “to present the system of morals all the more completely and apprehensibly and to present its rules in a form more convenient for use (still more for disputation)”. Kant’s philosophy is unquestionably an attempt to accomplish this very task Kant himself prescribes for philosophy. Most commentators, however, come to interpret Kant with their own prior understanding of the difference between the moral and the political, which is most notably reflected in their sharp demarcation of Kant’s works into his moral philosophy and his political philosophy in the debate. Such premature demarcation not only leads to the first flaw, i.e. a fragmented understanding of practical reason, but also leads to a fragmented understanding of Kant’s practical philosophy and the political philosophy it contains. For instance, the popular identification of Kant’s political philosophy primarily in his *Doctrine of Right* and his various “political” writings may seriously miss out on the roles of other elements in other writings that are equally important in his political philosophy. In the rest of this chapter, I will illustrate through the new interpretive approach what is missing in the fragmented interpretation of Kant’s political philosophy.

4. **The Moral Law as the Limiting Condition of the Freedom of Action**

This section and the next will deal specifically with the debate on the problematic unity of Kant’s moral philosophy and his *Doctrine of Right*. As I argued in the previous section, all we need to establish a seamless unity of the two doctrines is a single coherent account of practical reason in which both the moral law and the law of right are necessary and require each other. In view of this, this
section will prepare for the next by introducing the concept of the limiting condition of the freedom of action of every human being in Kant’s account of the moral law. In the next section, I will then argue why the limiting condition of the freedom of action of every human being implies the limiting condition of the freedom of action of human beings in their external relations. To begin with, let us read the well-known three formulations of the moral law in *Groundwork*, including the first *Formula of Universal Law* (with its variant, the *Formula of the Law of Nature*), the second *Formula of Humanity* and the third *Formula of Autonomy* (with its variant, *Formula of the Realm of Ends*):14

*Formula of Universal Law* (FUL): “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (Gr 4:421; cf. Gr 4:402).

*Formula of the Law of Nature* (FLN): “act as if the maxim of your action were to become by your will a universal law of nature” (Gr 4:421; cf. Gr 4:436).

*Formula of Humanity* (FH): “so act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (Gr 4:429; cf. Gr 4:436).

*Formula of Autonomy* (FA): “… the idea of the will of every rational being as a will giving universal law” (Gr 4:431; cf. Gr 4:432).

*Formula of the Kingdom of Ends* (FKE): “act in accordance with the

---

14 Please also see Wood (2008, 66-67) for a complete list of these formulae.
maxims of a member giving universal laws for a merely possible kingdom of ends” (Gr 4:439; cf. Gr 4:433, 437, 438).

“The above three ways of representing the principle of morality”, Kant asserts, “are at bottom only so many formulae of the very same law, and any of them of itself unites the other two in it” (Gr 4:436). In *Groundwork*, Kant articulates various characteristics of the moral law through his elaborations of these three formulae. These explications have been under extensive scrutiny ever since its publication and, in Höffe’s words back in 1989, “one might be tempted to consider the subject exhausted” (Höffe 1989, 149). Yet, as is evident in the second section of this chapter, in spite of the large number of studies on Kant’s account of the moral law, the current literature still falls short of offering a conclusively convincing account of its relationship with the *Doctrine of Right*. It is therefore doubtful whether the interpretations in the current literature are unproblematic, if the failure to unite the moral law and the law of right cannot be attributed to Kant’s own fault. Before I discuss the idea of a limiting condition in Kant’s account of the moral law, I will first deal with a commonplace misreading of the three formulae in the literature, in which the formulae are wrongly interpreted as three different “tests” of maxims rather than as three ways of representing the same principle of morality as Kant clearly intends.

After the explications of the three formulae in *Groundwork*, Kant acknowledges that “there is nevertheless a difference among the [three formulae],

15 Wood argues that the German original, “*deren die eine die anderen zwei von selbst in sich vereinigt*”, should not be translated “as saying that each of the formulas unites the other two in itself” (Wood 2008, 80). I will argue below that each of them does in fact unite the other two in itself, and that they are really three different ways of representing the same principle of morality. My position is that Kant’s position is unmistakably clear, for he claims that the three formulae are three ways of representing the *principle* of morality, which is singular.
which is indeed subjectively rather than objectively practical, intended namely to bring an idea of reason closer to intuition (by a certain analogy) and thereby to feeling” (Gr 4:436). The statement indicates the presupposition of and thereby prepares his reader for the connection between the three formulae and the three elements in every maxim: (1) With regard to the form of a maxim, the universality of the maxim is at stake, the moral imperative is the FUL. (2) With regard to the matter of a maxim, the end of the maxim is at stake, the moral imperative is thus the FH. (3) With regard to the complete determination of all maxims, the moral imperative is the FA (see Gr 4:436). By the end of these explications, Kant reminds us that:

A progression takes place here, as through the categories of the unity of the form of the will (its universality), the plurality of the matter (of objects, i.e., of ends), and the allness or totality of the system of these. But one does better always to proceed in moral appraisal by the strict method and put at its basis the universal formula of the categorical imperative: act in accordance with a maxim that can at the same time make itself a universal law. If, however, one wants also to provide access for the moral law, it is very useful to bring one and the same action under the three concepts mentioned above and thereby, as far as possible, bring it closer to intuition. (Gr 4:436-437)

In his interpretation of this passage, Wood (2008) suggests that the three formulae “constitute a developmental progression” and that “we need all the formulae to have a complete account of the content of the supreme principle” (69). However, these two claims reflect a misunderstanding on Wood’s part, which is
most evident in his additional claim that the three formulae “differ both ‘objectively’, in what they command, even more ‘subjectively’, in the aspect of the law they present to the moral agent” (69). Contrary to his reading, which essentially suggests that we have three different moral imperatives, Kant clearly maintains unmistakably in the passage that the three formulae are the results of bringing “one and the same action under the three concepts mentioned above” in order to provide access for the moral law, which is in the singular. The whole purpose of the developmental progression is no more than to bring the same moral imperative “closer to intuition”, by elucidating the same moral imperative with regard to the form, the matter and the complete determination of a maxim. In other words, they do not differ objectively as Wood suggests.

In order to derive the law of right from the moral law, we must first offer a coherent account of the three formulae of the supreme principle of morality in response to a (mis-)interpretation like Wood’s. As is evident from his arguments, his misreading primarily stems from his failure to distinguish between the method Kant adopts in the philosophical exposition of the moral law and the fundamental concept of a will under the necessitation of a self-given categorical imperative. In the former, the three formulae are unquestionably different, as Kant himself recognizes. In the latter, however, it can be shown that the three formulae necessarily imply each other. For instance, the FUL unites the FH in it because the moral imperative with regard to the form of a maxim rejects any determination of the will merely by relative and arbitrary matters, i.e. relative and arbitrary ends. Yet, since the will is nothing but a faculty of ends and therefore cannot be indifferent to ends, i.e. take no interest in them, it is nonetheless necessary and possible to determine an objective end with regard to the formal requirement of universality in the FUL (see MM 6:395; Rel 6:5). A moral imperative with regard
to the *matter* thus arises and it can be nothing but the FH. The FUL also unites the FA in it because the former repudiates all maxims that “are inconsistent with the will’s own lawgiving of universal law” (Gr 4:431). In accordance with the FUL, therefore, “the will is not merely subject to the law but subject to it in such a way that it must be viewed as also giving the law to itself and just because of this as first subject to the law (of which it can regard itself as the author)” (Gr 4:431). The FUL thus implies the autonomy of the will and hence the FA.

That FH has the other two formulae united it can be defended similarly. By prescribing *an end in itself*, the FH is the supreme *limiting condition* of the freedom of action of every human being and all merely subjective ends (see Gr 4:431). Its being a *limiting condition* indicates the formal implication of the FH; and its being the *limiting condition* of the freedom of action of *every* human being and *all* subjective ends indicates its *universality*. The FH thus unites the FUL in it. The *self-prescription* of an end in itself by the will itself as the limiting condition of all merely subjective ends in turn implies the autonomy of the will and therefore unites the FA. The FA can be shown to be uniting the other two through the same line of argumentation, i.e. if we properly consider the three formulae not as three different tests of maxims but as ways of representing the same moral law the will imposes on itself in its determination of a maxim.

---

16 In the fourth chapter of his *Kantian Ethics*, Wood clearly interprets the three formulae as three different tests of maxims (see 2008, 66-84). Allison also speaks of “the relatively straightforward testing procedure that Kant sketches” in *Groundwork* (1990, 94). In a later work, Allison similarly claims that “[a]s is clear to any reader of the *Groundwork*, [Kant’s view of moral deliberation] consists in subjecting one’s maxim to the universalizability test in order to determine whether it can be universalized without contradiction” (1996, 180). Allison’s claim, however, is followed by his acknowledgement of the weakness of such interpretation (see Allison 1996, 180). In the next section, I will produce direct textual evidence to defy the “universalizability test” reading, including Wood’s and Allison’s variant.
As promised at the beginning of this section, this section will introduce the concept of the limiting condition of the freedom of action in Kant’s account of the moral law, and the concept will serve as a preparation for the derivation of the law of right from the moral law in the next section. The above discussion offers the necessary justification for this approach because it establishes that any one of the three formulae in Kant is sufficient to represent the supreme principle of morality. As briefly mentioned above, the concept of limiting condition appears in *Groundwork* when Kant designates the FH as the supreme limiting condition of the freedom of action of every human being and the supreme limiting condition of all subjective ends (Gr 4:430-431). To understand what it means to be a limiting condition, let us review how Kant elaborates the FH:

[1] If, then, there is to be a supreme practical principle and, with respect to the human will, a categorical imperative, [2] it must be one such that, from the representation of what is necessarily an end for everyone because it is an *end in itself*, it constitutes an *objective* principle of the will and thus can serve as a universal practical law. The ground of this principle is: *rational nature exists as an end in itself*. [3] The human being necessarily represents his own existence in this way; so far it is thus a subjective principle of human actions. [4] But every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me; thus it is at the same time an *objective* principle [5] from which, as a supreme practical ground, it must be possible to derive all laws of the will. The practical imperative will therefore be the following: [FH] (Gr 4:428-429)
In (1), Kant states that the whole argumentation is made on a presupposition, namely, if there is to be a supreme practical principle, or a categorical imperative, with respect to the human will. It is clear from Kant’s statement that the reality of the categorical imperative is not vindicated but merely presupposed here in this passage (and actually in the whole *Groundwork*). This declaration notably echoes his position in another passage in *Groundwork* where he unequivocally holds that “in the idea of freedom we have actually only presupposed the moral law, namely the principle of the autonomy of the will itself, and could not prove by itself its reality and objective necessity” (Gr 4:449). This is also in line with his later claim in the *Critique of Practical Reason* that our cognition of the unconditionally practical cannot start from freedom but from the moral law:

It is therefore the moral law, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves), that first offers itself to us and, inasmuch as reason presents it as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them, leads directly to the concept of freedom. (CPrR 5:29-30)

Our consciousness of the moral law thus vindicates the reality not only of itself but also the idea of freedom in the *Critique of Practical Reason*. In the rest of this chapter, I will not discuss any further the reality of the moral law, for the aim of this chapter is merely to offer an account of the distinction between the moral and the political in Kant’s philosophy. Its reality will be seen as unproblematic and given in the following discussion.

The rest of the passage can be unpacked as follows. (2) Such a supreme
practical principle, for it to serve as a universal practical law, must bring with it a necessary *end in itself* for everyone.\textsuperscript{17} Or in other words, the ground of the principle must be that: *rational nature exists as an end in itself*, for the *end in itself* can be nothing but the existence of every rational being.\textsuperscript{18} (3) A human being, as a rational being, necessarily represents his or her own existence as an *end in itself* and this representation is at first a subjective principle of human actions. (4) Since the principle serves as a practical law for us only as *rational beings*, it must also hold for all rational beings (see also Gr 4:447). Every other human being therefore should also represent his or her existence in this way. The acknowledgement of this fact renders the subjective principle of human actions an objective principle at the same time. Points (2) to (5) thus constitute the primary arguments as to how the FH qualifies as a formula to represent the principle of morality as a supreme practical ground. In (2) and (3), Kant makes explicit reference to a necessary subjective end, and then in (4), he explicitly explains the “public” and “objective” dimension of the supreme practical ground. These two features of the FH make it the most appropriate formulation of the moral law to start with in the derivation of the law of right. My approach to derive the law of

\textsuperscript{17} This is because, as Kant argues in *Groundwork*, relative ends can furnish no universal principles: “The ends that a rational being proposes at his discretion as effects of his actions (material ends) are all only relative; for only their mere relation to a specially constituted faculty of desire on the part of the subject gives them their worth, which can therefore furnish no universal principles, no principles valid and necessary for all rational beings and also for every volition, that is, no practical laws. Hence all these relative ends are only the ground of hypothetical imperatives.” (Gr 4:428)

\textsuperscript{18} Kant arrives at this conclusion by arguing that, first, the worth of any object to be acquired by our action is always conditional; and second, for other things the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth (see Gr 4:428). The worth of these objects are therefore merely contingent, whereas the worth of an end in itself must be *absolute* (see Gr 4:428). Therefore, if there is to be a supreme practical principle, the ground must be that *rational nature exists as an end in itself*. 26
right from the FH is further confirmed by Kant’s claim in (5) that “it must be possible to derive all laws of the will” from the FH as a supreme practical ground, since the law of right is unquestionably also a law of the will in Kant.

The passage thus serves to explain why the FH represents the principle of morality as the supreme practical ground. As noted above, Kant also calls this principle of humanity the supreme limiting condition of the freedom of action of every human being and the supreme limiting condition of all subjective ends (see Gr 4:430-431). In the following section, I will build on this concept of limiting condition and reconsider its imperatival implications on the form, the matter and the complete determination of all maxims of action. I will then illustrate that these implications necessarily lead to the idea of a public will, from which the law of right can be inferred.

**5. From the Moral Law to the Idea of a Public Will**

As I argued in the last section, Kant rightly intends to have the FH uniting the other two formulae in it and regards the FH as representing the limiting condition of the freedom of action of every human being and of all subjective ends. As we shall see in this section, it can be further argued that the moral law is nothing but the limiting condition of the freedom of action in general and it subsequently leads to an important limiting condition of the freedom of action in our external relations, which is nothing but the law of right in Kant’s *Doctrine of Right* and his political writings (see MM 6:449). I will begin by reconsidering the imperatival implications of the limiting condition on the form, the matter and the complete determination of all maxims of action.

The formal implication of the limiting condition on all maxims of action is unmistakably the universality of a maxim in Kant’s account of the moral law,
which is best illustrated in his FUL:

*Formula of Universal Law* (FUL): “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (Gr 4:421; cf. Gr 4:402).

This imperative, as straightforward as it might seem at first sight, could be misinterpreted at times. As noted briefly above in a footnote, Allison claims that “[a]s is clear to any reader of the *Groundwork*, [Kant’s view of moral deliberation] consists in subjecting one’s maxim to the universalizability test in order to determine whether it can be universalized without contradiction” (1996, 180). The “universalizability test” interpretation, however, fails to grasp the essence of the imperative, but has its attention misplaced on a theoretical construal of the imperative. In essence, as Kant argues in *Groundwork*, “[t]he representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an imperative” (Gr 4:413). The primary function of the moral law as an imperative to the will of every human being is not the (theoretical) testing of the universalizability of a given maxim but the necessitation of the will by its self-given objective principle. The difference is subtle but important. If we wrongly regard the “universalizability test” reading as an account of the only formal imperatival implication of the moral law, we would be led to the awkward conclusion that, as Allison detects in the literature, “virtually any maxim, if subtly formulated, can be made to pass the test” (Allison 1996, 180). 19

---

19 The charge dates back at least to the nineteenth century when Mill, in his *Utilitarianism*, accuses Kant of failing to “show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral
The so-called “universalizability test” reading endorsed by Allison, in Kant’s view, is not defensible without the additional support of a correct interpretation of the FUL. In *Groundwork*, Kant explains specifically the differences between the FUL and the “universalizability test”:

We must be able to will that a maxim of our action become a universal law: this is the canon of moral appraisal of action in general. Some actions are so constituted that their maxim cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* that it *should* become such. In the case of others that inner impossibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself. It is easy to see that the first is opposed to strict or narrower (unremitting) duty, the second only to wide (meritorious) duty; and so all duties, as far as the kind of obligation (not the object of their action) is concerned, have by these examples been set out completely in their dependence upon the one principle. (Gr 4:424)

In this passage, Kant clearly distinguishes two imperatival implications of the same FUL. First, we should not act on any maxim that cannot be *thought* as a universal law of nature without contradiction. This requirement is clearly the same as the “universalizability test” reading which holds that Kant’s view of moral deliberation “consists in subjecting one’s maxim to the universalizability test in order to determine whether it can be universalized without contradiction” (Allison rules of conduct” (Mill 1988, 4).
As is evident from the passage, however, Kant unambiguously asserts that this first requirement is merely a preliminary one, for even when a maxim can be thought theoretically as a universal law of nature without contradiction, it might still be impossible to will practically that it should become such. The spirit of the FUL therefore lies within its second and higher imperatival implication explicated in the passage, which requires also that we will the maxim’s being a universal law of nature. Kant’s view here is especially unmistakable if one considers the formulation of the FUL carefully. The FUL explicitly requires that we will practically that the maxim should become a universal law of nature, not only that it can be thought theoretically as a universal law without contradiction.20

20 The “universalizability test” reading thus leads to a significant misinterpretation of Kant’s FUL and the misreading mainly arises from the difficulty to distinguish between the two different notions of contradiction found in the passage quoted in the main text (Gr 4:424), namely, a maxim in contradiction with itself when made a universal law and a will in contradiction with itself. Unfortunately, the misreading is so common in the current literature that it is very difficult to find a single Kant interpreter who does not think that the FUL is nothing more than a “universalizability test”. The problem is further complicated by the fact that many of these interpreters are dissatisfied by the efficacy of the FUL as a result of the misreading and propose various defenses against the charge first made by Mill in the nineteenth century (see footnote 19). For instance, Allison’s “universalizability test” reading (see footnote 16) is supplemented by his remark that we should attend faithfully to the “morally salient features of a situation” and adopts a hierarchical view of maxims (see Allison 1990, 91-94; 1996, 116) when we conduct the “universalizability test”. Wood similarly interprets the FUL as a “universalizability test” of maxims, but he acknowledges at the same time that “it would be radically defective as a general moral criterion, since it systematically yields both false positives and false negatives when we try to employ it generally” (Wood 2006, 345). In Wood’s view, “Kant’s actual account of ordinary moral reasoning… turns not on figuring out which maxims are universalizable, but on reasoning from a system of duties…” (Wood 2006, 345). In his view, the quest for an interpretation of the “universalizability test” that enables it to serve as a general moral criterion is “worse than a waste of time”, since it encourages “critics of Kant’s ethics to continue thinking, falsely, that something of importance for Kantian ethics turns on whether there is a universalizability test for maxims that could serve as such a general moral criterion” (Wood 2006, 373). Apart from the Kant interpreters, the procedural reading of Kant’s categorical imperative introduced by O’Neill (1975) and Rawls (1989) in the seventies and eighties also interprets the FUL as a test of maxims. In Rawls’s “highly schematic rendering of Kant’s conception of the categorical imperative”, he proposes a four-step
It should be clear from the discussion above that the moral law, or the categorical imperative with respect to the human will, is the limiting condition of the freedom of action because it is “the sole condition under which a will can never be in conflict with itself” (Gr 4:437). By “limiting condition”, Kant does not mean merely that the will or the freedom of action is limited by a rule that is simply (self-)given, but he also means that this is the rule only under which the will, and thus the freedom of action, can be brought into a harmonious unity with itself in all its undertakings. Or in other words, the limiting condition is the sole condition under which the will can determine itself thoroughly and consistently. In this regard, the “universalizability test” serves merely as a preliminary test which, if we understand the FUL correctly, is not absolutely necessary, for the thorough self-consistency of the will obviously requires that its maxim be able to be thought without contradiction as a universal law of nature. After all, the will cannot be unconditionally commanded by itself to determine itself to a self-contradiction.

The notion that the moral law (as the limiting condition of the freedom of action) is the sole condition under which a will can be thoroughly consistent with itself can also be seen in Kant’s other writings. For instance, in *Critique of Practical Reason*, Kant alleges that the determinations of a practical reason take place “only in order to subject a priori the manifold of desires to the unity of “categorical imperative procedure” that is mainly meant to be a procedural interpretation of the FUL as a test (see Rawls 1989, 498-506). His account of the procedure, as he admits, in turn follows closely that of O’Neill’s (see Rawls 1989, 498; O’Neill 1975, 136-193), who later also formulates her own defense of the FUL as a test in her *Constructions of Reason* (see O’Neill 1989, 81-104). In my opinion, none of these variant forms of “universalizability test” reading is a faithful interpretation of Kant’s FUL and categorical imperative. In order to explain what Kant means by the phrase “able to will”, we need to have a very deep and holistic understanding of the will in Kant’s philosophy. I will deal with the problem of universalizability in another paper.
consciousness of a practical reason commanding in the moral law, or of pure will” (CPrR 5:65). In other words, the moral law is the sole condition under which (and also the law through which) practical reason (the will) can be consistent with itself amidst the manifold of desires. In the absence of the moral law, practical reason would always be empirically conditioned and be in contradiction with itself amidst the manifold of desires. In Religion within the Boundaries of Mere Reason, Kant speaks of “the unity of maxims in general” as the characteristic of the moral law (see Rel 6:36-37), which means nothing but the thoroughgoing consistency of the will with itself in all its maxims. We can also discern this implication when Kant speaks of “the unity of the form of the will (its universality)” (Gr 4:436).

It might be already quite obvious to some readers as to why the above notion of the limiting condition of the freedom of action should lead to the law of right in the Doctrine of Right. A more cautious elucidation of their connection nonetheless requires also a more in-depth examination of the imperatival implications of the moral law on the other two elements of a maxim, i.e. its matter and its complete determination.

As noted in the last section, the FH specifies the moral imperative with regard to the matter of a maxim by determining the necessary objective end, i.e. rational nature exists as an end in itself, pursuant to the supreme practical principle (Gr 4:428-429). It was also noted in the last section that “[t]he human being necessarily represents his own existence in this way; so far it is thus a subjective principle of human actions” (Gr 4:429). In other words, according to Kant’s elaboration of the FH, the objective end (it is so because it is the necessary object of an objective principle of the will) can be seen as serving as a subjective principle of human actions. Or to put it differently, the transition from an objective principle to a subjective principle is made possible by the concept of objective end.
While Kant’s explanation is clear on this point, it is worthwhile to add one more point here in this section in order to preempt any potential misreading with regard to the concept of objective end in the moral law. Let us read another passage about the FH:

… in [the principle of humanity as an end in itself] humanity is represented not as an end of human beings (subjectively), that is, not as an object that we of ourselves actually make our end, but as an objective end that, whatever ends we may have, ought as law to constitute the supreme limiting condition of all subjective ends, so that the principle must arise from pure reason. (Gr 4:431)

This passage appears in Kant’s explication of the FH and covers an important characteristic of the moral law that is critical for the derivation of the law of right. According to the passage, the objective end ought as law to constitute the supreme limiting condition of all subjective ends, but it is not thereby to be represented in the law as an object we of ourselves actually make our end. This clarification reveals an important basic principle in Kant’s practical philosophy, which is that the formal principle of practical reason must undoubtedly take precedence over its material principle (see PP 8:376-377). The primacy of the formal principle is important because the practical reality of the objective end, which is a necessary object of the supreme practical principle, depends on the practical reality of reason, which proves its own reality by determining the will immediately by the condition of a universal lawful form of its maxims (see CPrR 5:48), i.e. the moral law. The moral law, which is merely formal (see CPrR 5:109), is therefore by itself the sole determining ground of the will. Though humanity is an objective
end pursuant to the moral law, it is not on that account to be taken as the determining ground of the will.\textsuperscript{21} This is basically the rationale behind Kant’s claim in the passage that “humanity is represented not as an end of human beings (subjectively), that is, not as an object that we of ourselves actually make our end”, for this wrongly implies that the categorical imperative commands that the end be the determining ground of the will.

The objective end of humanity thus serves as the \textit{a priori} limiting condition of all subjective ends, which, not unlike the limiting condition of the freedom of action discussed before, means nothing more than this: the objective end as a subjective principle is the sole condition under which the whole of all ends can be brought into a unity under one principle. Or, in Kant’s own words in \textit{Critique of Practical Reason}, the moral law has under it the whole of all ends, for it “elevates a human being above himself (as a part of the sensible world)” (see CPrR 5:86-87; see also CPrR 5:132). The importance of this clarification to this chapter is remarkable, for if the moral law prescribes an end \textit{subjectively}, it would contradict the fundamental concept of freedom and render impossible the whole project of deriving the law of right from the moral law.

If we consider the fact that actions are related to the effecting of ends in general, the two limiting conditions with regard to the form and the matter of a maxim are in fact one and the same condition. In essence, the limiting condition of the freedom of action (Gr 4:431) and the limiting condition of “all \textit{merely

\textsuperscript{21} In the \textit{Critique of Practical Reason}, Kant claims that “though the highest good may be the whole object of a pure practical reason, that is, of a pure will, it is not on that account to be taken as its determining ground, and the moral law alone must be viewed as the ground for making the highest good and its realization or promotion the object” (CPrR 5:109). In Kant, the highest good is the \textit{whole} object of a pure practical reason, while humanity is an objective end. However, whichever object it is, Kant is unequivocally clear that the moral law is merely formal and is by itself the sole determining ground of the will.
relative and arbitrary ends” (Gr 4:436) are both depicting the same “sole condition under which a will can never be in conflict with itself” (Gr 4:437). The next question is then: what else can be added to this concept of limiting condition by considering the third element of a maxim, i.e. its complete determination? In *Groundwork*, Kant holds that the moral imperative with respect to the complete determination of all maxims is that “all maxims from one’s own lawgiving are to harmonize with a possible kingdom of ends as with a kingdom of nature” (Gr 4:436). This statement points to an imperatival implication of the moral law that might be relatively inconspicuous in the discussion of the last two elements, namely, its implication in relation to the aim and the development of the use of reason. To highlight the peculiarity of this implication, let us read a passage from the *Critique of Pure Reason*:

> The aim of reason with its ideal is, on the contrary, a thoroughgoing determination in accordance with a priori rules; hence it thinks for itself an object that is to be thoroughly determinable in accordance with principle… (A571/B599)

And another from *Idea for a Universal History*:

> In the human being (as the only rational creature on earth), those predispositions whose goal is the use of his reason were to develop completely only in the species, but not in the individual. Reason in a creature is a faculty of extending the rules and aims of the use of all its powers far beyond natural instinct, and it knows no boundaries to its projects. (IUH 8:18-19)
As noted above, the moral law is the law through which reason determines the will immediately (by the condition of a universal lawful form of its maxims). Through its own reality, the law “is able for the first time to give objective though only practical reality to reason” (CPrR 5:48). The aim of reason and the development of the use of reason therefore also obtain *objective though only practical reality* through their connection with the moral law. Their implications on the moral law are noticeably reflected in the FKE, which reads:

*Formula of the Kingdom of Ends* (FKE): “act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends” (Gr 4:439; cf. Gr 4:433, 437, 438).

Its connection with the aim and the development of the use of reason will become clearer if we examine Kant’s concept of the kingdom of ends in detail:

By a kingdom I understand a systematic union of various rational beings through common laws. Now since laws determine ends in terms of their universal validity, if we abstract from the personal differences of rational beings as well as from all the content of their private ends we shall be able to think of a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and of the ends of his own that each may set himself), that is, a kingdom of ends, which is possible in accordance with the above principles. (Gr 4:433)

---

22 Admittedly, the first passage (A571/B599) is merely considering the aim of reason in its *theoretical* use (“it thinks for itself an object…”), but the aim of reason in its *practical* use is actually not different from this, for we have only one and the same reason (see CPrR 5:121).
In other words, the kingdom of ends is a whole of all ends in systematic connection which we can think of as possible only in accordance with the moral law. Or, put still differently, the kingdom of ends is the thoroughgoing determination of all maxims in accordance with the moral law. It is thus indubitably clear that the kingdom of ends is the object, or the aim, of reason in Kant’s practical philosophy, which Kant mentioned as early as in the Critique of Pure Reason (see A571/B599) as cited above, although its practical character was not yet developed there. That the kingdom of ends is the aim of reason is only further supported by Kant’s statement near the end of the Groundwork when he speaks of “the noble ideal of a universal kingdom of ends in themselves (rational beings)” and adds to it the remark that “we can belong as members only when we carefully conduct ourselves in accordance with maxims of freedom as if they were laws of nature” (see Gr 4:462-463); for this is essentially a reiteration that the aim of reason is attainable only when we conduct ourselves in accordance with the law of (practical) reason.

The aim of reason is accompanied by another equally important implication of the moral law: the development of the use of reason. In association with the teleological doctrine of nature, Kant in Idea for a Universal History argues that: 

All natural predispositions of a creature are determined sometime to develop themselves completely and purposively (IUH 8:18). As the predisposition to personality in human nature, “the susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (Rel 6:26) is thus also to be developed completely and purposively.23

23 This chapter cannot possibly discuss Kant’s teleological doctrine of nature. For the purpose of this chapter, it is enough to note that the will would contradict with itself if it is necessitated by the categorical imperative on one hand and does not seek to develop the
These imperatival implications are only illustrated more conspicuously in the FA and the FKE through their explicit reference to the aim of reason and an implicit one to the development of the use of reason. The FA and the FKE, as Kant rightly alleges, therefore have to do with the \textit{allness} or the \textit{totality} of the system of all ends made possible through the moral law (Gr 4:436). These implications are coherently in line with the other two formulae of the moral law, for the \textit{allness} or the \textit{totality} of the system of all ends is possible only in accordance with the limiting condition of all subjective ends and the limiting condition of the freedom of action in the FH and the FUL. Conversely, the limiting condition of all subjective ends and that of the freedom of action, when serving as a categorical imperative with respect to the human will, is by no means merely a passive theoretical test which awaits to be employed to test the permissibility of a maxim. Instead, as a categorical imperative with respect to the will, it is by itself the motive of actions effecting the aim of reason through the development of the use of reason and in accordance with the condition of a universal lawful form of all maxims.

Now since the moral law, or the categorical imperative with respect to the human will, requires that its limiting condition be applicable not only in a particular will but \textit{also} in the thoroughgoing determination of all maxims in a “kingdom of \textit{ends in themselves} (rational beings)” (Gr 4:462), the idea of a \textit{public will} emerges as the general legislative will (MM 6:318, 6:320) of a people and the whole of their ends. Such a notion of public will is thus derivable only from the idea of the union of a multitude of rational beings whose wills are limited by the same self-given limiting condition which renders such a union possible in the first corresponding predisposition to personality on the other.
place. In view of the peculiar nature of the concept of public will, there is a distinctive kind of lawgiving Kant needs to associate with it. In the next section, I will offer a comprehensive account of the idea of a public will and examine specifically why the law of right is especially important in the idea.

6. The Concept of Public Will and the Two Kinds of Lawgiving

The idea of a public will is the idea of a general legislative will (MM 6:318, 6:320) of a people and the whole of their ends. The idea follows from the moral imperatives of FA and FKE, which have to do with the complete determination of all maxims in accordance with the moral law. In essence, the concept of public will is therefore derived from the concept of a human will determinable by the self-given moral law, which first accounts for the sole limiting condition under which a will can be thoroughly consistent with itself in every human being; and then subsequently extends itself to account for the limiting condition of the freedom of action for the harmonious coexistence of all human beings in a possible kingdom of ends. The idea of a general legislative will is thus grounded on the lawgiving capability of reason in the human will to determine the will immediately by the condition of a universal lawful form of its maxims (see CPrR 5:48). In other words, the idea does not introduce another independent concept of lawgiving. As illustrated in the previous section, this extension of the moral law is by no means external to its intrinsic nature but is one of its inherent features, because reason by its nature “is a faculty of extending the rules and aims of the use of all its powers far beyond natural instinct, and it knows no boundaries to its projects” (see IUH 8:18-19). This claim can be confirmed by a cautious interpretation of the three formulae, especially the FA and the FKE, which are the moral imperatives in relation to the allness or totality of all maxims. It is also
further supported by considering Kant’s account of the aim of reason and the
development of the use of reason. The natural conclusion of the above discussion
is therefore that Kant’s political philosophy is an integral part of his moral
philosophy, or even his philosophy as a whole, because the concept of public will
arises from the concept of a will subject to the moral law. These considerations
also notably underscore the importance of the concept of public will, or the united
will of a people, in Kant’s political philosophy, for it is the only appropriate
collective to account for the specific limiting condition of the freedom of action in
the extended context of a union of a multitude of people.

Before we further consider the concept of public will in detail, let us first
prepare ourselves by considering Kant’s distinction of the two elements in all
lawgiving in the *Metaphysics of Morals*:

In all lawgiving (whether it prescribes internal or external actions, and
whether it prescribes them a priori by reason alone or by the choice of
another) there are two elements: first, a law, which represents an action
that is to be done as *objectively* necessary, that is, which makes the
action a duty; and second, an incentive, which connects a ground for
determining choice to this action *subjectively* with the representation of
the law. (MM 6:218)

The distinction of the two elements makes possible the subsequent distinction of
ethical lawgiving and juridical lawgiving:

All lawgiving can therefore be distinguished with respect to the
incentive (even if it agrees with another kind with respect to the action
that it makes a duty, e.g., these actions might in all cases be external).
That lawgiving which makes an action a duty and also makes this duty the incentive is ethical. But that lawgiving which does not include the incentive of duty in the law and so admits an incentive other than the idea of duty itself is juridical. (MM 6:218-219)

The above distinction makes possible the examination of the concept of a general legislative will in two different dimensions. On one hand, with respect to internal lawgiving to oneself, the law associated with the public will is not different from the fundamental moral law discussed in the previous sections. The significance of the internal lawgiving to the public will is threefold. First, in order to be a member of a universal kingdom of ends in themselves (rational beings), which is what a union of a multitude of people aspires to be, one must “carefully conduct oneself in accordance with maxims of freedom as if they were laws of nature” (Gr 4:462-463). Second, not unrelated to the first, the moral law requires that “every rational being must act as if he were by his maxims at all times a lawgiving member of the universal kingdom of ends” (Gr 4:438). Third, since the moral law in the above internal lawgiving is the law in accordance with which one can become a lawgiving member of the kingdom of ends, it is the foundation upon which a general legislative will is possible, for the concept of a general legislative will is nothing but “the concurring and united will of all, insofar as each decides the same thing for all and all for each” (MM 6:313-314). Or in other words, the defensibility of the concept of public will rests upon the capability of each to be a lawgiving member of a people in accordance with the moral law. Internal lawgiving is thus indispensable in the concept of public will. With respect to oneself, the lawgiving is therefore ethical, for the incentive is contained in the pure
On the other hand, with respect to a possible lawgiving enforceable on others, a human will subject to the moral law cannot be indifferent to the actions of other human beings, for the complete determination of all maxims by the moral law consists in a possible kingdom of ends which we can only achieve as a species but not as individuals. In other words, the thoroughgoing determination of all maxims requires not only that I myself as an individual belong to the kingdom of ends as a member, but also that others belong to the same. Yet, as noted earlier, “we can belong as members [to the kingdom of ends] only when we carefully conduct ourselves in accordance with maxims of freedom as if they were laws of nature” (see Gr 4:462-463). This legitimate expectation of reason on the lawful use of the freedom of action by others is therefore limited by the impossibility to require others (or be required by others) by coercion to make the moral law the incentive of their maxims at the same time. Any external lawgiving enforceable on others that can be seen as “the act of the public will” (see TP 8:294), therefore, has to be juridical in nature.

This leads us to consider a limiting condition of the freedom of action in the external relations of human beings in accordance with the moral law. This limiting condition, in Kant’s own words, is “the formal condition of choice that is to be limited in external relations in accordance with laws of freedom, without regard for any end (the matter of choice)” (MM 6:375). In other words, the limiting condition of the freedom of action in external relations represents the sole condition under which the external use of choice, or outer freedom, can be consistent with itself in accordance with the moral law. Since the moral law is the

---

24 The moral law lets oneself discover the sublimity of one’s own supersensible existence and subjectively effects respect for one’s higher vocation (see CPR 5:88; cf. Rel 6:26).
limiting condition of the freedom of action in general, as discussed earlier, this latter limiting condition in external relations has to be an integral part of the former, for the thorough self-consistency of the will requires both in the complete determination of all maxims.

In the *Metaphysics of Morals*, the aforementioned limiting condition is called *right* and is expressed as follows in Kant’s *universal principle of right*:

Any action is *right* if it can coexist with everyone’s freedom in accordance with a universal law, or if on its maxims the freedom of choice of each can coexist with everyone’s freedom in accordance with a universal law. (MM 6:230)

A few points should be noted in the principle of right. First, the limiting condition conveyed by the principle is unmistakably one that applies only to the external relations of human beings, for it is merely about the *coexistence* of everyone’s freedom. It does not require, as in the formal condition indicated in the FUL, that we can at the same time *will* that the maxim become a universal law (see Gr 4:421; cf. Gr 4:402). Second, the principle of right is not presented as an imperative, as in the case of the three formulae of the moral law, for in the external relations of human beings, as I argued above, *ethical or internal* lawgiving is out of the question. In *juridical or external* lawgiving, *incentive* is not represented as an integral part of the law, and hence the corresponding principle cannot be seen as an *imperative.*²⁵ Third, although the principle is not represented as an imperative, it nonetheless “lays an obligation on me, but it does not at all expect, far less

²⁵ I therefore disagree with Höffe when he interprets Kant’s principle of right as a categorical imperative of law (see Höffe 1989).
demand, that I *myself should* limit my freedom to those condition just for the sake of this obligation” (MM 6:231). This last claim might sound confusing at first sight, but its importance also lies in a correct understanding of this seemingly puzzling claim, for it best illustrates the peculiar nature of the principle of right. The claim is best explicated in Kant’s own words:

…we know our own freedom (from which all moral laws, and so all rights as well as duties proceed) only through the moral imperative, which is a proposition commanding duty, from which the capacity for putting others under obligation, that is, the concept of right, can afterwards be explicated. (MM 6:239)

Kant’s message is clear in this passage and echoes what I argued earlier in this section. The principle of right, as a law, is to put others (and to be put by others) under an obligation to act in a way conformable to the limiting condition of the freedom of action in our external relations. By itself, the law of right does not, and should not, contain any reference to the incentive behind the obligation. From this it therefore also follows that the law of right by itself does not even demand oneself to limit one’s own freedom just for the sake of this obligation. However, since the limiting condition of the freedom of action in our external relations is required by, and actually part of, the limiting condition of the freedom of action in general, which is represented by the moral law, every human being is demanded by the moral law to recognize the principle of right as a law to put others (including oneself, for one is also put by others) under obligation. Or in other words, the moral law and its internal lawgiving is why the principle of right is recognized as a law in the first place.
The above account of the relationship between the law of right and the moral law can then be compared with those of other commentators discussed in the second section of this chapter. In Gregor’s account, the necessity of the principle of right and its corresponding external lawgiving is largely presupposed as the starting point without any specific explanation (see Gregor 1993). From the principle of right and its external lawgiving, she argues that since right is a capacity to put another under obligation, Kant’s *Doctrine of Right* is dependent upon his analysis of obligation, i.e. his moral philosophy (see Gregor 1993, 52-53). However, since she never attempts to explain why practical reason and its moral law necessarily lead to the law of right, the status of the principle of right as a law prescribing an obligation remains dubious and unexplained.26 Gregor’s weakness is shared by Kersting and Mulholland for essentially the same reason. In his interpretation, Kersting argues merely that the practical necessity of the law of right presupposes Kant’s doctrine of the fact of reason and the ensuing thesis of the reality of transcendental freedom (see Kersting 1992, 347). Mulholland, on the other hand, similarly starts with an idea of a system of rights which necessarily involves universal laws and claims subsequently on that basis a connection with Kant’s moral doctrine because “in the analysis of the concept of moral law that Kant has developed, it has been shown to be analytic that persons can be morally subject only to laws that are universal” (Mulholland 1990, 173).27 None of them ever tried to explain how the law of right can be derived from the moral law.

26 As noted earlier, in the interpretive account detailed in this chapter, the principle of right is recognized as a law to *put others* under obligation because it represents the limiting condition of the freedom of action in our external relations and is thus required by the moral law, which is the limiting condition of the freedom of action in general.

27 Reiss (1991, 20) shares much of Mulholland’s arguments. Please see the review in the second section of this chapter.
In the second section of this chapter, I reviewed the interpretations of a few commentators who argue either that the *Doctrine of Right* can stand on its own, independently from Kant’s moral philosophy (see, for example, Pogge 2002, 153), or that the two are mutually independent (see, for example, Ripstein 2009; Wood 2002; Willaschek 1997, 2002). Now since in this section I have already explained the connection between Kant’s moral philosophy and the law of right in the *Doctrine of Right*, I am not going to discuss their weaknesses again here. For the purpose of this chapter, I am not going to discuss the substantial content of the *Doctrine of Right* either. Instead, I will return to the idea of a public will as the legislative will of the union of a multitude of people and reconsider whether the concept of right discussed in this section suffices by itself to explain all the implications of the idea of a public will in Kant’s philosophy.

7. **Positive and Negative Duties in Kant’s Political Philosophy**

Given the centrality of the concept of right in modern political philosophy, the *Doctrine of Right* is sometimes considered the (or even the only) appropriate piece of Kant’s writing to approach his political philosophy. For instance, Ripstein (2009) in his monograph on Kant’s legal and political philosophy dedicates each and every chapter to the problem of rights. Gregor considers it advantageous to approach Kant’s political philosophy through the *Doctrine of Right* for “we may expect to find [his political philosophy] integrated with the enterprise he began in the *Groundwork* and carried through over a period of some twelve years” (Gregor 1993, 51). Pogge goes so far as to designate the *Doctrine of Right* as Kant’s liberalism (see Pogge 2002, 135). This view is met with the Rawlsian view that the cornerstone of Kant’s political thoughts lies in his doctrine of autonomy, which Rawls himself adopts and adapts in his *A Theory of Justice* (Rawls 1999).
However, as I will discuss in the second chapter, Rawls’s understanding of autonomy does not at all reflect Kant’s own theory of autonomy. The endeavor to develop a comprehensive understanding of Kant’s political philosophy starting from his moral philosophy is therefore being led in the wrong direction if one attempts to read Kant through Rawls’s interpretive framework. Apart from these two approaches, we certainly see in the current literature much research on Kant’s specific political thoughts in his various political writings. However, these studies are mostly conducted within too narrow a context of a specific concept or idea in politics, without at the same time offering a holistic view of Kant’s political philosophy. If we consider Kant’s philosophy as a whole, it is often dubious as to whether these attempts, in the absence of a holistic view, can really do justice to Kant’s political thoughts in their interpretations. In the previous section, a specific aspect of the concept was examined, namely, the concept of right and external lawgiving in the concept of a general legislative will. In Kant’s view, however, the law of right determines merely what is permissible in our external relations:

If then my action or my condition generally can coexist with the freedom of everyone in accordance with a universal law, whoever hinders me in it does me wrong; for this hindrance (resistance) cannot coexist with freedom in accordance with a universal law. (MM 6:230-231)

That the law of right merely establishes what is permissible in our external relations has a very profound implication in the holistic interpretation of Kant’s political philosophy. Since the law of right is a formal principle, its peculiarity is best illustrated by a comparison with the FUL of the moral law, which has to do
primarily with the form of a maxim:

*Formula of Universal Law* (FUL): “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (Gr 4:421; cf. Gr 4:402).

As I discussed earlier in this chapter, many commentators pay little regard to the idea in the FUL that we have to “at the same time will that it become a universal law”. The FUL is not merely a “universalizability test” of maxims, as many commentators believe, but a moral imperative conveying the sole condition under which the will can be thoroughly consistent with itself. In essence, it commands not only that we *should not* act in accordance with a maxim that cannot be *thought* without contradiction as a universal law of nature. It also does not command merely that we *can* act in accordance with whatever maxims that can be *thought* without contradiction as a universal law of nature. What it commands is that we *should* act *only* in accordance with the maxim which we at the same time *will* that it becomes a universal law through our action. The distinction is extremely important, for it accurately shows that in Kant’s moral philosophy, the moral law has to do with the *determinability*, not *permissibility*, of the will by the condition of a universal lawful form of its maxims in accordance with a law of reason (see CPrR 5:48).

On the other hand, the equally formal principle of right does not concern itself with the *determinability* of the will (or the public will) through the formal condition specified by itself, as is evident from its detachment from the *incentive* of one’s maxim. The concept of right concerns merely the possible external

---

28 Please refer to the discussion in the fourth section of this chapter.
coexistence of one’s external use of freedom with everyone else’s. The law of right therefore can be regarded as a “permissibility test” through which one can determine whether a maxim or an action is permissible or not, or in other words, right or not. Also, since the concept of right prescribes only a limiting duty, i.e. a duty to restrain from acting in accordance with certain maxims, it can be said that “[t]he duty corresponding to right is always negative”, as Kant himself asserts in his Drafts for the Metaphysics of Morals (see KGS 23:246).²⁹ In view of this, the next question is then whether the concept of right, along with its negative duty of right, is all what the concept of public will is about? Does it represent in entirety the whole idea of a limiting condition of the freedom of action for the union of a multitude of people? Or put differently, is this how the idea of the human will subject to the categorical imperative extends itself to be the legislative will of a possible kingdom of ends?

In the previous section, before I examined the concept of right specifically, I distinguished between two kinds of lawgiving that are both present in the concept of public will. While the law of right is the law for all human beings in their external relations (see MM 6:449), i.e. a law that puts others under obligation, the moral law is the law for the will of every human being internally. Now since the law of right is not, and cannot be, the law in accordance with which the public will determines itself,³⁰ the moral law is the only law through which the public will determines itself. In other words, in contrast to what many might be tempted to believe by intuition, the public will, despite its being “public”, is not

²⁹ One can also compare this claim with Kant’s distinction of formal and material duties to oneself. Kant calls the former “limiting (negative) duties” and the latter “widening (positive) duties to oneself” (MM 6:419).

³⁰ The law of right is merely the law in accordance with which the public will limits itself.
determinable by the law of right, which puts others under obligation, but only by the fundamental moral imperative, which, according to Kant, “is a proposition commanding duty, from which the capacity for putting others under obligation, that is, the concept of right, can afterwards be explicated” (MM 6:239). This clarification gives rise to certain substantial implications for the interpretation of Kant’s moral and political philosophy. First and foremost, any attempt to separate Kant’s political philosophy from his moral philosophy or vice versa is thus rendered indefensible. What we have is a single enterprise of Kant’s practical, or moral, philosophy within which his political philosophy is an integral part.31

Second, on top of the negative duty in accordance with the law of right, it is now possible to determine a positive duty of every human being in political matters in accordance with the moral law.

In brief, the positive duty essentially follows from the moral imperative: “every rational being must act as if he were by his maxims at all times a lawgiving member of the universal kingdom of ends” (Gr 4:438). As noted in the previous section, the duty belongs to the concept of public will because a public will, or a general legislative will, is “the concurring and united will of all, insofar as each

31 My view should be distinguished from Palmquist’s in his Kant’s System of Perspectives (1993). Palmquist stipulates that Kant’s practical philosophy has two “sides”. First, Kant’s “practical standpoint”, as Palmquist calls it, “is concerned with individual morality, and aims at the realization of the “highest good”. Second, Kant’s political philosophy “is concerned with what could be called communal morality, and aims at the realization of the ‘right’ in a legal system” (Palmquist 1993, 252). Palmquist believes that “Kant includes both sides in [Metaphysics of Morals], but only the former in his specifically Critical works” (Palmquist 193, 252). The problem with Palmquist’s reading is twofold. First, Kant’s concept of right has its foundation on morality, but it is not morality (or communal morality), for it is not concerned with incentive. Kant’s concept of right is a concept of duty that belongs to the doctrine of morals (see MM 4:406-407). Second, Kant’s political philosophy should not be limited to the concept of right and the aim of the realization of the “right” in a legal system. There is indeed a concept of positive duty in Kant’s political philosophy. Please read the rest of this chapter for this point.
decides the same thing for all and all for each” (MM 6:313-314) and it is possible only when a particular will is determined by the above moral imperative. The positive duty in the concept of public will is therefore exactly the same duty commanded by the moral imperative, which is nothing but a variant of the moral law. In theory, therefore, the concept of positive duty in the idea of a public will can be derived from any of the three formulae of the moral law, for they are “at bottom only so many formulae of the very same law, and any of them of itself unites the other two in it” (Gr 4:436). However, given the following hint found in Kant’s Drafts for the Metaphysics of Morals, it is clear we should look to the FH, which corresponds to the matter of a maxim:

The duty corresponding to right is always negative. The duty corresponding to the end is always affirmative. (see KGS 23:246)

As discussed earlier in this chapter, the FH specifies the ground of the supreme practical principle as: rational nature exists as an end in itself (see Gr 4:428-429). The FH, as the principle of humanity as an end in itself, in turn constitutes the supreme limiting condition of all subjective ends (Gr 4:431). In order to unveil the positive duty commanded by the FH, however, we must look further than its imperatival implication as a limiting condition. In this regard, let us turn to a point that was not mentioned in detail before in this chapter, namely, the full effect of the FH in the human will:

Fourth, concerning meritorious duty to others, the natural end that all human beings have is their own happiness. Now, humanity might indeed subsist if no one contributed to the happiness of others but yet did not intentionally withdraw anything from it; but there is still only a negative
and not a positive agreement with *humanity as an end in itself* unless everyone also tries, as far as he can, to further the ends of others. For, the ends of a subject who is an end in itself must as far as possible be also my ends, if that representation is to have its *full* effect in me. (Gr 4:430)

This important passage appears in Kant’s explication of the FH when he tries to explain the concept of *meritorious duty to others*. As discussed earlier, the idea of a public will arises from the extension of the idea of a human will subject to the moral law to account for the limiting condition of the freedom of action for the harmonious coexistence of all human beings in a possible kingdom of ends. In the idea of such a general legislative will (see MM 6:318, 6:320), the *full* effect of the moral law must be assumed, for the idea of such a will must be one that is thoroughly determined in accordance with the moral law for it to be considered the legislative will in the possible kingdom of ends, or else we would be left with an idea of a more qualified higher will which is determined by the *full* effect of the law. The supreme practical principle in the concept of a public will, therefore, cannot command merely a *negative agreement* but also a *positive agreement* with the FH. In other words, the concept of *positive duty* in the idea of a public will must be the *positive agreement* with the FH rather than a mere *negative agreement*.

In order to illustrate the implication of this conclusion on the interpretation of Kant’s political thoughts, I shall end this section with an example. In *On the Common Saying*, Kant speaks of a universal principle by which a people has to appraise its rights *negatively*: “*What a people cannot decree for itself, a legislator also cannot decree for a people*” (TP 8:304). This principle clearly follows from
the negative concept of right discussed above, for it merely determines what cannot be decreed by a legislator but not how a legislator should decree for a people. According to the above concept of positive duty in the public will, which corresponds to a positive agreement with humanity as an end in itself, we can now say that it is the positive duty of a legislator to try, as far as he or she can, to further the ends of others when he or she decrees for a people. Since this duty can only follow from the moral law in the internal lawgiving of the legislator himself or herself, it is indeed also an ethical duty on the part of the legislator. Due to limit of space, I shall not discuss the positive and ethical duty on the part of the citizens. However, if we consider the fact that in the concept of public will, everyone is united under a general legislative will, which is nothing but “the concurring and united will of all, insofar as each decides the same thing for all and all for each” (MM 6:313-314), their positive duty is indeed not as different as one might expect. Notwithstanding their similarities, however, it is possible to further develop the concept of positive duty in the concept of public will by considering its implications on the negative duty of right. It is then also possible to discern more specific duties for legislators and citizens by taking into account their specific roles and interactions with each other. These studies require a series of papers and therefore cannot be done here.

8. Closing Remarks

The conclusion of this chapter does not really offer anything new to non-philosophers who manage to develop their faculty of appraising and uphold their fortunate simplicity (see Gr 4:404) against any attempt to deflect their judgments. It is unmistakably clear to the most common understanding that what we want from political philosophy is not merely the reassurance that humanity
shall subsist even when no one contributes to the happiness of others but yet does not intentionally withdraw anything from it (Gr 4:430). We do not need to be told either that we should not act on any principle which cannot be made a universal law without contradiction, for even the commonest understanding can tell that it simply does not work. The problem we are facing is that we know what does not work but yet we at times “mutually corrupt each other’s moral disposition and make one another evil” (Rel 6:94). This is therefore not a problem of ignorance on the part of ordinary people, but a “battle that every morally well-disposed human being must withstand in this life, under the leadership of the good principle, against the attacks of the evil principle” (Rel 6:93). We certainly see a significant portion of our fellow men who fail to contain their selfish animal inclination and try whenever they can to except themselves from the only workable principles for the whole human race (see IUH 8:23). In such a dreadful situation, all we need to do is to “[unfurl] a banner of virtue as rallying point for all those who love the good, that they may congregate under it and thus at the very start gain the upper hand over evil and its untiring attacks” (Rel 6:94).

If we can agree that this is what we need to do, then which principle should be the keystone of Kant’s political philosophy? Is it the negative principle of right? Or the same old moral law from his moral philosophy? Which one of these can remind us of the sublimity of our own existence and effects our respect for our higher vocation in our attempt to contain our selfish animal inclination? (see CPrR 5:88) Or, which one of these is the keystone of our (co)existence? I spent more than half of this chapter on the clarification of Kant’s account of the moral law before proceeding to the extended idea of a public will. In the examination of the idea of a public will, I spent more than half of it on the discussion of its negative principle of right. Only in the last section when I spoke of the full effect of the
moral law on us did I associate a positive and ethical duty with the idea of a public will. I hope this late mention would not mislead my reader to believe that it is not important, for this is what all the preceding discussions attempt to prove.

In this first chapter of the dissertation, I explained the inseparability of the moral and the political in Kant’s philosophy. In the next chapter, I will turn to Rawls’s theory and see why and how he separates the political from the moral. The contrast between Kant and Rawls is instructive because Rawls allegedly invokes Kant’s notion of autonomy as a starting point of his theory and yet he insists upon a rigorous separation between his own political conception of justice and Kant’s moral doctrine. Rawls’s aspiration therefore seems to suggest that in a certain interpretation of Kant’s philosophy, the political can indeed be separated from the moral. In the following chapter, however, I will argue that Rawls’s theory is built upon a misreading of Kant’s notion of autonomy and that the subsequent separation of the political from the moral is hardly defensible.
II The Covert Subversion of Kant’s Account of Practical Reason in Rawls’s Understanding of Morality

1. Introduction

In his *Political Liberalism*, Rawls stresses that a basic thesis of his political liberalism is that, for the conception of justice to serve as “a basis of a reasoned, informed and willing political agreement”, the conception should be, as far as possible, “independent of the opposing and conflicting philosophical and religious doctrines that citizens affirm” (9). With respect to this unusual aspiration, Rawls proposes a peculiar idea of a “freestanding” view of a political conception of justice, meaning of which is best explained in his own words:

While we want a political conception to have a justification by reference to one or more comprehensive doctrines, it is neither presented as, nor as derived from, such a doctrine applied to the basic structure of society, as if this structure were simply another subject to which that doctrine applied… it means that we must distinguish between how a political conception is presented and its being part of, or as derivable within, a comprehensive doctrine. I assume all citizens to affirm a comprehensive doctrine to which the political conception they accept is in some way related. But a distinguishing feature of a political conception is that it is presented as freestanding and expounded apart from, or without reference to, any such wider background. To use a current phrase, the political conception is a module, an essential constituent part, that fits into and can be supported by various reasonable comprehensive
doctrines that endure in the society regulated by it. This means that it can be presented without saying, or knowing, or hazarding conjecture about, what such doctrines it may belong to, or be supported by. (PL 12-13)

In other words, in Rawls’s view, a viable political conception of justice should be freestanding in the sense that it can be accepted and supported by a society without the citizens agreeing on its justifying doctrine at the same time. In “Justice as Fairness: Political not Metaphysical”, Rawls calls this strategy the “method of avoidance” (JFPM 231) because it aims to “avoid disputed philosophical, as well as disputed moral and religious, questions” (JFPM 230). The goal is admittedly formidable, for it aspires to offer an impartial political conception of justice over which political agreement could be expected. However, this chapter will start by arguing that Rawls’s understanding of morality defeats the purpose of his method of avoidance in his own theory of justice, because his understanding of morality notably serves as the philosophical premise of his theory and adopts a foundational role that is not particularly different from the role of any comprehensive doctrine. This chapter will then argue that this understanding of morality is manifested in Rawls’s design of the original position through his peculiar procedural interpretation, or misinterpretation, of Kant’s moral doctrine. In the comparison between Rawls’s and Kant’s views on morality, I will show that Rawls wrongly attributes to Kant two characteristic features of his

---

32 Galston (2008) also argues that Rawls’s political theory embodies a comprehensive theory. She stops short of arguing that the comprehensive theory contains an underlying theory of morality, but her arguments do strongly hint that it does. First, she points out that Rawls’s theory specifies what individuals enduringly desire, willingly adopt and freely endorse. Second, Rawls’s theory specifies what citizens must deeply believe. Third, Rawls formulates the original position to explain the principles of political life (see 1849-1850).
own account of the autonomy of the will which Kant himself could not possibly endorse. The two features, namely, the capability of principle-choosing and the concept of expression, puts the autonomy of the will in a false light from a Kantian perspective, for they covertly replace the self-lawgiving and self-determining capability of practical reason vindicated by Kant in his philosophy. Since the latter capability is the autonomy of the will and is essentially morality itself in Kant, Rawls’s understanding of morality, which is allegedly based on a procedural interpretation of Kant, amounts to a covert subversion of Kant’s account of practical reason and morality.

This covert subversion does not come with no cost and is related to three major attacks against Rawls’s theory.33 I will review the background and the relevant arguments in two of these three attacks and relate these debates with Rawls’s understanding of morality. It will be shown that Rawls’s understanding of morality is the direct cause of these disputes because his understanding mistakenly deprives practical reason of its essential capability. I will also briefly sketch the resolutions of these debates if Kant’s original account of morality, and hence of practical reason, is faithfully applied.

2. The Foundational Role of Rawls’s Understanding of Morality

The reliance of Rawls’s political liberalism on a specific philosophical understanding of morality does not come as a surprise. After all, an understanding of morality that grounds a theory of justice unquestionably amounts to a philosophical stance in its own right. Rawls’s political liberalism indirectly but noticeably presupposes his understanding of morality in several respects. First, his

33 Only two of them will be discussed in this chapter, for reasons I will present later in the relevant section.
distinction between moral philosophy and moral theory, which is most clearly explained in “The Independence of Moral Theory”, is grounded on his conviction that the study of “substantive moral conceptions” and the way “they are related to our moral sensibilities and natural attitudes” can proceed independently without the guidance of moral philosophy (IMT 286). The distinction and the independence of moral theory from moral philosophy set the stage for his theory of justice by offering the theoretical support necessary for the justification of his method of avoidance. Second, his puzzling distinction between moral conceptions and political conceptions is the keystone of his political liberalism, without which a correct interpretation of his theory would be virtually impossible. Third, his understanding of morality also informs the partial assimilation of Kant’s moral philosophy in his original formulation of Kantian constructivism, in which Rawls’s own views of morality are especially discernible when he intentionally or inadvertently departs from Kant’s original philosophical insights.

The three distinctions clearly constitute three fundamental premises of Rawls’s political conception of justice by distinguishing the “political” from the “moral”, thereby determining what are and what are not relevant in his peculiar notion of the “political” domain. Moral theory, as Rawls defines it, is relevant and moral philosophy is not. Political conception is relevant, while moral conception (at least in the “narrower” sense coined by Rawls) is not. Part of Kant’s moral philosophy is relevant, and the rest is not. These distinctions are philosophically determinate and their justifications cannot rest only on an understanding of the “political” but also on an understanding of the “moral”. In other words, in spite of

34 In fact, Rawls goes so far as to suggest that “…the further advance of moral philosophy depends upon a deeper understanding of the structure of moral conceptions and of their connections with human sensibility…” (IMT 302), rather than the other way around.
Rawls’s method of avoidance, his understanding of morality evidently serves as a philosophical foundation of his theory by grounding at least these three fundamental premises. Furthermore, since the first distinction between moral theory and moral philosophy offers the justification necessary for the method of avoidance, and Rawls’s understanding of morality justifies this distinction in the first place, Rawls cannot legitimately invoke his method of avoidance in defense of his understanding of morality, or else the whole argument would become a case of circular reasoning. Rawls’s understanding of morality is thus itself unavoidably subject to challenge with regard to its philosophical defensibility.\textsuperscript{35}

Admittedly, it is difficult to identify and give a satisfactory interpretive account of Rawls’s understanding of morality, for morality never enters as a distinct subject in his major works. In \textit{A Theory of Justice}, Rawls dedicates a whole chapter on “The Sense of Justice” (TJ 397-449), but the chapter is merely meant to be a study of “the special case” of “the course of moral development as it might occur in a well-ordered society realizing the principles of justice as fairness” (404) and is concerned with “the acquisition of the sense of justice” in support of the stability of a “well-ordered society” and a “just scheme” (see 397-405). Morality is not discussed in its own right in the chapter. Though the theme of the chapter is moral development, Rawls stresses that he does not want to commit himself to either one of the two conceptions of moral learning adopted by the empiricists and the rationalists, let alone any specific understanding of morality per se (see TJ 401-404). His method of avoidance is pursued even more thoroughly after his political turn. In PL, the word “morality” appears in the main

\textsuperscript{35} The unequivocal denial of the potential foundational role of any possible full-fledged moral doctrine in the political domain, as suggested by Rawls’s method of avoidance, also constitutes a far-reaching philosophical claim that could be subject to challenge.
text only five times. In these rare occasions where “morality” is mentioned, it is associated either with “personal morality” (15, 236), “the ideals and virtues of morality” (236), “political morality” (254) or Habermas’s opposition against Locke’s and Kant’s views (410). In the passages where “personal morality” and “the ideals and virtues of morality” are mentioned, Rawls merely intends to deny their roles in “political discussion of constitutional essentials and basic questions of justice” (PL 16) without elaborating any further as to the meanings of the two terms. Rawls does not acknowledge any specific role of “political morality” in his political conception of justice either, for when he mentions the term he claims that the “content of public reason is not given by political morality” (PL 254), again without any further explanation. Needless to say, in the last case where Rawls tries to explain Habermas’s view, Rawls’s own understanding of morality is not in sight either.

The only remaining clue to Rawls’s understanding of morality is thus his idea of “free and equal moral persons”, which is an important element in his theory both before and after his political turn. In A Theory of Justice, Rawls alleges that the members of a well-ordered society “are, and view themselves as, free and equal moral persons”, in the sense that:

---

In PL, which is often regarded as the fullest expression of his renewed form of conception of justice after his political turn, the idea of free and equal moral persons still plays a prominent and foundational role in his theory (see, for example, PL 257-288). In JFPM, however, Rawls, in accordance with his political turn, emphasizes that although this conception of the person is admittedly a moral conception, it is “adapted to a political conception of justice and not to a comprehensive moral doctrine” (232). The conception is thus “in effect a political conception of the person” (232). This chapter deals with the defensibility of his conception of the person as a moral conception, the latter claim therefore shall not interfere with the following discussion. For more about Rawls’s political turn, please see Why Political Liberalism: On John Rawls’s Political Turn (Weithman 2010).
…they each have, and view themselves as having, fundamental aims and interests in the name of which they think it legitimate to make claims on one another; and they each have, and view themselves as having, a right to equal respect and consideration in determining the principles by which the basic structure of their society is to be governed.

(TJ 475)

In order to determine the principles that could be chosen by the free and equal moral persons, Rawls introduces a peculiar justificatory methodology called reflective equilibrium and invokes the original position as the “device of representation” (see JFPM 401; PL 24-27, 35, 75) with an aim to depict “the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted” (TJ 18). The resultant principles which match our considered judgments, according to Rawls, are expected to be “involved in the formation of [our] regulative sentiment” (TJ 430), i.e. our sense of justice, which is one of the two moral powers proposed by Rawls. I do not attempt to give a complete interpretive account of this methodology here. For the purpose of this section, we merely need to note that, his idea of the original position is meant to be a device of representation to determine the principles that should be involved in the formation of our moral power of the sense of justice. Therefore, it is reasonable to assume that if we examine the grounds of the design of the original position, we should be able to determine Rawls’s understanding of morality.

A useful hint to the grounds of the design of the original position is Rawls’s assertion that the “original position may be viewed… as a procedural interpretation of Kant’s conception of autonomy and the categorical imperative
within the framework of an empirical theory” (TJ 226). Actually, Rawls is always unhesitating, at least in *A Theory of Justice* and “Kantian Constructivism in Moral Theory”, to acknowledge the Kantian roots of his conception of justice as fairness. In *A Theory of Justice*, Rawls proclaims that “there is a Kantian interpretation of the conception of justice from which this principle [of equal liberty] derives” and that “[t]his interpretation is based upon Kant’s notion of autonomy” (TJ 221). In KC, Rawls even makes it the theme of the whole paper to “set out more clearly the Kantian roots” of the conception of justice of fairness (515). It is thus reasonable to hold that Rawls’s procedural interpretation of Kant’s conception of autonomy should be a good account of his understanding of morality. Yet, a more complete picture requires a clear distinction between Kant’s moral philosophy and Rawls’s “procedural interpretation” of Kant within the framework of an empirical theory. Only after such distinction is made can we see the most distinctive features of Rawls’s understanding of morality. In the next section, I will argue that the distinction between the two is so substantial that Rawls’s understanding of morality is not Kantian in any defensible way.

3. **Rawls’s Misinterpretation of Kant’s Autonomy of the Will**

   Very early in his elucidation of the idea of the original position, Rawls asserts that:

   It also turns out to be desirable to characterize the original position so that the parties are to choose principles that hold unconditionally whatever the circumstances. This fact is connected with the Kantian interpretation of justice as fairness. But I leave this matter aside until later (§40). (TJ 108-109)
The efficacy of the Kantian interpretation of justice as fairness is thus crucial to the overall efficacy of the original position, for the interpretation shall account for how “the parties are to choose principles that hold unconditionally whatever the circumstances”. The Kantian interpretation of justice as fairness also clearly assumes a foundational role in the design of the original position and the interpretation should reflect Rawls’s understanding of morality, as discussed in the previous section. However, when Rawls clarifies his Kantian interpretation of the original position near the end of §40 in *A Theory of Justice*, he reminds us that:

…the Kantian interpretation is not intended as an interpretation of Kant's actual doctrine but rather of justice as fairness. Kant’s view is marked by a number of deep dualisms, in particular, the dualism between the necessary and the contingent, form and content, reason and desire, and noumena and phenomena. To abandon these dualisms as he understood them is, for many, to abandon what is distinctive in his theory. I believe otherwise. His moral conception has a characteristic structure that is more clearly discernible when these dualisms are not taken in the sense he gave them but recast and their moral force reformulated within the scope of an empirical theory. What I have called the Kantian interpretation indicates how this can be done. (TJ 226-227)

In this short passage, Rawls is telling us two beliefs, or claims, of his. First, as is admitted by Rawls in the passage, he believes that Kant’s moral “conception” has a “characteristic structure” that can be “recast and their moral force reformulated within the scope of an empirical theory”. This belief, I believe, is one that Rawls attempts to justify by the overall efficacy of his theory. Second,
Rawls makes the seemingly self-evident claim that “Kant’s view is marked by a number of deep dualisms” without any elaborations here or in any of his writings. By using wordings like “abandon”, “recast” and “reformulated”, Rawls notably avoids making further philosophical assessment of Kant’s original moral philosophy and downplays the philosophical significance of his partial adaptation of Kant’s view in his empirical theory. The dualisms are simply named and then abandoned. This is in line with his method of avoidance, for a direct critique of Kant’s philosophical doctrine would more conspicuously amount to a philosophical stance which Rawls certainly prefers to avoid. However, such selective or partial assimilation, no matter how it is worded or presented, nonetheless suggests that Rawls identifies certain faults or deficiencies in Kant which he finds necessary to avoid. Rawls never tries to justify this avoidance in his writings, and interestingly I know of no interpreters who ever criticize him for the lack of justification either. The only explanation for this unusual state of affairs is that, the so-called dualisms identified by Rawls do not come from nowhere but from a thinker as influential as Hegel. In KC, Rawls hints at a possible connection between his own, Dewey’s and Hegel’s view on Kant:

We tend to think of [John Dewey] as the founder of a characteristically American and instrumental naturalism and, thus, to lose sight of the fact that Dewey started his philosophical life, as many did in the late nineteenth century, greatly influenced by Hegel; and his genius was to adapt much that is valuable in Hegel’s idealism to a form of naturalism congenial to our culture. It was one of Hegel’s aims to overcome the

37 The same passage on Kant’s various dualisms also appears in his “A Kantian Conception of Equality” (264).
many dualisms which he thought disfigured Kant's transcendental idealism, and Dewey shared this emphasis throughout his work, often stressing the continuity between things that Kant had sharply separated. This theme is present particularly in Dewey’s early writings, where the historical origins of his thought are more in evidence. In elaborating his moral theory along somewhat Hegelian lines, Dewey opposes Kant, sometimes quite explicitly, and often at the same places at which justice as fairness also departs from Kant. Thus there are a number of affinities between justice as fairness and Dewey’s moral theory which are explained by the common aim of overcoming the dualisms in Kant’s doctrine. (KC 516)

The whole passage is quoted here, for Rawls is very careful with his wording. While Rawls readily acknowledges the Hegelian root of Dewey, he merely admits that “there are a number of affinities” between himself and Dewey, and the “affinities” are limited to the places where they both depart from Kant. Rawls therefore does not admit or deny any indirect Hegelian roots of his view on Kant. I do not intend to pursue any further the similarities, or “affinities” as Rawls calls it, between the three thinkers, for such an analysis would lead to nothing more than a historical account that is unnecessary yet controversial.38 Instead, I merely want to point out that Rawls’s attempt to recast Kant’s moral philosophy “within the scope of an empirical theory” is in fact grounded on and informed by a

38 In a footnote that appears in the middle of the above citation, Rawls points his readers to two early writings of Dewey, namely “Outlines of a Critical Theory of Ethics” (Dewey 1891) and “The Study of Ethics: A Syllabus” (Dewey 1894) for a glimpse of Dewey’s critique of Kant (Dewey 1891, 290-300) and Dewey’s statement of his own form of the self-realization doctrine (Dewey 1894, 300-327) respectively (see KC 516). A lot of “affinities” between Dewey’s and Rawls’s view can be observed in these writings.
critique of Kant’s view on morality that is at least as philosophically fundamental as those of Hegel and Dewey. If we acknowledge this fact, we might reasonably anticipate certain fundamental differences between Rawls’s and Kant’s account of morality that could be as significant as the differences between Kant, Hegel and Dewey.

Near the end of the last section, I argued that Rawls’s understanding of morality is manifested in the design of his idea of the original position and that, according to Rawls, the “original position may be viewed... as a procedural interpretation of Kant’s conception of autonomy and the categorical imperative within the framework of an empirical theory” (TJ 226). Yet, now we are also prompted to believe that Rawls’s understanding of morality might be significantly different from Kant’s account of morality. If both statements are true, then the natural conclusion would be that Rawls’s “procedural interpretation” of Kant probably distorts Kant’s view to a substantial extent. In the rest of this section, I will examine a few discrepancies between Kant’s moral philosophy and Rawls’s procedural interpretation adopted in A Theory of Justice. We will see that these discrepancies do not originate from conscious departures from Kant but from an unintended fundamental misinterpretation of Kant’s view.

In §40 of A Theory of Justice, a few pages before Rawls reminds us of the dualisms he identifies in Kant, Rawls concludes his discussion of Kant’s notion of autonomy by stating that:

Kant held, I believe, that a person is acting autonomously when the principles of his action are chosen by him as the most adequate possible expression of his nature as a free and equal rational being. The principles he acts upon are not adopted because of his social position or
natural endowments, or in view of the particular kind of society in which he lives or the specific things that he happens to want. To act on such principles is to act heteronomously. Now the veil of ignorance deprives the persons in the original position of the knowledge that would enable them to choose heteronomous principles. The parties arrive at their choice together as free and equal rational persons knowing only that those circumstances obtain which give rise to the need for principles of justice. (TJ 222)

Here Rawls holds that the veil of ignorance in the original position is directly derived from what he believes to be Kant’s notion of autonomy. His belief, however, apparently stems from a fundamental misreading of Kant’s notion of autonomy, for his interpretation fails to affirm the most essential element in Kant’s doctrine of freedom: the capability of pure reason itself to legislate unconditionally in the faculty of desire, which is the single most important and fundamental insight Kant repeatedly emphasizes in his different writings:

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? (Gr 4:446-447)

39 The reliance of a moral theory on a notion of autonomy is unusual for an analytic political philosopher like Rawls, even from his own perspective. In his IMT, Rawls himself notes that: “Moral theory is the study of how substantive moral conceptions, that is, the study of how the basic notions of the right, the good, and moral worth may be arranged to form different moral structures” (IMT 286). This is in line with Schneewind’s observation that “…analytic moral philosophy generally [does] not give autonomy any special place” (2013, 152).
...for, freedom and the will’s own lawgiving are both autonomy and hence reciprocal concepts... (Gr 4:450)

This Analytic shows that pure reason can be practical - that is, can of itself, independently of anything empirical, determine the will - and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds. (CPrR 5:42)

For, this moral law is based on the autonomy of his will, as a free will which, in accordance with its universal laws, must necessarily be able at the same time to agree to that to which it is to subject itself. (CPrR 5:132)

“The will’s property of being a law to itself”, “the will’s own lawgiving”, the will determining itself independently of anything empirical and the autonomy of the will in accordance with its own universal laws are clearly different expressions of the capability of the pure reason itself to legislate unconditionally in the faculty of desire.40 These are only some of the shorter passages in which Kant explicitly talks about the autonomy of the will. If we consider the fact that freedom and the autonomy of the will are one and the same thing (Gr 4:450) and include those about the freedom of the will, we will end up with many pages of citations in which Kant explains the same idea repeatedly in his different writings. In fact, the

---

40 In Kant, the will is the faculty of desire (see, for example, CPJ 5:172) and the will is nothing but reason in its practical use, i.e. practical reason: “Everything in nature works in accordance with laws. Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will. Since reason is required for the derivation of actions from laws, the will is nothing other than practical Reason” (Gr 4:412).
capability of pure reason to legislate in the faculty of desire is the most important subject matter of the *Critique of Practical Reason*, or even the whole critical philosophy. In the introduction of the second *Critique*, Kant unequivocally asserts that the first question in his moral philosophy is “whether pure reason of itself alone suffices to determine the will or whether it can be a determining ground of the will only as empirically conditioned” (CPrR 5:15). In the Second Critique, Kant answers this question in the affirmative, for the property of the will’s being a law to itself is proved by “what it does” (CPrR 5:3) in its own lawgiving, which we become immediately conscious of “as soon as when we draw up the maxims of the will for ourselves” (see CPrR 5:29), for it prescribes the a priori necessity that cannot be derived from experience alone in the absence of an unconditional moral law. Once the reality of this property of autonomy of the will is established as such, as Kant notes in the preface of the Second Critique, “the concept of freedom, insofar as its reality is proved by an apodictic law of practical reason, constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason” (CPrR 5:3-4).

Recall once again what Rawls believes to be Kant’s view of the autonomy of the will: “a person is acting autonomously when the principles of his action are chosen by him as the most adequate possible expression of his nature as a free and equal rational being” (TJ 222). Contrary to Kant’s original notion of autonomy that rests upon the self-lawgiving and the self-determination of the will, Rawls endorses a notion of autonomy that focuses on the choice of principle and the expression of one’s own nature as a free and equal rational being. The capability to legislate unconditional law gives place to an apparent capability to choose from different principles, and the direct self-determination of one’s own will gives place to a concern merely with the expression of one’s own nature through the
chosen principles of action. Nowhere in his writings does Kant seem to endorse such an interpretation of the autonomy of the will, and nowhere in his TJ, KC or PL does Rawls seem to endorse Kant’s original notion of the autonomy of the will either.\footnote{In some passages in his \textit{Lectures on the History of Moral Philosophy} (2000) and “Themes in Kant’s Moral Philosophy” (1989), Rawls seems to be aware of Kant’s original notion of the autonomy of the will. However, when he tries to interpret Kant’s moral philosophy with his unique procedural approach, his misreading is conspicuous. The misreading will be evident by the end of this chapter.} The next question is then: whether Kant could possibly approve Rawls’s notion of autonomy\footnote{Admittedly, as Shell rightly notes, Rawls invokes “autonomy” in at least three different senses, namely, \textit{rational autonomy}, \textit{full political autonomy} and \textit{moral autonomy} (see Shell 2009, 5). Here in this chapter when I speak of Rawls’s notion of autonomy, I am always referring to \textit{moral autonomy}, which is also \textit{full autonomy} in Rawls.} if Kant were presented with such an interpretation? This is the question I am trying to answer in the next two sections.

4. Lawgiving or Choosing Principles?

In contrast to what Rawls believes to be Kant’s view, Kant unmistakably argues that the notion of the autonomy of the will lies not in the ability of our power of choice to choose from different principles, but in the capability of the will, which is practical reason, to be of itself practical, and hence independent from being determined by sensible impulses. Admittedly, apart from the notion of the autonomy of the will, Kant also speaks of the autonomy of the power of choice in some occasions. However, even with the power of choice, Kant unfailingly maintains that its autonomy, or freedom, is only inferred from its determinability through the moral law as unconditional command, and that the principle of autonomy is to choose only in such a way that the maxims of the choice are also included as universal law in the same volition:
We can quickly be convinced that the concept of freedom of the power of choice does not precede in us the consciousness of the moral law but is only inferred from the determinability of our power of choice through this law as unconditional command. (Rel 6:50)

Notwithstanding the somewhat misleading translation of Kant’s *Willkür* to “the choice” or “the power of choice” in English (see, for example, the discussions in *Practical Philosophy*, 89 and Allison 1990, 129-136), one must not figure that its autonomy lies in its power to choose principles. If one avers that Kant may allow a notion of autonomy like this, as Rawls seemingly assumes, Kant foresaw that and explicitly rejects the idea of defining freedom of the power of choice as an “ability” to make a choice for or against the moral law in *The Metaphysics of Morals*:

We can also see that freedom can never be located in a rational subject’s being able to choose in opposition to his (lawgiving) reason, even though experience proves often enough that this happens (though we still cannot comprehend how this is possible). - For it is one thing to accept a proposition (on the basis of experience) and another thing to make it the *expository principle* (of the concept of free choice) and the universal feature for distinguishing it (from *arbitrio bruto s. servo*); for the first does not maintain that the feature belongs necessarily to the concept, but the second requires this. - Only freedom in relation to the internal lawgiving of reason is really an ability; the possibility of deviating from it is an inability. How can the former be defined by the latter? It would be a definition that added to the practical concept the
exercise of it, as this is taught by experience, a *hybrid definition* (*definitio hybrida*) that puts the concept in a false light. (MM 6:227)

Kant is not thereby rejecting the obvious empirical fact that a rational subject does sometimes choose in opposition to his own lawgiving reason. He merely holds that the fundamental concept of the autonomy of the will, which includes the power of choice, requires that the concept of free choice be defined by its determinability through the universal law as unconditional command (Rel 6:50), or else the capability of the power of choice will be put “in a false light”. Or in other words, a genuine concept of autonomy, whether we are talking about the autonomy of the will or the power of choice, cannot be defined by anything taught *merely* by experience, for this directly contradicts the very idea of Kantian autonomy. In philosophy, we need to determine the *expository principle* of the concept of free choice. The *hybrid definition* given by Rawls, on the other hand, is not a philosophically tenable definition but merely an empirical proposition.

Rawls’s misreading is even more noticeable when he cites an objection against Kant by Sidgwick. According to Rawls, Sidgwick associates Kant’s notion of autonomy with choice-making, and thus “it seems to him on Kant’s view the lives of the saint and the scoundrel are equally the outcome of a free choice (on the part of the noumenal self) and equally the subject of causal laws (as a phenomenal self)” (see TJ 224). Rawls explicitly supports the criticism by Sidgwick, ironically one of the most important classical utilitarian philosophers whom he seeks to oppose, when he says “Kant never explains why the scoundrel does not express in a bad life his characteristic and freely chosen selfhood in the

43 “Insofar as reason can determine the faculty of desire as such, not only *choice* but also mere *wish* can be included under the will” (MM 6:213).
same way that a saint expresses his characteristic and freely chosen selfhood in a
good one” (TJ 224). This objection confirms Rawls’s misinterpretation on one
hand, and manifests the deficiency of Rawls’s understanding of Kantian morality
on the other. In Kant’s philosophy, the autonomy of the will, the freedom of the
will and the moral capability of the will are one and the same thing, and the artful
question raised by Rawls and Sidgwick reveals nothing but their misreading. As a
sympathizer of Kant’s philosophy, however, Rawls feels prompted to propose a
fictitious but ungrounded defense by Kant against this objection he raises. This
leads to another misinterpretation discussed in the following section.

5. Determination or Expression?

Failing to identify a defense in Kant against his own unfounded criticism,
Rawls proposes a potential defense as follows:

Kant’s reply must be that though acting on any consistent set of
principles could be the outcome of a decision on the part of the
noumenal self, not all such action by the phenomenal self expresses this
decision as that of a free and equal rational being. Thus if a person
realizes his true self by expressing it in his actions, and if he desires
above all else to realize this self, then he will choose to act from
principles that manifest his nature as a free and equal rational being. The
missing part of the argument concerns the concept of expression. Kant
did not show that acting from the moral law expresses our nature in
identifiable ways that acting from contrary principles does not. (see TJ
224)
I do not attempt to consider why Rawls is so concerned about the “expression” of our nature as a free and equal rational being in “identifiable ways”, or why he attributes a new interpretation of the noumenal and the phenomenal to Kant that is clearly different from Kant’s original account. As in the previous case, I will only examine whether Kant could possibly endorse Rawls’s reading. In contrast to Rawls’s belief that our autonomy must be expressed in identifiable ways in our actions, Kant holds that:

Whatever concept one may form of the freedom of the will with a metaphysical aim, its appearances, the human actions, are determined just as much as every other natural occurrence in accordance with universal laws of nature. (IUH 8:17)

According to Kant, therefore, expressing a decision in identifiable ways as that of a free and equal rational being, especially when considered as one made by the so-called phenomenal self, is thus never an issue, or even a wrong request to make. Outer experience, Kant says, “does not disclose the inwardness of the [moral] disposition but only allows inference to it, though not with strict certainty” (Rel 6:63). In fact, Kant makes it clear that it is impossible to express even to oneself “an entirely reliable cognition of the basis of the maxims which he professes, and of their purity and stability”, let alone to others:

Indeed, even a human being’s inner experience of himself does not allow him so to fathom the depths of his heart as to be able to attain,

---

44 These questions can be answered by a thorough examination of Rawls’s procedural interpretation of Kant’s moral philosophy in “Themes in Kant’s Moral Philosophy” (1989). This examination requires a separate paper and could be done independently.
through self-observation, an entirely reliable cognition of the basis of
the maxims which he professes, and of their purity and stability. (Rel
6:63)

The concept of expression is thus by no means a “missing part” in Kant’s
philosophy, as Rawls argues (see TJ 224). In Rawls’s misinterpretation, the
determining ground is the desire to express oneself as a free and rational being in
identifiable ways in one’s own action, and one has to choose one’s principles in
order to fulfill this desire. In Kant’s view, however, “all precepts of pure practical
reason have to do only with the determination of the will” and “the practical a
priori concepts in relation to the supreme principle of freedom at once become
cognitions and do not have to wait for intuitions in order to receive meaning”
(CPrR 5:66). No concept of expression is thus necessary or even allowed to enter
as the determining ground of the will as in Rawls’s misinterpretation. No higher
desire of expressing or realizing one’s true self is required to guide us in the
choice of principles either. In Kant, the concept of expression has no place but
must give way to “the necessary determination of the will a priori” (CPrR 5:5),
which is manifested not in any inner or outer expression in identifiable ways⁴⁵ but
in the inner consciousness of the moral law:

I ask instead from what our cognition of the unconditionally practical
starts, whether from freedom or from the practical law. It cannot start
from freedom, for we can neither be immediately conscious of this,

⁴⁵ In his philosophy, Kant alleges that conscience is an internal court in the human being
“before which his thoughts accuse or excuse one another” (MM 6:438), and that conscience “could
also be defined as the moral faculty of judgment, passing judgment upon itself” (Rel 6:186). In a
certain sense, conscience therefore offers a certain “expression” of practical reason holding the
human being’s duty before him.
since the first concept of it is negative, nor can we conclude to it from experience, since experience lets us cognize only the law of appearances and hence the mechanism of nature, the direct opposite of freedom. It is therefore the moral law, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves), the first offers itself to us and, inasmuch as reason presents it as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them, leads directly to the concept of freedom. (CPrR 5:29)

According to Kant, the consciousness of the freedom of the will and the consciousness of the moral law are one and the same thing (see CPrR 5:46), and the consciousness of the moral law is nothing but “the necessity with which reason prescribes to us and to the setting aside of all empirical conditions to which reason directs us” (CPrR 5:30) “as soon as we draw up maxims of the will for ourselves” (CPrR 5:29). This consciousness cannot be seen as an “expression” of our nature in identifiable ways. It is not “an entirely reliable cognition”, through self-observation, “of the basis of the maxims which [one] professes, and of their purity and stability” either (see Rel 6:63). The consciousness of the freedom of the will is merely the consciousness of our existence as determinable in an intelligible order of things:

…consciousness of freedom of the will, whereby the will of a rational being that, as belonging to the sensible world cognizes itself as, like other efficient causes, necessarily subject to laws of causality, yet in the practical is also conscious of itself on another side, namely as a being in
itself, conscious of its existence as determinable in an intelligible order of things - conscious of this not, indeed, by a special intuition of itself but according to certain dynamic laws that can determine its causality in the sensible world… (CPrR 5:42)

Kant unambiguously holds that there is no “special intuition” of ourselves, in identifiable ways, through which we can express our nature as a free and equal rational being. In his practical philosophy, the freedom, or autonomy, of the will is “identifiable” not in any “expression” but in the consciousness of the moral law and the accompanying consciousness of our existence as determinable in an intelligible order of things. In other words, the notion of autonomy in Kant is about the determination of our existence rather than the expression of our nature.

6. Three Criticisms Against Rawls

In the last three sections, I argued that Rawls’s understanding of morality is grounded on a misinterpretation of Kant’s notion of autonomy of the will. As I stated earlier in the introduction to this chapter, the misinterpretation comes with a high price tag, for it is related to three major attacks against Rawls’s theory. In this section, I will review three of the most important criticisms against Rawls. In the following sections, I will then show that these criticisms are actually the unavoidable consequences of the weaknesses of Rawls’s understanding of morality, and that they can be resolved by accommodating Kant’s concepts of self-lawgiving and self-determination of the will, rather than Rawls’s concepts of principle-choosing and expression.

While Rawls’s understanding of morality serves to justify his method of avoidance on the surface, the strategy ironically provokes philosophical criticisms
from different directions on its way to circumvent the battlefield of the competing comprehensive doctrines. The first wave of criticism comes most notably from the communitarians. While they are not worried by Rawls’s attempt to avoid competing comprehensive doctrines in the justification of his conception of justice, they are not convinced by the apparent independence of his justificatory strategy from social convention and shared understanding in the community. In order to show the indefensibility of Rawls’s approach, the communitarians invoke, as Gutmann (1985) notes, either Aristotelian or Hegelian ideas in their criticisms against Rawls. For instance, MacIntyre (2007) invokes the Aristotelian idea that justice is rooted in “a community whose primary bond is a shared understanding both of the good for man and the good of that community” (232-233) in his criticism of Rawls for the neglect of desert. The same Aristotelian idea also explicitly inspires Taylor (1979) in his attack on “atomistic” liberals, including Rawls, who “try to defend... the priority of the individual and his rights over society” (39). Unger (1975, 85, 191-231) and Sandel (1982, 179-180), on the other hand, seem to be inspired by the Hegelian conception of man as historically conditioned in their rejection of the liberal view of man as free and rational being adopted by Rawls. The force of these critiques is questionable though, for evidently, as Gutmann (1985) argues, they fail to appreciate the peculiar features of Rawls’s method of avoidance and his idea of a freestanding conception of justice. This chapter, however, will not review the history of the liberal-communitarian debate. Instead, I would like to point to a remarkable feature of the critiques raised by the communitarians, which is that they are

46 Amy Gutmann gives a very comprehensive review of the communitarian critics of Rawls in her “Communitarian Critics of Liberalism” (1985). Some of the following citations from the communitarian critics are quoted by Gutmann in her paper.
clearly inspired by certain Aristotelian ideas and Hegelian conceptions. If we consider the fact that Rawls’s method of avoidance and his justificatory strategy are informed by an understanding of morality that is also notably inspired by Aristotle and Hegel,\textsuperscript{47} then we are witnessing a debate in which both sides believe that they are at least partially following the insights of the same two great thinkers.

Friendly fire does not come only from the communitarian critics, but also from thinkers who are equally committed to the aspiration of a freestanding political conception of justice. The debate in this case is over the notorious problem of the unconstructed starting point of Kantian constructivism proposed by Rawls. In KC, Rawls alleges that Kantian Constructivism “holds that moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept”, and that “[a]part from the procedure of constructing the principles of justice, there are no moral facts” (519). In other words, the goal of Kantian constructivism is to construct the freestanding principles of justice from a freestanding unconstructed starting point. However, O’Neill, one of Rawls’s students, contends that Rawls fails to give a constructivist (and hence freestanding) justification for its starting point, i.e. the procedure of practical reason itself, but instead offers only a coherentist justification for the procedures (O’Neill 2003a, 323; O’Neill 2003b, 351, 357). In O’Neill’s view, Kantian constructivism is thus constructivist in a “limited” sense at best, since it fails to satisfy the constructivist

\textsuperscript{47} See the discussion in Section 3 of this chapter for the indirect and unacknowledged Hegelian root of Rawls’s understanding of morality. In TJ, Rawls explicitly acknowledges that his methodology to the conception of the subject (in which contingent assumptions and general facts are employed) goes back to Aristotle’s procedure in the \textit{Nicomachean Ethics} (see TJ 44-45). Rawls also admits that his theory of the good is a familiar one going back to Aristotle (TJ 79). In PL, Rawls’s invokes the “Aristotelian principle” in his moral psychology (203) to “spell out the ways in which the well-ordered political society of justice as fairness is intrinsically good” (207).
criteria of using only minimal and abstract premises or starting materials to construct principles (O’Neill 1996, 38–40). O’Neill’s accusation is shared by many other commentators of Rawls. For instance, Cohen (2008) argues that Rawls’s Kantian constructivism is “fact-infested” (287), while James (2005) asserts that Rawls assumes that constructivist moral judgment is authoritative only when “grounded in independent judgments about what kind of social practices exist and what kinds of agents participate in them” (282). These accusations fundamentally threaten the viability of the method of avoidance, for if a freestanding starting point cannot be determined, it is clearly impossible to construct a freestanding political conception of justice.

Another attack that is relevant to this chapter is the criticism against Rawls’s idea of public reason and overlapping consensus. In PL, Rawls significantly modifies his account of constructivism by distinguishing between Kant’s moral constructivism and his renewed political constructivism. In his political constructivism, Rawls maintains that there is an unavoidable dualism “between the point of view of the political conception and the many points of view of comprehensive doctrines” (xxi), and thus we should not “put forward more of our comprehensive view than we think needed or useful for the political aim of consensus” (153). Our comprehensive view, which is an exercise of practical reason according to Rawls (PL 59), thus has to be restrained in view of a necessity to respect “the limits of public reason”, without which we cannot possibly reach “an overlapping consensus on a conception of political justice” (PL 153). Rawls’s theory therefore posits an inevitable dualism between “private” practical reason and public reason. In view of this, Koukouzelis (2009) argues that Rawls’s public

reasoning is unreasonably “based on an alleged ‘method of avoidance’ from revealing our comprehensive doctrines in the public realm” and Koukouzelis suggests that public reasoning should be based, “in the opposite direction, on a conception of the public use of reason... which helps in disclosing, rather than concealing the principle of our deliberation on justice” (843). I agree with Koukouzelis’s view on this matter, but the arguments involved, I believe, should be situated in a more far-reaching discussion of Kant’s philosophy than what Koukouzelis offered in his paper. While I would like to argue that the dualism between “private” practical reason and public reason is caused by Rawls’s (mis)understanding of morality, a Kantian resolution of the dualism is not possible without a comprehensive and convincing elucidation of the necessary connection between Kant’s moral philosophy and political philosophy. I will therefore tackle the dualism in a separate paper in the future.

In fact, this chapter does not attempt to engage in these debates directly. Instead, I will argue in this chapter that the first and the second debates stem from the same problem in Rawls, namely, the deprivation of the self-lawgiving and self-determining capability of practical reason in Rawls’s account of autonomy and understanding of morality. The following sections will vindicate this claim by illustrating that Rawls’s understanding of morality is responsible for the dualism between the right and the good, and the indefensibility of the starting point for the conception of justice underlying the two debates respectively. A thorough discussion of each of these two debates certainly also merits an individual paper, but since the purpose of this chapter is to demonstrate the fundamental problem of Rawls’s understanding of morality, I will only give a brief sketch of the corresponding resolutions if Rawls is to adopt a more Kantian understanding of morality.
7. Dualism between the Right and the Good

The dualism between the right and the good is not a unique feature of Rawls’s theory but the focal point of controversy in the history of political philosophy ever since the introduction of the idea of inalienable natural right by the classical liberal thinkers as the foundation of their social contract theories. Natural right in classical liberalism is, as Hart (1955) puts it, “the equal right of all men to be free” (175). The natural right is “one which all men have if they are capable of choice; they have it qua men and not only if they are members of some society or stand in some special relation to each other” (Hart, 1955, 175). It is “not created or conferred by men’s voluntary action; other moral rights are” (Hart, 1955, 175-176). The theory of natural right thus assigns primacy to the right and does not specify the good independently from the right. The notion of the primacy of natural right, however, is denied by the utilitarians and the legal positivists in later eighteenth and early nineteenth century. Bentham (1987 [1816]), the most prominent critic at the time, famously repudiates the classical theories of natural right as nothing more than “nonsense, rhetorical nonsense, - nonsense upon stilts” (53) and “a perversion of language” (60). For Bentham, as Hart (1983) concludes, “the idea of a right not created by positive law was a contradiction in terms like ‘cold heat’ or ‘resplendent darkness’: rights, [Bentham] claims, are all fruits of positive law” (185), which is in turn guided by the principle of utility. In utilitarianism and legal positivism, therefore, primacy is assigned to the good and the right is not specified independently from the good.

49 Utilitarians reject the primacy of natural right because in utilitarianism the right cannot be understood independently from the utilitarian good, which is the good produced as the consequence of an action. In other words, an action is right only if it produces a good consequence.
The pendulum continues to swing between the right and the good in the debate between Rawls and his communitarian critics. On one hand, Rawls attempts to advance a freestanding political conception of justice to limit the various conceptions of the good (see PL 174-176). Priority is notably given to the right, for he claims that “comprehensive conceptions of the good are admissible, or can be pursued in society, only if their pursuit conforms to the political conception of justice” (PL 176). On the other hand, “the communitarian critics, unlike modern liberals, make the case for a politics of the common good” (Sandel 1984, 16). They “question the liberal claim for the priority of the right over the good, and the picture of the freely choosing individual it embodies” (Sandel 1984, 16). Sandel (1984) accuses the Kantians, by which he means to include Rawls, of defining the self as a “choosing self, independent of the desires and ends it may have at any moment” (17). MacIntyre (2007) goes so far as to claim that we should reject any doctrine of right the same way we reject the existence of witches (67). For the communitarians, as MacIntyre (2007) puts it, justice is rooted in “a community whose primary bond is a shared understanding both of the good for man and the good of that community” (232-233). In other words, in their view, the right cannot be specified independently from the good.

Centuries of debate have not left us with a satisfactory resolution of the dualism between the right and the good, and this could be partly why MacIntyre (2007) complains that “we have – very largely, if not entirely – lost our

50 In PL, Rawls holds that, despite the priority of the right, a political conception of justice must combine the right and the good in a definite way and therefore cannot neglect the latter (see PL 173-174). He therefore proposes “five ideas of the good found in justice as fairness” in PL (see PL 173-211). However, it is Rawls’s unmistakable position that the state should be neutral “with respect to comprehensive doctrines and their associated conception of the goods” (see PL 192-194). In other words, priority is always given to the right in Rawls’s philosophy.
comprehension, both theoretical and practical, of morality” (2). On the surface, the disagreement between Rawls and his communitarian critics might be seen as stemming from their different conceptions of the self. Rawls’s theory is grounded on a political conception of the self as free and equal that is “adapted to a political conception of justice” (PL 18). The adaptation is made possible by presupposing that part of our moral sensibility conforms to his idea of the reasonable (see PL 51). By specifying what the citizens endurably desire, willingly adopt, freely endorse and deeply believe, the idea of the reasonable serves as the all-encompassing idea that accommodates all the elements required in the subject for Rawls’s theory to succeed.  

Discontented with Rawls’s abstract and arguably dogmatic conception of the self, Sandel contends that the Rawlsian self is “an antecedently individuated subject, standing always at a certain distance from the interests it has” and is thus “beyond the reach of experience” (Sandel 1982, 62). Sandel’s view is shared by other communitarian thinkers and is the principal


52 Sandel goes on to argue that: “But a self so thoroughly independent as this rules out any conception of the good (or of the bad) bound up with possession in the constitutive sense. It rules out the possibility of any attachment (or obsession) able to reach beyond our values and sentiments to engage our identity itself. It rules out the possibility of a public life in which, for good or ill, the identity as well as the interests of the participants could be at stake. And it rules out the possibility that common purposes and ends could inspire more or less expansive self-understandings and so define a community in the constitutive sense, a community describing the subject and not just the objects of shared aspirations. More generally, Rawls’ account rules out the possibility of what we might call ‘intersubjective’ or ‘intrasubjective’ forms of self-understanding, ways of conceiving the subject that do not assume its bounds to be given in advance. Unlike Rawls’ conception, intersubjective and intrasubjective conceptions do not assume that to speak of the self, from a moral point of view, is necessarily and unproblematically to speak of an antecedently-individuated self” (Sandel 1982, 62).

53 See, for example, MacIntyre (2007): “…the individual is identified and constituted in and through certain of his or her roles, those roles which bind the individual to the communities in and through which alone specifically human goods are to be attained; I confront the world as a member of this family, this household, this clan, this tribe, this city, this nation, this kingdom. There is no
reason why they are not convinced by Rawls’s theory. The liberal-communitarian debate, therefore, can be seen as revolving around the idea of a legitimate conception of the self for the purpose of political philosophy.

This chapter does not attempt to engage in the debate on the conception of the self between the two camps, partly because of limit of space, but mainly because I am not convinced that the debate is the proper path to the resolution of the dualism between the right and the good. Ever since Rawls defines the notion of autonomy as choosing the principles of one’s action as the most adequate possible expression of his nature as a free and equal rational being (see TJ 222), the conception of the self as free and equal is assigned a foundational role in his theory of justice. Without reflecting on Rawls’s methodology thoroughly, his communitarian critics follow his way of thinking by centering the dispute around their differences on the conception of the self, which lead only subsequently to their different views on the priority of the right and the good. In essence, the liberal-communitarian debate arises because the right in Rawls is derived from his “moral” (in TJ) or “political” (in PL) conception of the self, while the good in communitarianism is determined by an empirical concept of the self. According to Kant, however, the debate might lead them nowhere because the foundation of their theories is misplaced since the very beginning.

---

54 Or consider the definition of Kantian constructivism as one with “a particular conception of the person as an element in a reasonable procedure of construction, the outcome of which determines the content of the first principle of justice” (KC, 516). The foundational role of the conception of the person is also unmistakable.
In response to the debate, Kant would say, I believe, the dualism cannot be resolved by adopting any conception of the self as the foundation not only of political philosophy but also of practical reason. First, in Kant’s view, the communitarian view that the self is unavoidably empirically conditioned by the good (whether it be the desires and the ends of oneself or the common good) unquestionably produces heteronomy:

…if one assumes any object under the name of a good as a determining ground of the will prior to the moral law and then derives from it the supreme practical principle, this would always produce heteronomy and supplant the moral principle. (CPrR 5:109)

Since the passage seems like it is especially targeted at the communitarians and since the following critique against Rawls applies equally well to them, I shall not pursue any further Kant’s probable response to the communitarians.

While the communitarians openly object to the idea of an autonomous self, Rawls’s notion of autonomy and the associated idea of the reasonable serves as the keystone of his political liberalism both before and after his political turn. Yet, according to Kant, Rawls’s conception of the self also produces heteronomy and is not a correct account of what human being is. In Kant’s view, a philosophical conception of the self, if he were to adopt this terminology, is possible only after a comprehensive examination of the principal capabilities and the essential characteristics of humans. Rawls, however, covertly reverses Kant’s approach by grounding the autonomy of the will on his conception of the self. In Rawls’s notion of autonomy, practical reason, or the will, of the subject is required to choose its principles of action as the most adequate possible expression of its
nature as a free and equal rational being (see TJ 222) in identifiable ways, or else, as is absurdly claimed by Rawls, practical reason is not autonomous. In other words, for Rawls, autonomy is autonomy dictated by his conception of the self to be expressed and identified. Practical reason is thus left with no other capacity than to choose the principles for the most adequate possible expression of a nature of the self determined by Rawls. Such an attempt, in Kant’s view, could be the most dangerous of all.

As I admitted earlier, this chapter can only give a brief sketch of the resolution of the dualism between the right and the good if we are to adopt a more Kantian understanding of morality. Now the proper path can be sketched as follows. In the liberal-communitarian debate, the dualism between the right and the good is not an inevitable dualism of the two concepts per se, but is caused by the contrasting conceptions of the self held by the two camps. If the dualism could ever be resolved at all, one must not precede an account of practical reason, or an understanding of morality, with any specific conception of the self. Instead, one must first examine the primary capability of practical reason, as Kant did, and then determine the concept of the right and the concept of the good based on the facts established by such an examination. Only when both concepts are determined by considering the same capability of self-lawgiving and hence self-determination of practical reason (in contrast to Rawls’s principle-choosing and concept of expression) shall the dualism vanish and cease to hinder the advance of political philosophy. There is simply no alternative.

8. Indefensibility of the Starting Point for Rawls’s Conception of Justice

As is evident from the discussion in the last section, the dispute on the
priority of the right and the good is essentially a dispute on the conception of the self in Rawls, which serves as the foundation, or starting point, of his constructivist theory. The dispute, therefore, is indeed a variant of a debate in a broader sense on the defensibility of the starting point in constructivism. Admittedly, this broader debate involves a set of complicated and interrelated considerations amidst the new possibilities and challenges brought about by the constructivist approach to the first principles of justice. Yet, their views can be roughly categorized as follows. First, we have the starting point of Rawls’s theory already discussed above. Second, we have the critics of constructivism who argue that constructivism cannot possibly avoid arbitrariness without relying on an unconstructed foundational basis (Cohen 2003; Raz 1990, 15; 2003; Timmons 2003) and hence “it is not a coherent, stable, and self-standing metaethical doctrine” (Bagnoli 2014, 312). Third, apart from the outright critics, we have other constructivists who attempt to propose alternative accounts of constructivism. For instance, O’Neill, being the most prominent among them, proposes her own starting point for constructivism because she believes that Rawls’s constructivism fails to satisfy the constructivist criteria of using only minimal and abstract premises or starting materials to construct principles (O’Neill, 1996, 38–40). This section shall focus on the challenge posed by the third type of critics. Since O’Neill’s constructivism is often considered the most systematic and philosophically rigorous attempt to develop a more universal alternative to Rawls’s (see Sutch 2001, 86; Budde 2009), I will focus specifically on O’Neill’s challenge against Rawls in this section.

The first thing to note is that O’Neill admits for constructivism a more ambitious metaethical role of offering an alternative to the realist and the relativist views, which Rawls seeks merely to avoid (see O’Neill 1988, 1). This apparently
more ambitious role is, in my opinion, just a more accurate account of what Rawls’s constructivism aspires to achieve in the first place, and the move marks O’Neill’s intention to hold constructivism answerable to the philosophical questions Rawls merely leaves aside. As noted above, O’Neill’s primary contention against Rawls is her dissatisfaction with the starting point of Rawls’s constructivism. Unlike Rawls’s communitarian critics, who contend that Rawls’s conception of the self is too abstract, O’Neill argues that Rawls’s starting point involves an ideal of the person that is too determinate to be philosophically defensible (see O’Neill 1989, 212). The diagnosis is accompanied by a brief but remarkable defense for the method of abstraction in ethical reasoning (see O’Neill 1989, 208-209) found in Rawls and in her variant constructivism. O’Neill’s objection is therefore specifically targeted at the idealization, rather than the abstraction, found in Rawls’s conception of the person.

O’Neill’s diagnosis is supported by her insightful observations of Rawls’s conception of the original position and the veil of ignorance. O’Neill forcefully argues that Rawls actually does “more than abstraction behind the construction of the original position”, for the ways he abstracts in the design of the veil of ignorance are governed by a certain ideal (O’Neill 1989, 209). In other words, as O’Neill puts it, “[t]he artful tailoring of the veil of ignorance is determined by a highly selective abstraction from actual human choosing, which reflects a certain ideal of the human subject”; and the ideal “is used not just to determine how ‘thick’ or ‘thin’ the veil of ignorance should be, but rather to tear some carefully placed holes in it” (O’Neill 1989, 209). O’Neill’s position is thus unmistakable. “Idealization masquerading as abstraction”, in O’Neill’s view, “produces theories that may appear to apply widely, but in fact covertly exclude from their scope those who do not match a certain ideal” (O’Neill 1989, 210).
O’Neill’s variant constructivism is primarily an attempt to reinvent the constructivist theory of justice with abstract starting points that are free from any idealization or moral ideals of a specific culture (O’Neill 1989, 212). O’Neill’s starting point is, as she puts it in “Constructivism in Ethics” (1989, 206-218), “a meager and indeterminate view of rationality” which credits agents only with “the capacity to understand and follow some form of social life, and with a commitment to seek some means to any ends (desired or otherwise) to which they are committed” (212). In Towards Justice and Virtue (O’Neill 1996) and Bounds of Justice (O’Neill 2000), O’Neill further elaborates her starting points,55 which seemingly do not rely on any idealization or moral ideals of a specific culture. In her constructivism, however, O’Neill also acknowledges that “[a]n adequate account of ethics will need not only convincing starting points but convincing ways of proceeding from those starting points, that is to say an adequate conception of practical reasoning” (O’Neill 1996, 39). This additional element clearly complicates the problem of the defensibility of her constructivism. At least, one may wonder why the “adequate conception of practical reasoning” is not one of the starting points in O’Neill’s constructivism. One may also wonder why the “meager and indeterminate view of rationality” is separated from “the adequate conception of practical reasoning”. The answers to these two questions might not

---

55 In these two works, as Budde (2009, 202-203) notes, O’Neill “is emphatic in her insistence to use only weak, uncontroversial, merely abstract and empirically realistic starting points and material” (see O’Neill 1996, 7, 48, 179, 210), which include “a sober and accurate, indeed critical, view of the builders within the relevant domain(s) and of the materials and skills to be available to them” (O’Neill 1996, 62), “an abstract account of others’ capacities, capabilities and vulnerabilities” (O’Neill 1996, 64), “an empirically realistic view of the capacities and capabilities agents have, of ways in which they are vulnerable to others” (O’Neill 2000, 7) and “materials available in all contexts (e.g. the results of instrumental reasoning and of empirical inquiry, knowledge of social realities and human abilities)” (O’Neill 1996, 210).
be very important if O’Neill’s conception of practical reasoning is also as weak, uncontroversial, merely abstract and empirically realistic (see note 21) as is required for her starting points. However, if this is not the case, O’Neill would fail her own constructivist requirements that ground her attack on Rawls’s constructivism, though the weakness now lies not in the starting points but in her conception of practical reasoning.

In O’Neill’s constructivism, the “adequate” conception of practical reasoning is called the “Kantian” or “Critical” conception (see O’Neill 1996, 50),56 which is chosen as the “adequate” one out of four alternatives she identifies, including two conceptions of end-oriented reasoning (“Platonist” and “Instrumental”) and two conceptions of act-oriented reasoning (“Kantian” or “Critical” and “Particularist”) (see O’Neill 1996, 50). According to O’Neill, the “Kantian” conception is the “adequate” conception because it is the only one among the four that fulfills her criterion of followability:

For anything to count as practical reasoning it should, presumably, meet at least certain quite simple standards. It should, in particular, at least aim to be followable by others for whom it is to count as reasoning.

(O’Neill 1996, 51)

In *Towards Justice and Virtue*, O’Neill gives a concise explanation why the other three conceptions fail to meet the requirement of followability. All these explanations, however, essentially rely on the same argument: any practical reasoning that is conditioned (either by a conception of the objectively good, a

56 It is also sometimes called “the critical account of practical reason” (see, for example, O’Neill 2000, 26).
subjective end, or actual norms and commitments in the three types of practical reasoning respectively) is arbitrary and is not followable, for the objectivity and the authority of the condition cannot be established (see O’Neill 1996, 51). O’Neill’s criterion for followability, as Budde (2009) points out, is thus not defined merely in the weaker sense that the principles could be followed, which “one might assume that all forms of practical reason(ing) would claim to fulfil” (208). Instead, the criterion is defined to conform to “a certain intuitive understanding of what we might hope that reason can provide either for practice or for theory” (O’Neill 2000, 12). In O’Neill’s own words, the intuitive understanding is that “reason will be no more than the term we use for the necessary conditions for any coordination, however minimal, by those among whom the reasoning is to count” (O’Neill 1996, 60). The criterion thus “turns out to be the demand to make no arbitrary moves and to provide an accessible authority, in short the standard of followability” (Budde 2009, 211; see also O’Neill 2000, 26). O’Neill’s constructivism is therefore explicitly grounded on the end of coordination. Or, as Budde (2009) puts it, the principle of followability, and hence the whole of O’Neill’s constructivism, is merely “constitutive” or “instrumental” to the end of coordination (212).

I therefore agree with Budde’s claim that “O’Neill cannot avoid the same shortfalls that she diagnosed in Rawls’s Kantian constructivism” (2009, 220). After all the thoughtful attempts she made, the result is apparently nothing more than a move from Rawls’s idealized conception of the person to O’Neill’s teleological conception of practical reason. Both conceptions serve as the starting point of their corresponding constructivist theories without any acceptable constructivist justifications, thereby defeating the purpose of the constructivist approach. In the previous case of the liberal-communitarian debate, I asserted that
if the dualism between the right and the good could ever be resolved, one must not precede an account of practical reason, or an understanding of morality, with any specific conception of the self. In view of the more general problem of the defensibility of the starting points of constructivism, my diagnosis can be generalized as follows: if the constructivist approach could be defensible at all, the only viable starting point is the authentic account of the self-lawgiving and self-determining capability of practical reason in Kant’s philosophy. Any attempt to apply any condition on practical reason, whether it be an end (as in O’Neill’s theory\textsuperscript{57}) or an idealized conception of the person (as in Rawls’s theory), the defensibility of the condition itself will be controversial. In other words, we are only left with two choices. Either no constructivist theory of justice is possible, or we must start with the self-lawgiving and self-determining capability of practical reason in Kant’s philosophy.

### 9. Closing Remarks

The first argument this chapter seeks to make is that, despite his method of avoidance, Rawls’s understanding of morality does serve a foundational role in his political theory of justice and thereby also amounts to a philosophical stance in its own right. In particular, his moral conception of the person and his procedural interpretation of Kant’s conception of autonomy undeniably together constitute a

\textsuperscript{57} Watkins and Fitzpatrick also argue that “O’Neill’s account of reason… would have us inappropriately define reason in terms of one of its consequences rather than in terms of its nature” (2002, 356). Watkins and Fitzpatrick also rightly note that “Kant’s own broader account of rationality seems to provide precisely what must be foreign to O’Neill’s view, an account of the nature of reason” (356). They go on to introduce their interpretation of Kant’s broader account of reason, which they say is intended to be understood as a version of realism rather than constructivism (357). My position is that Kant’s philosophical insight cannot be categorized either as realism or constructivism.
substantial account of morality, which is then applied to develop, or construct, his political conception of justice.

This chapter goes on to argue that Rawls’s understanding of morality is reflected in his procedural interpretation of Kant’s conception of autonomy and the categorical imperative within the framework of an empirical theory (see TJ 226). In Rawls’s own words, his interpretation is an attempt to render Kant’s moral conception more discernible by recasting the “dualisms” he identifies in Kant within the scope of an empirical theory (see TJ 227). As argued in this chapter, Rawls’s misinterpretation deprives practical reason of its primary and essential capability of self-lawgiving and self-determining, and covertly replace it with Rawls’s principle-choosing and concept of expression. The authority of practical reason in Kant is thereby covertly subverted. When practical reason is not capable of self-lawgiving and self-determination as in Kant, and is merely capable of principle-choosing and expressing as in Rawls, practical reason is no longer autonomous and must be subject to another authority. This covert subversion of practical reason therefore unavoidably leads to subsequent debates on the proper authority to which practical reason should be subject.

In the liberal-communitarian debate, Rawls’s communitarian critics “question the liberal claim for the priority of the right over the good, and the picture of the freely choosing individual it embodies” (Sandel 1984, 16). For the communitarians, justice is not rooted in Rawls’s antecedently individuated self but in “a community whose primary bond is a shared understanding both of the good for man and the good of that community” (MacIntyre 1981, 232-233). The debate is thus evidently one on the legitimate authority to which the practical should be subject. For Rawls, it is his political conception of the self as free and equal. For the communitarians, it is the community and its shared understanding.
In the debate on the defensibility of the starting point of Rawls’s constructivism, O’Neill is also dissatisfied with Rawls’s idealized starting point. Unlike the communitarians, O’Neill argues that Rawls’s conception of the self is an idealization masquerading as abstraction (O’Neill 1989, 210) and that the proper starting points for constructivism should be weak, uncontroversial, merely abstract and empirically realistic (O’Neill 2002, 7, 48, 179, 210). As illustrated in this chapter, O’Neill fails to avoid the same shortfalls she diagnosed in Rawls, for her starting points clearly presuppose an end for practical reason. In this case, therefore, the competition for the authoritative role for practical reason is between Rawls’s conception of the self and O’Neill’s end of coordination.

As I argued near the end of the last section, whenever we attempt to apply any condition on practical reason in a political theory, the defensibility of the condition itself will be controversial. The foremost question for the constructivists, or all political philosophers in general, is therefore this: is practical reason, or the will, able to determine itself unconditionally in accordance with a law given by itself? If one answers in the negative and attempts to propose one’s own conditions for practical reason, one must answer another question: why should we accept the conditions one prescribes for practical reason? No answer to this second question will be uncontroversial, for it is impossible to establish once and for all as to why all free rational beings must submit to the same heteronomous condition unanimously. On the other hand, if one answers the first question in the positive and acknowledges the authority of practical reason, the proper question to ask in political philosophy should be this: what limiting conditions of politics can be inferred from the self-lawgiving and self-determining practical reason? The answer to this question lies in the first chapter when I derived the principle of right from the moral law. The question thus takes us back to Kant’s political
philosophy.
III When Political Philosophy Becomes Too Applied: A Kantian Appraisal of the Roles of Political Philosophy Proposed by Rawls

1. Introduction

This chapter will consider a fundamental but rarely asked question: do Kant and Rawls have different views on the role of political philosophy? In most comparative studies of Kant’s philosophy and Rawls’s theory of justice, the focus is usually the alleged Kantian root of Rawls’s Kantian constructivism and the justifiability of his partial assimilation of Kant. Most studies thus mainly focus on the starting point of Rawls’s Kantian constructivism and Kant’s moral doctrine, making only passing reference to Kant’s various political writings.\(^{58}\) In lieu of this popular perspective, this chapter will uncover the fundamental differences between Rawls’s political philosophy and Kant’s political, rather than moral, philosophy by assessing the philosophical implications of the roles of political philosophy set out by Rawls in the first section of his *Justice as Fairness: A Restatement* and in the introduction of his *Lectures on the History of Political Philosophy*. In these two works, Rawls introduces four roles political philosophy “may have as part of a society’s public political culture” (JFR 1) and they reflect certain important convictions of Rawls about political philosophy. In this chapter, the four roles will be scrutinized and contrasted with Kant’s views on the same subject matters, with a view to clarify the deepest differences in the ways of

\(^{58}\) As Flikschuh (2000) notes, “the current absorption of Kant into mainstream liberalism” after Rawls’s use of Kant “is based almost exclusively on Kant’s ethical writings… Kant’s political writings… continue to be neglected by contemporary liberals” (3).
thinking of the two philosophers. The expected finding will not only serve as an instructive foundation for further comparative studies of Kant and Rawls, but will also guide the way towards an answer to a more profound question: how is Rawls’s way of thinking different from Kant’s and which one of them is preferable in face of the challenges acknowledged by Rawls?

In each of the following sections, one of the four roles of political philosophy proposed by Rawls will be contrasted with Kant’s views respectively. Needless to say, I cannot possibly deal with all the passages in which Rawls mentions the roles of political philosophy, but I believe these four roles can faithfully represent his views, given the fact that they are all included in the introductions of two important works without any amendments.

2. Rawls’s Practical Role of Political Philosophy

In LHPP, Rawls proposes the first role of political philosophy as follows:

The first is its practical role arising from divisive political conflict when its task is to focus on deeply disputed questions and to see whether, despite appearances, some underlying basis of philosophical and moral agreement can be uncovered, or differences can at least be narrowed so that social cooperation on a footing of mutual respect among citizens can still be maintained. (LHPP 10)

The practical role thus aspires to attend to actual political conflicts. For some, including me, the role might not sound distinctive enough from the role of politician. The distinction is only further blurred by Rawls’s well-known complete exclusion of metaphysics, normative ethics and metaethics from political
theorizing (Hampton 1989, 795). As Hampton notes, it amounts to an attempt to detach political philosophy from philosophy and such attempt is alarming for many political philosophers (1989, 795). In “The Idea of Overlapping Consensus”, Rawls gives a clear response to this query:

Some may think that to secure stable social unity in a constitutional regime by looking for an overlapping consensus detaches political philosophy from philosophy and makes it into politics. Yes and no: the politician, we say, looks to the next election, the statesman to the next generation, and philosophy to the indefinite future. (IOC 24)

Apparently, Rawls tries to make room for the practical role of political philosophy by limiting the role of the politician to pleasing the voters. It is, however, doubtful whether politicians would openly admit that they merely look to the next election in their undertakings, and it is even more improbable for politicians to publicly declare that this is the proper role not only of themselves but also of the whole profession.

The maxim that a politician looks merely to the next election is, according to Kant, incompatible with publicity and violates the transcendental formula of public right:

All actions relating to the rights of others are wrong if their maxim is incompatible with publicity. (PP 8:381)

The maxim that a politician looks merely to the next election is one that he or she “cannot divulge without thereby defeating [his or her] own purpose, one that absolutely must be kept secret if it is to succeed and that [he or she] cannot
publicly acknowledge without unavoidably arousing everyone’s opposition to [his or her] project” (PP 8:381). Needless to say, this maxim cannot serve as the universal maxim for the profession of politician, let alone the definition of its role. No matter how many examples we see in our experience where politicians make it their maxim in secret, we do not say, contrary to what Rawls suggests, this is what the profession of politician is all about.

On one hand, Rawls relies on the miserable examples in the profession of politician. On the other hand, he highly boasts the profession of political philosopher by claiming that philosophy “looks to the indefinite future” (IOC 24), although the number of examples to the contrary is arguably no less than in the former case. In fact, if the practical role of political philosophy arises, as Rawls maintains, from “divisive political conflict and the need to settle the problem of order” (see JFR 1), we may for the same reason allege that political philosopher looks no further than the current conflict.59 The only clue that might lead us to make the same conclusion as Rawls’s, that philosophy looks to the indefinite future, then lies in the peculiar nature of the conflicts Rawls sets out to settle.

Let us take Rawls’s own political philosophy as an example. The conflict his political liberalism seeks to settle is a “deep doctrinal conflict with no prospect of resolution” (PL xviii), which arises from “the diversity of reasonable comprehensive religious, philosophical, and moral doctrines found in modern democratic societies” that “is not a mere historical condition that may soon pass away” but “a permanent feature of the public culture of democracy” (PL 36). In Rawls’s opinion, philosophy looks to the indefinite future because the doctrinal

59 This is apparently the view of Rawls when he evaluates the roles of political philosophers, including Hobbes, Locke, Montesquieu and himself, in their respective historical situations (see JFR 1-2).
conflict it looks to can never be resolved in the future and is a permanent feature of the present culture. By devaluing the role of politician on one hand and asserting in a dogmatic tone the perpetuity of the deepest unresolvable conflict he sees in the democratic culture on the other, Rawls manages to make room for a practical role of political philosophy. The substantial role of political philosophy, in Rawls’s thoughtful design, is thus to mediate society’s deepest conflict (see IOC 24) that is proclaimed to be unresolvable from the beginning, and the political philosophers should refrain at the same time from addressing the moral topics on which the doctrines divide, under the politically correct banner of impartiality (see PL xxviii).

With regard to Rawls’s practical role, several responses can be consolidated as follows on behalf of Kant. First, in Kant’s view, political philosophy cannot be seen as arising from actual conflict as Rawls suggests (LHPP 10). In Kant, political philosophy arises from a comprehensive philosophical account of human nature, including the sensible and the supersensible nature (CPrR 5:43), although it often has to take as its object a certain particular nature of human beings that is cognized only by experience (see MM 6:217). In his comprehensive interpretation of human nature, Kant identifies “the maxim of violence in human beings and of their malevolent tendency to attack one another” (MM 6:312) that is “woven into human nature” (Rel 6:30). Kant’s account of of human nature thus necessitates the need for an account of the limiting condition of politics (PP 8:372) even before the actual experience of conflicts, hostilities or threats.

Admittedly, conflict does have a role in Kant’s political philosophy, yet it is never the doctrinal conflict between different comprehensive doctrines Rawls consistently looks to in his endeavor. In relation to his political philosophy, Kant speaks of conflict in a few occasions. First, as briefly mentioned above, Kant’s
account of human nature acknowledges the malevolent tendency inherent in human nature against each other, which gives rise to “the abuse of reciprocally conflicting freedom” (CPJ 5:432). Kant’s political philosophy can be seen as arising from this malevolent tendency to conflict, for the tendency must be “opposed by lawful power in a whole, which is called civil society” (CPJ 5:432).

Second, Kant speaks of “the unavoidable conflict of culture with the nature of the human species as a physical species” (CBHH 8:116) in the progress of the human race when he tries to articulate a correct interpretation of Rosseau’s works. This conflict is again obviously not the doctrinal conflict identified by Rawls. Third, Kant also maintains in his “The Conflict of the Faculties” that:

…there will be a conflict between the higher [(including the faculties of medicine, law and theology)] and lower [(the philosophy faculty)] faculties which is, first, inevitable, but second, legal as well; for the lower faculty has not only the title but also the duty, if not to state the whole truth in public, at least to see to it that everything put forward in public as a principle is true. (CF 7:32)

In Kant, the conflict brought up mainly by the lower faculty of philosophy is inevitable and legal, for it is in line with the duty of the lower faculty to “answer for the truth of the teachings it is to adopt or even allow” (CF 7:27). However, since Rawls asserts that political philosophy has no special access to fundamental truths (LHPP 1), the duty of philosophers proposed by Kant cannot be one of the roles of political philosophy for Rawls.

The first objection thus leads to the second objection. By attending to actual conflicts without any duty “to answer for the truth of the teachings it is to adopt or
even allow”, the practical role of political philosophy is limited to the
determination of applied principles of politics, which, according to Kant, is not
the primary role of political philosophy. In Kant’s view, an applied principle of
politics is a principle that “applies [the] concepts [in the metaphysics of right] to
cases of experience” (SRLP 8:429) and it is “drawn from experiential cognition of
human beings, that have in view only the mechanism for administering right and
how this can be managed appropriately” (SRLP 8:429). Given the diversity,
complexity and contingency of experiential cognition of human beings, however,
the determination of an applied principle of politics cannot be regarded as the
primary role of philosophy but must be left to the better judgments of the
politicians and the citizens in a society. As rightly reminded by Ripstein (2009), in
Kant’s political philosophy “[t]he systematic implications of that right [of every
person being his or her own master] have to be worked out first, before any
‘principle of politics’ incorporating information based on experience can be
introduced” (5). Principles of politics, according to Kant, must presuppose “a
metaphysics of right”, which “abstracts from all conditions of experience” (SRLP
8:429). “Right must never be accommodated to politics”, Kant maintains, “but
politics must always be accommodated to right” (SRLP 8:429). Rawls, however,
arguably accommodates his theory entirely to politics without any metaphysics of
right. Kant might therefore contend that Rawls’s practical role of political
philosophy is practical in the wrong way by taking up the role of politician and
citizen, leaving aside the proper role of political philosophy in Kant’s view at the
same time.

In response to the two points discussed above, Rawls might declare that
political philosophy can be positioned to determine certain principles of politics
that are applicable only in a particular culture in a specific time and place:
The aim of political philosophy, when it presents itself in the public culture of a democratic society, is to articulate and to make explicit those shared notions and principles thought to be already latent in common sense; or, as is often the case, if common sense is hesitant and uncertain, and doesn’t know what to think, to propose to it certain conceptions and principles congenial to its most essential convictions and historical traditions. (Rawls 1980, 518)

The third point that could be made on behalf of Kant is that, if a metaphysics of right is lacking, political philosophy cannot assume even such a humble role proposed by Rawls. In the absence of a metaphysics of right, Rawls invents an unusual theoretical device called reflective equilibrium as part of his justificatory strategy, and he claims that the justification of his political conception of justice “rests upon the entire conception and how it fits in with and organizes our considered judgments in reflective equilibrium” (TJ 507). Considered judgments, according to Rawls, “enter as those judgments in which our moral capacities are most likely to be displayed without distortion” (TJ 42). In order to have our moral capacities displayed without distortion, we should, Rawls advises, discard those judgments “made with hesitation”, “in which we have little confidence”, “made when we are upset or frightened”, “or when we stand to gain one way or the other”, since judgments like these “are likely to be erroneous or to be influenced by an excessive attention to our own interests” (TJ 42). The notion of considered judgments thus presupposes an idea of correctness which is in turn based on the notion of moral capacities in Rawls’s theory, including the capacity for a sense of justice and the capacity for a conception of the good (TJ xii).

If all these elaborations sound sophisticated yet confusing, we might safely
conclude, in Rawls’s own words, that he essentially seeks to determine certain principles of justice which happen to lead to judgments that “match our considered judgments duly pruned and adjusted” (TJ 18) to display our sense of justice without distortion. In Kant’s view, however, this might be the most ignorant strategy a political philosopher could attempt in the absence of a proper metaphysics of right, when Rawls tries to justify a certain conception of justice merely by its being “congenial to [common sense’s] most essential convictions and historical traditions” (Rawls 1980, 518). In fact, for Kant, the very notion of a sense of justice is already indicative of a misled philosophical effort to investigate the nature of justice, for the concept of justice “cannot have [its] seat in a sense, and… even less could such a sense have the slightest capacity for the expression of universal rules” and “a representation of truth, suitability, beauty, or justice could never enter our thoughts if we could not elevate ourselves above the senses to higher cognitive faculties” (CPJ 5:293). Without going into a lengthier discussion than necessary to show their differences on this point, I will invoke another observation to support my claim that, in Kant’s view, Rawls cannot fulfill his artfully formulated practical role in the limited scope of the public culture of a democratic society.

When he explains the first practical role of political philosophy in JFR, Rawls specifically mentions an important example of conflict he identifies in the tradition of democratic thought between the claims of liberty and the claims of equality (JFR 2). He asserts that: “[d]ebates over the last two centuries or so make plain that there is no public agreement on how basic institutions are to be arranged so as to be most appropriate to the freedom and equality of democratic citizenship” (JFR 2). If Kant ever read this, the most probable comment he might make is that: the debates merely make plain that there is no agreement among the
political philosophers who misrepresent these ideas in different ways.\textsuperscript{60} In Kant’s philosophy, freedom, equality and independence are three a priori principles under the idea of the general will (see TP 8:290) which do not conflict with each other. They are different expressions of the same limiting condition of politics, i.e. the concept of right, and imply each other (see TP 8:289-296). It is only when the principle of freedom becomes claims for freedom, when the principle of equality becomes claims for equality and the principle of independence becomes claims for independence, as in Rawls, do we start to struggle in the unresolvable conflict between them. In Kant’s view, therefore, it might be justified to conclude that, at least in this case, Rawls’s practical role of political philosophy is its role to settle the conflict it causes, without any hope of success if a metaphysics of right is nonetheless missing, for, in Kant’s words, “[I]ike the wooden head in Phaedrus’s fable, a merely empirical doctrine of right is a head that may be beautiful but unfortunately it has no brain” (MM 6:230).

3. Rawls’s Orientation Role of Political Philosophy

In LHPP, Rawls summarizes the second role of political philosophy as follows:

The second role, which I call orientation, is one of reason and reflection. Political philosophy may contribute to how a people think of their political and social institutions as a whole, of themselves as citizens, and of their basic aims and purposes as a society with a history - a

\textsuperscript{60} Kant is definitely not alone if he really raised this complaint. For instance, Dworkin also complains that: “Nothing is easier than composing definitions of liberty, equality, democracy, community, and justice that conflict with one another. But not much, in philosophy, is harder than showing why these are the definitions that we should accept” (Dworkin 2006, 116).
nation - as opposed to their aims and purposes as individuals, or as members of families and associations. (LHPP 10)

In JFR, Rawls further elaborates that:

The idea is that it belongs to reason and reflection (both theoretical and practical) to orient us in the (conceptual) space, say, of all possible ends, individual and associational, political and social. Political philosophy, as a work of reason, does this by specifying principles to identify reasonable and rational ends of those various kinds, and by showing how those ends can cohere within a well-articulated conception of a just and reasonable society. (JFR 3)

The first feature worth noting about the role is that it arises from the need, as Rawls puts it, for a conception that enables the members of any civilized society “to understand themselves as members having a certain political status - in a democracy, that of equal citizenship - and how this status affects their relation to their social world” (JFR 3). Since Rawls is specifically referring to a democracy in his specification of the orientation role, we may safely assume that his own theory of justice is at least an exemplar of this role. However, if Rawls’s doctrine is an empirical doctrine of right that accounts merely for our sense of justice and considered judgment without reflecting or elevating to higher cognitive faculties, how could political philosophy, at least in the specific case of Rawl’s own doctrine, possibly orient us as he envisages?

Owing to the absence of a metaphysics of right that indicates an elevation to an account of the higher cognitive faculty of pure practical reason, the best Rawls can do with his political philosophy is to “propose” a “conception” for others to
“understand” their political status. At its very best, such a conception can only illustrate the objective theoretical necessity of the conception, but not its subjective practical necessity, for Rawls never attempts to show its connection with an apodictic law by which practical reason unconditionally and necessarily determines itself. In other words, people may be convinced by the conception and the theory, but the subjective practical necessity of the subjects actually adopting the conception in their actions nonetheless remains unexplained. This motivational deficit, as Krause (2005, 363) calls it, is a perennial problem in Rawls and in other cognitivist theories of justice that subsequently leads to other debates and criticisms.61 Also, since Rawls acknowledges no formal law of the pure practical reason serving as the proper starting point, the starting point in Rawls’s idea of political philosophy must be certain unjustified materials. This results in the notorious problem of the unconstructed starting point in Rawls’s Kantian constructivism and also in the subsequent developments of political constructivism by Rawls and other scholars (see Bagnoli, 2014; Cohen, 2003; Raz, 1990, 15; 2003; Timmons, 2003).

These problems are arguably the unavoidable consequences of Rawls’s failure to appeal to an accurate account of practical reason. This accusation clearly stands

61 For instance, Krause (2005) rightly acknowledges that “[t]o understand what is involved in desiring justice is to know that this desire never could be satisfied by perfecting the theory of justice” (381). However, in contrast to my diagnosis, she believes that the study of justice “leans on our knowledge of moral psychology and human agency, central to which must be an awareness of the irrepressible place of affectivity and conceptions of the good” (381). In response to her diagnosis, I would add that our knowledge of moral psychology and human agency also requires a preceding account of the higher cognitive faculty of pure practical reason. Krause also points out that the problem of motivational deficit is noted by many commentators (see Sandel 1982 1996; Williams 1985; Taylor 1989, 1995; Baynes 1995, 208; Solomon 1995; Larmore 1996; Blackburn 1998; Markell 2000; Bennett 2001; Hall 2003; Walzer 2003) and is linked to the debate of the priority of the right over the good.
in stark contrast to Rawls’s understanding of theoretical and practical reason, for he believes that his theory is a work of both theoretical and practical reason (JFR 3):

Following Kant’s way of making the distinction, we say: practical reason is concerned with the production of objects according to a conception of those objects - for example, the conception of a just constitutional regime taken as the aim of political endeavor - while theoretical reason is concerned with the knowledge of given objects. (PL 93)

Rawls’s explanation of Kant’s distinction, however, is partial and inaccurate, for, apart from a few imprecise wordings which are not going to be discussed here, it does not recognize the important distinction in Kant between a will (practical reason) subject to a nature and a nature subject to a will:

Hence the difference between the laws of a nature to which the will is subject and of a nature which is subject to a will (as far as the relation of the will to its free actions is concerned) rests on this: that in the former the objects must be the causes of the representations that determine the will, whereas in the latter the will is to be the cause of the objects, so that its causality has its determining ground solely in the pure faculty of reason, which can therefore also be called a pure practical reason. (CPrR 5:44)

In Kant’s view, only when the determining ground resides solely in the pure faculty of reason, can we regard the will as the cause of the objects. Otherwise,
the will is determined by the representations of objects and cannot be seen as the
cause of the objects. In other words, only when the determining ground resides
solely in the pure faculty of reason is reason truly practical (see CPrR 5:24-25). A
detailed discussion of Rawls’s understanding of morality merits an individual
paper, so I am not going into further details of the distinction here. For the
purpose of this chapter, it suffices to illustrate that Rawls’s practical reason is not
Kant’s practical reason, for Rawls ignores, or avoids, the primary function of
practical reason as its own determining ground in Kant. In Rawls, Kant’s account
of pure practical reason is given a more “restricted” sense as follows:

The distinction between the reasonable and the rational goes back, I
believe, to Kant: it is expressed in his distinction between the
categorical and the hypothetical imperative in the Foundations and his
other writings. The first represents pure practical reason, the second
represents empirical practical reason. For the purposes of a political
conception of justice, I give the reasonable a more restricted sense and
associate it, first, with the willingness to propose and honor fair terms of
cooperation, and second, with the willingness to recognize the burdens
of judgment and to accept their consequences. (PL 48-49)

In brief, Rawls’s “restricted sense” of the reasonable is restricted in the sense
that practical reason is prescribed with two notions of willingness as its
determination ground. Or in other words, pure practical reason is now being
“restricted” by the objects of, first, proposing and honoring fair terms of
cooperation, and second, recognizing the burdens of judgment and accepting their
consequences. Rawls’s idea of the reasonable in the restricted sense therefore
cannot represent pure practical reason, for practical reason is now being conditioned by two objects. In essence, Rawls’s formulation of the orientation role renders the problem of politics “a mere technical problem” and is reminiscent of Kant’s idea of political moralist (PP 8:377). For Rawls, the technical problem is how a people ought to think in a democracy in order to attain the end of a stable and just society of a democracy. In Kant’s view, Rawls’s practical reason is thus nothing but the lower faculty of desire, where practical reason is empirically conditioned and determined by the representations of objects.

Kant is unmistakably against any approach to political philosophy in which the end is prescribed as the determining ground of the first principles (see PP 8:377). In Kant’s view, this approach is like “putting the cart before the horse” (PP 8:376), for Rawls determines “the cart”, i.e. the end, before the orientation of “the horse”, i.e. practical reason. Rawls’s approach amounts to an attempt to orient practical reason, which can be of itself practical (see Gr 4:461, CPrR 5:105, MM 6:214), by prescribing a heterogeneous end for it to follow, and it is impossible to explain with absolute certainty why practical reason (“the horse”) must follow the orientation of the specific end (“the cart”), when the role of determining the orientation should belong to practical reason (“the horse”) according to Kant. Rawls’s approach betrays the positive concept of freedom, i.e. the ability of pure reason to be of itself practical (see MM 6:213-214), and deprives pure practical reason of its authority to make “a rule of [itself] the motive of an action (by which an object can become real)” (CPrR 5:60).

In contrast to this approach, Kant clearly alleges that in problems of practical reason, and hence also in problems of politics, one must begin from its formal principle, i.e. “[s]o act that you can will that your maxim should become a universal law (whatever the end may be)”, rather than its material principle, i.e.
“the end (as object of choice)” (see PP 8:376-377):

The [formal] principle must undoubtedly take precedence; for, as a principle of right, it has unconditional necessity, whereas the [material principle] necessitates only if the empirical conditions of the proposed end, namely of its being realized, are presupposed; and even if this end (e.g., perpetual peace) were also a duty, it would still have to be derived from the formal principle of maxims for acting externally. (PP 8:377)

Kant’s allegation is not merely targeted against ends like Rawls’s idea of a stable and just society, but even the ideal of perpetual peace, which is at the same time a duty, for its being a desirable end is nonetheless derived from the formal principle. Political moralists or philosophers like Rawls, who subordinate the principles to the end, might share the same ultimate goal as Kant. However, Kant criticizes their approach not because of the particular end they choose, but the procedure for leading to the end (see PP 8:377). When the end is wrongly or prematurely specified before the principles, the problem of politics becomes a mere technical problem. The end is wished for only as a natural good, and the necessity of the motivation for free citizens to pursue this end is inexplicable. On the contrary, if we begin with the formal principle of practical reason instead, like what the moral politician does according to Kant, the problem of politics becomes a moral problem, and the end of perpetual peace “is now wished for not only as a natural good but also as a condition arising from acknowledgement of duty” (PP 62).

Kant’s position regarding the precedence of the formal principle in problems of politics is remarkably consistent with his similar allegation in his moral philosophy, in which the highest good is the end but not the determining ground of the pure will (see, for example, CPrR 5:109 and TP 8:279).
Motivation for free citizens to realize the end is in this case explicable, for the end is now connected with the idea of duty, in which respect for the moral law is the motivation for its realization (see CPrR 5:76; CPrR 5:88; Rel 6:27; Rel 6:46). A Kantian objection against Rawls’s orientation role is thus clear and firm: Rawls’s empirical theory is an empirical doctrine of right that cannot possibly account for the self-orientation of practical reason.

4. Rawls’s Reconciliation Role of Political Philosophy

In LHPP, Rawls explains the reconciliation role of political philosophy as follows:

A third role, stressed by Hegel in his Philosophy of Right (1821), is that of reconciliation: political philosophy may try to calm our frustration and rage against our society and its history by showing us the way in which its institutions, when properly understood, from a philosophical point of view, are rational, and developed over time as they did to attain their present, rational form. When political philosophy acts in this role, it must guard against the danger of being simply a defense of an unjust and unworthy status quo. This would make it an ideology (a false scheme of thought), in Marx’s sense. (LHPP 10)

In the third role Rawls seemingly tries to take up the role of politician again by aspiring to calm our frustration and rage against our society and its history. The first thing to note about this aspiration is that it inevitably limits the validity of a doctrine of political philosophy to the specific historical situation faced by the philosopher, for it is unimaginable how a specific doctrine of political philosophy
can possibly calm our frustration and rage in any possible political situation. This observation is discernibly in line with the alleged Hegelian origin of the role of reconciliation, for Hegel explicitly states the following just before he introduces the idea of reconciliation in philosophy:

This treatise, in so far as it contains a political science, is nothing more than an attempt to conceive of and present the state as in itself rational. As a philosophic writing it must be on its guard against constructing a state as it ought to be. Philosophy cannot teach the state what it should be, but only how it, the ethical universe, is to be known… To apprehend what is is the task of philosophy, because what is is reason. As for the individual, every one is a son of his time; so philosophy also is its time apprehended in thoughts. It is just as foolish to fancy that any philosophy can transcend its present world, as that an individual could leap out of his time or jump over Rhodes. If a theory transgresses its time, and builds up a world as it ought to be, it has an existence merely in the unstable element of opinion, which gives room to every wandering fancy. (Hegel [1821] 2001, 19)

I choose to quote the long passage here because it is necessary to illustrate the essential connection between two ideas in Hegel. First, the task of philosophy is merely to apprehend what is, rather than what should be. Second, no philosophy can transcend its present world. For Hegel, the two ideas imply each other, and subsequently lead to his idea of reconciliation which is stated after the last passage:

The barrier which stands between reason, as self-conscious spirit, and
reason as present reality, and does not permit spirit to find satisfaction in reality, is some abstraction, which is not free to be conceived. To recognize reason as the rose in the cross of the present, and to find delight in it, is a rational insight which implies reconciliation with reality. This reconciliation philosophy grants to those who have felt the inward demand to conceive clearly, to preserve subjective freedom while present in substantive reality, and yet though possessing this freedom to stand not upon the particular and contingent, but upon what is self-originated and self-completed. (Hegel [1821] 2001, 19)

The satisfaction of the claims of reason (see Hegel [1821] 2001, 28) or the demands of reason (see Hegel [1821] 2001, 27) is an important concept in Hegel’s philosophy of right that is deeply woven into his complex metaphysical doctrine. The delight stated in the passage arises from the overcoming of the barrier, i.e. the abstraction, and the adoption of the rational insight that enables us to see “the rose in the cross of the present” rather than to look to the abstraction. According to Hegel, abstraction is the undesirable way, in contrast to the way of rational insight, to apprehend the real world that corresponds to “the negative side of the will”, which misleads us to reckon “any content as a limit, and [flee] from it” (Hegel [1821] 2001, 31). “This freedom is that of the void” (Hegel [1821] 2001, 31), which becomes actual when we submit ourselves to “the desire to teach the world what it ought to be” (Hegel [1821] 2001, 20) and “would destroy the established social order, remove all individuals suspected of desiring any kind of order, and demolish any organization” (Hegel [1821] 2001, 31). In the end, if we insist on the way of abstraction, “[o]nly in devastation does the negative will feel that it has reality” (Hegel [1821] 2001, 31).
On the surface at least, Rawls’s reconciliation role of political philosophy is arguably borrowed from Hegel’s idea of the reconciliation of reason and reality. In Hegel’s original idea, the reconciliation of reason and reality leads to the satisfaction of the claims of reason, which enables reason to find delight in reality. The idea clearly presupposes Hegel’s metaphysical doctrine which assigns philosophy the role of the rational apprehension of what is rather than what should be, and this is the fundamental reason why the retreat from abstraction is considered a reconciliation in the very first place. In his formulation, Rawls tactically tries to avoid most of the metaphysical burden behind Hegel’s idea of the satisfaction of the claims of reason, by focusing on the negative idea of dissatisfaction rather than the positive idea of satisfaction, and by turning to the negative emotions of frustration and rage from the positive emotion of delight.

Nonetheless, significant philosophical implications still remains in the reconciliation role of political philosophy adapted by Rawls that is clearly detectable in his elucidation. Rawls basically alludes to Hegel’s idea of a correct rational insight or apprehension of reality when he stresses that frustration and rage can be calmed if we adopt a proper understanding of the present reality of our society and its history from a philosophical point of view. The difference between Rawls and Hegel is merely rhetorical, for his thoughtful paraphrasing (“...when properly understood, from a philosophical point of view, are rational…”) essentially amounts to an attempt to propose a philosophically authentic way to apprehend reality, like what Hegel did. Although Rawls never admits that his theory presupposes any metaphysical doctrine, his claim in the reconciliation role seems to presuppose Hegel’s view on the role of philosophy, i.e. to apprehend what is rather than what should be. It is especially conspicuous when Rawls explicitly acknowledges that his view “fits one of Hegel’s well-known sayings:
‘When we look at the world rationally, the world looks rationally back.’” (JFR 3) and echoes Hegel’s philosophical insight by adding that “we are to accept and affirm our social world positively, not merely to be resigned to it” (JFR 3).

If we leave aside the Hegelian origin of the reconciliation role, it could be difficult to understand why Rawls places such an emphasis on the negative emotions of frustration and rage. Rawls remarkably proclaims that he and his doctrine embrace the fact of reasonable pluralism, and that pluralism is not to be “seen as disaster but rather as the natural outcome of the activities of human reason under enduring free institutions” (PL xxiv). If pluralism of comprehensive doctrines must be accepted as an unchangeable fact in a democracy, though Rawls acknowledges at the same time that it is “not always easy to accept” (JFR 3), why are the negative emotions of frustration and rage, which could be far easier to accept than the former, not accepted similarly? If Rawls believes that a society united under a single comprehensive doctrine is undesirable and requires the oppressive use of state power, why does he believe that political philosophy should seek to unite a people by prescribing a particular way of understanding the reality?

In Rawls’s idea of the reconciliation role, political philosophy seeks to advise how we should find delight by prescribing a certain way of apprehending or judging our society and its history. In Kant’s view, however, the role is plainly unacceptable and unjustifiable. For Kant, philosophy is never in a position to advise how people should be happy:

With respect to the former (happiness) no universally valid principle for laws can be given. For both the circumstances of the times and the highly conflicting but always changing illusion in which someone
places his happiness (though no one can prescribe to him in what he should place it) make any fixed principle impossible and [happiness] in itself unfit to be a principle of legislation. (TP 8:298)

In fact, according to Kant, not only is it impossible for a philosopher to formulate any reliable fixed principle for the pursuit of happiness or the avoidance of negative feelings, it is also impossible even for an individual to “determine with complete certainty what would make [himself or herself] truly happy, because for this omniscience would be required” (Gr 4:418). In his political philosophy, Kant forcefully argues that the concept of right, which is the limiting condition of politics (see PP 8:372) and the first principle in Kant’s political philosophy, “proceeds entirely from the concept of freedom in the external relation of people to one another and has nothing at all to do with the end that all of them naturally have (their aim of happiness) and with the prescribing of means for attaining it; hence too the latter absolutely must not intrude in the laws of the former as their determining ground” (TP 8:289). In political matters, therefore, the limiting condition is always the concept of right rather than any alleged principle of happiness. The pursuit of happiness is left to the judgment of the individual, as long as it conforms to the concept of right:

As for the freedom [of every member of a state] as a human being I express its principle for the constitution of a commonwealth in the following formula: No one can coerce me to be happy in his way (as he thinks of the welfare of other human beings); instead, each may seek his happiness in the way that seems good to him, provided he does not infringe upon that freedom of others to strive for a like end which can
coexist with the freedom of everyone in accordance with a possible universal law (i.e., does not infringe upon this right of another). (TP 8:290)

Kant’s philosophy is thus in stark contrast to Hegel’s and Rawls’s. By adopting Hegel’s idea in his reconciliation role, Rawls might become a political philosopher who looks merely to the past and the present, rather than to “the indefinite future” (see IOC 24) as he intends.

5. The Role of Probing the Limits of Practicable Political Possibility

In LHPP, Rawls introduces the fourth role of political philosophy as follows:

The fourth role is that of probing the limits of practicable political possibility. In this role, we view political philosophy as realistically utopian. Our hope for the future of our society rests on the belief that the social world allows at least a decent political order, so that a reasonably just, though not perfect, democratic regime is possible. (LHPP 10-11)

As this fourth role is only “a variation of the previous one” (JFR 4) and sufficient background has been provided in the preceding sections, I will simply examine Rawls’s explanation textually and see how Kant might want to revise Rawls’s specification of the role.

First, Kant does not agree with the idea of probing the limits of practicable political possibility. Practicable political possibility has to do with “the physical possibility of [the object] by the free use of our powers” (CPrR 5:57). If we
consider the practicable political possibility only, the implication is that the object is taken as the determining ground of our faculty of desire (see CPrR 5:57). The problem then becomes a problem of political prudence, the solution of which requires “much knowledge of nature… to make use of its mechanism for the end proposed” (PP 8:377). Philosophers, like any ordinary person, are not omniscience even about the present, let alone the indefinite future. It is therefore extremely dangerous to let philosophers define for all what is practicable and what is not, even in the simplest cases, let alone an extension to the limits. The practicable political possibility of an object should be left to the better judgment of the united will of the people, for “when someone makes arrangements about another, it is always possible for him to do the other wrong; but he can never do wrong in what he decides upon with regard to himself” (MM 6:313). Philosophers can do the most serious wrong when they try to dictate for all of us in a dogmatic tone what is practicable and what is not, as if they knew better about the physical possibility of an object than all of us, thereby determining directly what we might desire or hope for our societies.

Instead of focusing on the question of the limits of practicable political possibility, Kant focuses on the question of the practical, i.e. what human beings as rational natural beings realize through their will, which is “the ability to determine their causality by the representation of rules, hence insofar as they are capable of actions in accordance with principles and consequently also in accordance with a priori practical principles” (CPrR 5:32). In this case, in which the determining ground is the law of the will, “the moral possibility of the action must come first”, and the only question is “whether we could will an action which is directed to the existence of an object if the object were within our power” (see CPrR 5:58). This consideration is quite independent of the physical possibility of
the object, for what matters for the self-determination of the will is merely that what it determines according to its own law is necessarily within the power available for its determination, or else it would be in a contradiction with itself by determining itself to something beyond the power available. Kant does not claim a better knowledge of the physical possibility of any object, but he illustrates what we can hope for out of “what we know to reside within our power”, which can be determined by the moral law (Rel 6:98).

The difference between Rawls’s focus on physical possibility and Kant’s primacy of moral possibility is clearly noticeable in Rawls’s elaboration of the fourth role. First, Rawls asserts that: “Our hope for the future of our society rests on the belief that the social world allows at least a decent political order” (LHPP 11 and JFR 4). Our hope is thus subject to an appraisal of the physical possibility of what the social world allows, not to the moral possibility of what we can do within our power. Kant would contend in response that:

Now I claim to be able to predict to the human race - even without prophetic insight - according to the aspects and omens of our day, the attainment of this goal. That is, I predict its progress toward the better which, from now on, turns out to be no longer completely retrogressive. For such a phenomenon in human history will not be forgotten, because it has revealed a tendency and faculty in human nature for improvement such that no politician, affecting wisdom, might have conjured out of the course of things hitherto existing, and one which nature and freedom alone, united in the human race in conformity with inner principles of right, could have promised. But so far as time is concerned, it can promise this only indefinitely and as a contingent occurrence. (CF 7:88)
Kant thereby urges that no politician, no philosopher and nobody should conjure merely out of the present reality and its history what we can hope for the future, but must turn to the tendency and faculty in human nature for improvement and what we can do within our power when we consider what we can hope for our future. In Kant’s view, our hope cannot rest upon what we think the social world allows, as Rawls suggests, but must rest upon what we can do within our power to shape the social world. In LHPP, Rawls goes on to assert that with regard to its fourth role, political philosophy is going to answer the following questions:

So we ask: What would a just democratic society be like under reasonably favorable but still possible historical conditions, conditions allowed by the laws and tendencies of the social world? What ideals and principles would such a society try to realize given the circumstances of justice in a democratic culture as we know them? (LHPP 10-11)

The first question is clearly about the physical possibility of an object, i.e. a just democratic society, that is “conjured out of the course of things hitherto existing” (CF 7:88) by appealing to an account of “reasonably favorable but still possible historical conditions” taught by experience. The second question further confirms that in Rawls’s idea of political philosophy, the future of a society is basically determined by the present reality, for he acknowledges once again that the ideals and principles we try to realize depend on the circumstances as we know them. In Kant’s view, the way of thinking implicit in these questions is not correct, for the empirical and contingent conditions of carrying out the law are made conditions of the law itself:

For here it is a matter of the canon of reason (in the practical), where the
worth of practice rests entirely on its conformity with the theory underlying it, and all is lost if the empirical and hence contingent conditions of carrying out the law are made conditions of the law itself, so that a practice calculated with reference to an outcome probable in accordance with previous experience is given authority to control a self-sufficient theory. (TP 8:277)

Rawls’s fourth role of political philosophy is clearly an attempt to wrongly regard the empirical and contingent conditions, i.e. the historical conditions of the social world and its culture as we know them, as the conditions of the ideal of political philosophy. Political philosophy and the dignity of humanity are rendered worthless in Kant’s view, if all we can expect from political philosophy is a reinforcement of the belief that the empirical and contingent conditions should have primacy over any possible practical laws in political matters. Kant might also add that the attempt to define political philosophy as such is the scandal of philosophy:

For, to the scandal of philosophy, it is not uncommonly alleged of this theory that what may be correct in it is yet invalid in practice; and this is said in a lofty, disdainful tone, full of the presumption of wanting to reform reason by experience even in that in which reason puts its highest honor, and in a wisdom that can see farther and more clearly with its dim moles’ eyes fixed on experience than with the eyes belonging to a being that was made to stand erect and look at the heavens. (TP 8:277)

By rejecting any attempt to condition and reform reason by experience, Kant does
not thereby maintain that political philosophy should pay no regard to empirical conditions. In Kant’s view, experience is indispensable in a metaphysics of morals:

But just as there must be principles in a metaphysics of nature for applying those highest universal principles of a nature in general to objects of experience, a metaphysics of morals cannot dispense with principles of application, and we shall often have to take as our object the particular nature of human beings, which is cognized only by experience, in order to show in it what can be inferred from universal moral principles. But this will in no way detract from the purity of these principles or cast doubt on their a priori source. - This is to say, in effect, that a metaphysics of morals cannot be based upon anthropology but can still be applied to it. (MM 6:216-217)

According to Kant, it is therefore legitimate and actually necessary to consider empirical conditions in political philosophy, as long as the empirical conditions are not seen as the conditions of the highest principles but as the conditions of the application of these principles. Louden in his *Kant’s Impure Ethics* (2000) gives a very good account of “the empirical or impure side of Kant’s project in ethics” (6):

Kant referred to this second part variously as “moral anthropology”, “practical anthropology”, “applied moral philosophy”, and sometimes simply “anthropology,” but the important point is that it deals with the empirical (or what I call “impure”) study of human nature rather than with pure (non-empirical) principles. Although Kant was adamant that
the first or pure part of ethics was foundational and thus fundamentally more important than the second part, he was equally insistent that the second part was absolutely necessary whenever one wished to apply the results of the first part to human beings. (Louden, 2000, vii)

Louden rightly acknowledges the role of the second part in the application of the pure principles to human beings, but I hesitate to call the empirical study of human nature in Kant the “impure” part of his project, for it seems to convey an excessive sense of tension between the pure and the empirical. In Kant’s philosophy, the two parts are complementary and combined in a certain order, with the pure part serving as the highest principles and the empirical part serving as the conditions of the application of these principles. Rawls’s philosophy, however, has the order reversed, with the empirical wrongly serving as the condition of the first principles. While Kant’s ordering results in an empirically informed theory, Rawls’s ordering results in an empirically determined theory in which there is no authentic “pure” part whatsoever. Rawls’s erroneous ordering is thus reminiscent of Kant’s insightful reminder: “[s]o much depends, when we wish to join two good things, on the order in which we combine them!” (Rel 6:179)

---

63 Louden regards the pure part of Kant’s ethics as its first part, and its empirical part as its second part (see Louden 2000, vii, 6).
64 The distinction between an empirically informed theory and an empirically determined theory is borrowed Louden (2000, 170).
65 By “two good things”, Kant is referring to “statutory laws” and “moral conviction” in the passage (see Palmquist 2016, 446). However, as Palmquist (2016) notes, the statement also holds for Kant’s argument in the First Piece that “we err only when we adopt a deluded ‘conviction’, choosing to believe that the lesser, conditional good is more important than the higher, absolute good of morality” (447). Kant’s reminder turns out to be relevant also to the argument here in this chapter, for the empirical part of Kant’s moral and political philosophy is not bad by itself either
6. Closing Remarks

Rawls’s theory is usually seen as the revival of the systematic and philosophical discussion of justice (Höffe 1995, 4) and “marks a pivotal turning point in the most recent history of practical philosophy” (Habermas 1995, 109). Yet, this chapter argues that the roles of political philosophy proposed by Rawls might be considered unacceptable and unjustifiable by Kant, though Rawls ironically believes that his own doctrine “sufficiently resembles Kant’s in enough fundamental respects so that it is far closer to his view than to the other traditional moral conceptions that are appropriate for use as benchmarks of comparison” (Rawls 1980, 517). The findings of this chapter can not only serve as an instructive foundation for further comparative studies of Kant and Rawls, but also as a reminder for the Rawlsian political philosophers, or the liberal thinkers in general, that if they should encounter any difficulties in their endeavors, they might consider reforming their way of thinking by examining its differences from Kant’s philosophical insights.

and we err only when we err in the order in which we combine the pure part and the empirical part.
IV A Political Conception of the Human Being in Kant: An Alternative Interpretive Starting Point of Kant’s Political Philosophy from a Political Perspective

1. Introduction

There is a perennial tension in Rawls’s political liberalism as a result of his agenda to develop a political conception of justice that can be presented as freestanding from any comprehensive moral doctrines. The tension seems to be unavoidable, for if we consider the normative connotations of both justice and morality, the very idea of a conception of justice seems to suggest that there must be a corresponding moral conception of the person in the same theory. It is therefore unsurprising to see that, notwithstanding his method of avoidance, Rawls explicitly invokes a moral conception of the person as a starting point of his political liberalism both before and after his political turn.

On the other hand, Kant interpreters face a different challenge when they want to justify Kant’s political philosophy in modern pluralistic societies. Given

---

66 Here I mean a starting point in a weaker sense that the conception of the person is not a derivative but a starting element of his theory, without regard to its substantial role in the theory.

67 In his Political Liberalism, which is often regarded as the fullest expression of his renewed form of conception of justice after his political turn, the idea of free and equal moral persons still plays a prominent and foundational role in his theory (see, for example, PL 257-288). In “Justice as Fairness: Political not Metaphysical”, however, Rawls, in accordance with his political turn, emphasizes that although this conception of the person is admittedly a moral conception, it is “adapted to a political conception of justice and not to a comprehensive moral doctrine” (JFPM 232). The conception is thus “in effect a political conception of the person” (JFPM 232). For more about Rawls’s political turn, please see Why Political Liberalism: On John Rawls’s Political Turn (Weithman 2010). For more about the general problem of starting points in political constructivism, please see “Starting Points: Kantian Constructivism Reassessed” (Bagnoli 2014).
the increasing diversity of cultures in our societies and the general tendency to avoid metaphysics in academia, Kant interpreters usually find themselves torn between Kant’s own claims and prevailing liberal beliefs. In Kant’s own view, his political thoughts (or his political philosophy) stem from his moral philosophy, which is an integral part of his critical philosophy. His critical philosophy is in turn “the only one true system of philosophy” in his view, before which “there was as yet no philosophy at all” (see MM 6:207). However, it is now increasingly difficult to imagine how one can possibly convince everyone that Kant’s philosophy is the only correct philosophy for us.

As I mentioned in the introduction of this dissertation, this research is basically triggered by the tension in Rawls’s theory on one hand and the challenge faced by Kant interpreters on the other. It is the premise of the whole research that these two issues, which have more far-reaching implications than our interpretations of Kant and Rawls, can be resolved only with a decisive account of the connection between the moral and the political. While the previous chapters already served to tackle the problem from different angles, this last chapter will offer yet another perspective to the problem, which I believe would be most acceptable to those who are most accustomed to the liberal way of thinking.

In essence, the objective of this chapter is as follows: to interpret Kant’s philosophy starting from a purely political conception of the human being in his writings and ending in his moral philosophy. The advantage of this new interpretive approach is twofold. First, it effectively avoids the charge that Kant’s political philosophy is derived from a single comprehensive moral doctrine that might not be endorsed by everyone in a pluralistic society, for the new interpretive approach starts with a political conception of the human being. Second, it does not thereby betray Kant’s original plan of having his moral philosophy serving as
the ground of his political philosophy, for the new interpretive approach aspires to end with his moral philosophy. In other words, the interpretive approach will establish that the political conception of the human being found in Kant requires his moral philosophy.

This chapter will consist of the following sections. In the next section, I will contrast the justificatory approaches taken by Kant and Rawls and explain in further detail why the new interpretive approach to Kant’s moral and political philosophy points to a promising resolution. In the third section I will specifically review the controversies surrounding Rawls’s choice of starting points in his constructivist theory of justice. The review will further establish the controversial nature of the constructivist approach to justice. The fourth section will be a brief review of the general reception of Kant’s philosophy in the liberal tradition, in which Rawls’s partial assimilation plays a prominent and leading role. The fifth section introduces a political conception of the human being in Kant’s writings. The last two sections detail the alternative interpretive approach with the political conception of the human being as its starting point. Towards the end we will see why Kant’s moral doctrine is necessarily implied by his political philosophy.

2. The Third Way: To Approach the Moral from the Political

There are two distinguishable moral conceptions of the person in Rawls. Before his political turn, Rawls’s moral conception of the person is thicker. In A Theory of Justice (TJ), Rawls’s theory is arguably based on the premise that “[h]uman beings have a desire to express their nature as free and equal moral persons, and this they do most adequately by acting from the principles that they would acknowledge in the original position” (TJ 462-463). In TJ, the moral conception of the person is therefore not only assumed to be what reasonable
persons have a *desire* to express, but is also closely connected with Kant’s philosophical doctrine, for the original position is nothing but “a procedural interpretation of Kant’s conception of autonomy and the categorical imperative within the framework of an empirical theory” (TJ 226).

After his political turn, the moral conception of the person is considerably thinner. In PL, the idea of free and equal moral persons is largely transformed into a conception of citizen endowed with the moral powers of the reasonable and the rational (see PL 48). While the rational (which applies merely to a single agent in Rawls’s view) is indispensable in Rawls’s theory, the renewed form of political liberalism has its starting point mainly in the moral power of the reasonable, which accounts for the citizens’ *willingness* to propose and honor fair terms of cooperation on one hand, and their *willingness* to recognize the burdens of judgment and to accept their consequences on the other (see PL 49). This “thinner” moral conception of the person no longer makes direct reference to Kant’s philosophy. In PL, Rawls merely acknowledges that the *distinction* of the reasonable and the rational goes back to Kant (see PL 48).

By incorporating less in his moral conception of the person in his theory, Rawls seems to believe that the move could render his theory more readily acceptable in a democratic society marked by reasonable pluralism. He apparently fails to recognize, however, that for a *moral* conception of the person to serve as a starting point for a conception of justice, there is a minimal requirement from

---

68 In PL, Rawls speaks of Kant’s moral constructivism at times and claims that “as for Kant’s moral constructivism… a correct moral judgment is one that meets all the relevant criteria of reasonableness and rationality incorporated into the categorical imperative procedure for testing maxims” (PL 114). His idea of the reasonable, however, is not identical to Kant’s, for Rawls states that he gives his idea of the reasonable “a more restricted sense” than Kant (PL 49). In fact, Rawls’s idea of the reasonable may be more closely connected to Scanlon’s principle of moral motivation, as Rawls himself hints at a footnote (see PL 49).
which it cannot possibly escape. In order to serve as a normative conception\(^{69}\) in a theory of justice, a moral conception of the person must at least include in it an account of the corresponding desire or willingness on the part of the subject, in the absence of which a conception of justice would be nothing but a castle in the air, whose binding force on the subject is inexplicable at best or non-existent at worst. The defensibility and acceptability of a conception of justice therefore also rest upon the defensibility and acceptability of the corresponding desire or willingness inherent in the theory.

Let us take Rawls’s moral conception of the person as an example. In the thicker conception of the person before his political turn, the subject is supposed to have a primary desire to express Rawls’s moral conception of the person (see TJ 462-463), in the absence of which the whole moral conception would be impotent in practice. In the thinner conception of the person after his political turn, the subject is supposed to have a primary willingness (see above or PL 49), without which the subsequent conception of justice would be nothing more than a mere thought. In both cases, the primacy of the desire or the willingness is not itself vindicated by Rawls’s theory but merely presupposed as an indispensable element in the moral conception of the person. In Rawls’s case, the defensibility and acceptability of his conception of justice is thus established only for reasonable persons, who by definition voluntarily share the desire or willingness he describes in his theory (see PL 394-395). Rawls’s theory thereby excludes all the unreasonable persons and unreasonable doctrines by definition,\(^{70}\) without

\(^{69}\) As Rawls himself admits in JFP, 232.

\(^{70}\) To my knowledge, Rawls himself never speaks of unreasonable persons in his major writings. However, I share Freeman’s view that Rawls does not seek to address unreasonable persons or unreasonable doctrines (Freeman 2003,40). In other words, unreasonable persons are essentially excluded from Rawls’s theory, for they are not conformable to his starting point, i.e. his
giving an account at the same time as to why and how they ought to become reasonable. His moral, and supposedly normative, conception of the person is thus a conception borrowed from experience, which merely appeals to what a subject already desires or wills in experience, but does not seek to explain why and how such a desire or willingness is also binding on the will of the subject. In other words, Rawls’s theory of justice has no explanatory, let alone normative, value when we have an overwhelming number of unreasonable persons in a society who do not share the desire or willingness he presupposes.\(^1\)

Kant’s concept of right, in contrast to Rawls’s conception of justice, is meant to be applicable and binding on the will of every human being, without regard to what a subject happens to desire at any given moment. This universality is possible because Kant does not start with a moral conception of the person that is merely contingent,\(^2\) but with a fact of reason that is equally real to all human beings. In Kant, the fact of reason refers to our immediate consciousness of the moral law (as soon as we draw up maxims of the will for ourselves) (see CPrR 5:29), in which reason “presents [the moral law] as a determining ground [of our will] not to be outweighed by any sensible conditions” (CPrR 5:29-30). Reason therefore “of itself (not in the service of the inclinations) determines the will” and qualifies as a true higher faculty of desire (CPrR 5:24-25), which can therefore legitimately assume the role of the source of normativity. In contrast to Rawls’s approach, Kant thereby elucidates not a mere moral conception of the person but

---

\(^1\) Freeman admits that, in Rawls’s view, “there is no presumption that Social Darwinists, fundamentalists, neo-Nazis, or Southern slaveholders would be amenable to public reason; nor should any effort be made to address or accommodate their views” (Freeman 2003, 40).

\(^2\) Rawls’s moral conception of the person is contingent because it does not necessarily apply to all human beings.
an account of *personality*:

The idea of the moral law alone, together with the respect that is inseparable from it, cannot be properly called a predisposition to personality; it is personality itself (the idea of humanity considered wholly intellectually). The subjective ground, however, of our incorporating this incentive into our maxims seems to be an addition to personality, and hence seems to deserve the name of a predisposition on behalf of it. (Rel 6:28)

In Kant’s account of *personality*, the *respect* for the moral law of reason is an integral part and is also the sole and the undoubted moral incentive (CPrR 5:79). Its being the incentive cannot be proved (or disapproved) by any examples (or examples to the contrary) in experience, for the respect arises not from anything borrowed from experience but from the moral law itself, which “lets us discover the sublimity of our own supersensible existence and subjectively effects respect for [our] higher vocation” (CPrR 5:88). The respect for the moral law, as an integral part of *personality*, is therefore something we are never able to lose as a human being (Rel 6:46). In other words, the incentive, i.e. the respect for the moral law, is always there in every subject, and serves as the starting point from which the subsequent concept of right can be derived.\(^{73}\)

Now we have two radically different approaches by the two philosophers. While Rawls presupposes a *desire* or a *willingness* in the subject in order to vindicate his conception of justice, Kant tells us that all human beings indeed have a *higher faculty of desire* from which a concept of right can be derived. The

---

\(^{73}\) Please see Chapter I for a detailed discussion.
apparent conflict between the two approaches is characterized by two conflicting claims. First, with regard to Rawls’s approach, Kant might say: when it is not grounded by a moral philosophy, as in Rawls’s case, political philosophy cannot legitimately invoke a *moral* or *normative* conception of the person as a *starting element*, regardless of its role in the *justification* of the whole doctrine, for it is controversial not because it attempts to justify the doctrine but because it presupposes a *desire*, or *willingness*, on the part of the subject.74 Second, with regard to Kant’s approach, Rawls might hold that: Kant’s philosophical doctrine fails to offer an adequate conception of justice that can well order a constitutional democracy, for “a continuing shared understanding on one comprehensive religious, philosophical or moral doctrine can be maintained only by the oppressive use of state power” (PL 37). This amounts to the fact of oppression and directly contradicts with the essential nature of the democratic political culture. In other words, while Kant might contend that a political philosophy without a corresponding moral doctrine is untenable, Rawls maintains that a political philosophy dependent on a specific moral doctrine is not sustainable without the oppressive use of state power.

This chapter does not seek to argue for Kant’s approach and against Rawls’s, or vice versa. Instead, for the purpose of this chapter, we only need to note that the conflict between the two arises from their different views on the role of the *moral* in the *political*. In both Kant and Rawls, the *moral* (whether it be a full-fledged

74 This claim can be compared with Forst’s idea that “the moral point of view must combine cognitive (the capacity for justification), volitional (willingness to give justification and act justifiably), and affective (the sensorium for moral violations) components” (Forst 2012, 39). In other words, from a Kantian perspective, the moral conception of the person in Rawls is controversial in its volitional components. The lack of the affective element is also controversial, but it is beyond the ambit of this chapter. For more, please see “Desiring Justice: Motivation and Justification in Rawls and Habermas” (Krause 2005).
moral doctrine or a moral conception of the person) notably serves both as a starting and a normative element in the political, and this fact is arguably the underlying issue in the conflict between the two. The Kantian critique against Rawls introduced above is essentially that, in the absence of a supporting moral philosophy, Rawls cannot legitimately start with a moral conception that is normative. On the other hand, the Rawlsian critique against Kant is basically that the political cannot start with a normative doctrine of moral philosophy that might not be accepted by everyone. Their conflict, therefore, lies not in their disagreement over the starting point or the source of normativity but over the combination of both.

Logically, there are two possible solutions to the deadlock. First, in response to the Kantian critique against Rawls, one may propose starting points that presuppose less normatively. This aspiration is clearly the motivation behind the constructivist approach to justice initiated by Rawls. In the next section, I will review the controversies surrounding the approach and illustrate that the approach might be no less controversial than the conflict the constructivists set out to resolve. Second, in response to the Rawlsian critique against Kant, one may reverse the commonplace interpretive approach by approaching the moral from the political, thereby leaving aside the normative in the interpretive starting point. This potential solution to the conflict between Kant and Rawls is the primary object of this chapter. It will attempt to interpret Kant’s philosophy starting from a merely political conception of the human being extracted from his various writings on history and politics, leaving aside the problem of normativity in the beginning, and see if it will lead us to Kant’s moral philosophy and its normative implications in the end when it is absolutely necessary. Before we get to this, however, I need to show that the other possible solution, i.e. the constructivist
approach, is as problematic as the conflict it aspires to resolve.

3. **Controversies Surrounding Constructivist Normative Starting Point**

In his attempt to find an alternative to realism and relativism, Rawls invokes a peculiar assimilation of Kant, in which the metaphysical insights of Kant are bracketed and a procedural interpretation of Kant’s conception of autonomy is introduced. This unusual justificatory strategy is later named Kantian constructivism, a term coined by Rawls in his seminal essay “Kantian Constructivism in Moral Theory”. In Rawls’s definition, Kantian constructivism invokes “a particular conception of the person as an element in a reasonable procedure of construction”, which is to construct “the content of the first principle of justice” (KC, 516). The conception of the person therefore serves as a starting point in Kantian constructivism. The strategy is purported to offer an alternative to “familiar traditional moral conceptions, such as utilitarianism, perfectionism, and intuitionism”, and to serve as “the Kantian roots” (515) of Rawls’s conception of justice as fairness in TJ. Although Rawls never fails to stress that his whole project is situated in the context of a democratic society with deeply shared notions, principles and convictions (KC 518), his Kantian constructivism, which is said to justify his conception of justice from minimal and non-controversial starting points, is eagerly applauded by those in dire need of new justificatory resources to defend liberal universalism against relativist and post-modern critics in a wider, universalist context.\(^{75}\)

\(^{75}\) This is not surprising, for Rawls asserts that, in Kantian constructivism, the task of “justifying a conception of justice is not primarily an epistemological problem” and that “[t]he search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society replaces the search for moral truth interpreted as fixed by a prior and
Kantian constructivism subsequently becomes the focus of a growing debate on the validity of its starting points. On one hand, communitarian critics contend that Rawls’s Kantian constructivism in TJ illegitimately invokes a metaethical starting point, i.e. the moral conception of the person, which is independent from social convention and shared understanding in the community. In their view, such a starting point is not justifiable (see, for example, MacIntyre, 1984, 232-233; Sandel, 1984, 9). Similar charge is not only raised by these outright critics of Kantian constructivism, but also by one of Rawls’s students, O’Neill, who contends that Rawls fails to give a constructivist justification for the starting point, i.e. the procedures of practical reason itself, and that Rawls’s Kantian constructivism is thus constructivist in a limited sense at best (see O’Neill, 2000, 2003a, 2003b). On the other hand, as a generalization of the above criticism against the specific starting points adopted by Rawls, there are also contentions that the agenda of political constructivism is essentially self-defeating. In the views of these critics, political constructivism inevitably relies on *unconstructed* starting points in order to avoid arbitrariness and hence it cannot be coherent with its own constructivist claim (see Cohen 2003; Raz 1990, 15; 2003; Timmons 2003). After all, as Rawls himself asserts, “not everything, then, is constructed; we must have some material, as it were, from which to begin” (PL 104). This generalized challenge is thus not targeted at any particular starting points adopted by any specific thinker, but at the fundamental agenda of political constructivism in general. Various defenses are proposed in response to these objections. For independent order of objects and relations, whether natural or divine, an order apart and distinct from how we conceive of ourselves” (KC 519). For many, Kantian constructivism seems to offer a promising answer to the debate between realism and relativism, by retaining the benefits of the former while avoiding its epistemological and ontological costs (Darwall et al. 1992; Shafer-Landau 2003; Timmons 2003; Hussain & Shah 2006; Enoch 2009).
instance, O’Neill’s accusation is accompanied with a variant of Kantian
together with Rawls’s and Korsgaard’s versions (1996), constitute the three most
influential approaches to Kantian constructivism.

O’Neill’s constructivism is often considered a systematic and philosophically
rigorous attempt to develop a more universal alternative to Rawls’s (see Sutch
2001, 86; Budde 2009) and it is developed out of two dissatisfactions with
Rawls’s, of which Budde (2009) gives a good account. First, “Rawls, [O’Neill]
conjectures, does not give a constructive justification of the procedures of
practical reason itself, but instead offers only a coherentist justification for the
procedures…, which relies on shared intuitions and considered judgments” (see
Budde, 2009, 201). Second, “O’Neill also argues that Rawls fails to satisfy the
constructivist criteria of using only minimal and abstract premises or starting
materials to construct principles (O’Neill, 1996, 38–40) by using in fact
‘idealized’ and therefore, according to O’Neill, illegitimate starting materials”
(Budde, 2009, 201-202). O’Neill is thus prompted to develop a variant form of
Kantian constructivism in which the alleged non-ideal starting point is a recursive
principle of the (self-)justification of reason that is closer to Kant’s philosophy in
her interpretation (see O’Neill, 1989, 21-27). O’Neill’s constructivist position is
best illustrated by the following passage:

Yet it is not easy to see why Kant thinks the CI procedures constitute
requirements of practical reason or reasonableness. It may be that Kant
and Rawls may end up with parallel difficulties, that both propose a
constructive procedure for identifying specific practical principles, but
both fail to show convincingly that this procedure is grounded in or
expresses (an adequate conception of) reasonableness. Their positions may be constructive in the limited sense that they propose procedures which agents can use to establish principles for guiding action but not in the fuller sense of justifying those procedures and grounding objective normative judgments. If so, neither Rawls or Kant will justify ethical claims in the stronger sense to which constructivists aspire. (O’Neill 2003b, 355)

O’Neill’s contention against Rawls is forceful here, for it points out a fundamental weakness in Rawls’s Kantian constructivism: it merely proposes procedures, but does not justify the objective necessity of those procedures. A similar charge could be brought against the conception of the person in Kantian constructivism, for it is merely one of the many conceptions a free citizen might choose to adopt. The objective necessity of adopting that particular conception specified by Rawls is also not justified. In Rawls’s theory, the only justification available is, as Rawls’s puts it, “a matter of the mutual support of many considerations, of everything fitting together into one coherent view” (TJ 19 and cf. 507). For Rawls, the justification only “rests upon the entire conception and how it fits in with and organizes our considered judgments in reflective equilibrium” (TJ 507). O’Neill is thus entirely correct when she argues that Rawls offers only a coherentist justification for the procedures, which relies merely on shared intuitions and considered judgments.

Unfortunately, although her diagnosis is arguably accurate, O’Neill’s variant form suffers from the same weaknesses she identifies in Rawls.76 In the second

76 For more, please see “Constructivism all the way down – Can O’Neill succeed where Rawls failed?” (Budde 2009).
chapter, I already discussed O’Neill’s variant in detail. For the purpose of the present chapter, it suffices to note that, in place of Rawls’s moral conception of the person, O’Neill invokes a normative conception of practical reasoning as a starting element in her constructivism (see O’Neill 1996, 38-65). The presupposition of a normative conception of practical reasoning in O’Neill’s theory thereby leads to a difficulty that is not different from the one suffered by the presupposition of the normative conception of the person in Rawls’s. In both cases, the presupposition of a normative conception is itself unjustified. In other words, no explanation is given as to why the normative conception is binding on the will of a rational being.

As is evident from the above discussion, Kantian constructivism, or political constructivism in general, finds itself entangled in an apparently unresolvable antinomy: the aspiration to develop a freestanding view of a political conception of justice is met with the requirement to justify itself out of thin air. This paradoxical agenda is not an invention of Rawls, but one that is prevalent in the post-metaphysical age to which Rawls’s Kantian constructivism merely seeks to respond. Its paradoxical nature, however, is one that can be alleviated, if not tamed, if we note the following two points. First, according to Rawls, the adjective freestanding is used to describe a political conception that “is not presented as derived from, or as part of, any comprehensive doctrine” (PL xlii). The freestanding requirement therefore does not demand that the political conception be absolutely free from any dependence, or even less that the conception has no ground. In other words, the word freestanding should be strictly understood in connection to Rawls’s concept of comprehensive doctrine, which I will discuss in the next section. Second, the suggestion that Kantian constructivism should have its starting points justified constructively is misleading
and renders the justificatory method self-defeating. At its best, the endless pursuit of *self-justifiable starting points* can only lead to the recursive principle of the *(self-)*justification of reason proposed by O’Neill (see O’Neill, 1989, 21-27), which might be self-fulfilling in *theory* but impotent in *practice*. Put differently, O’Neill clearly has her focus misplaced when she sets out to look for a *self-justifiable starting point* rather than a *self-binding law of practical reason*. After all, the fact that a particular principle is able to justify itself (self-fulfilling in theory) does not imply that the principle is *unconditionally binding* on the will of a rational being (potent in practice). In other words, she notably and wrongly recast a *practical* problem into a *theoretical* problem.

These are thus two arguments which I believe may alleviate the problems faced by the constructivist approach to justice. This chapter will not pursue any further the implications of these arguments but will consider their implications on the liberal reception of Kant’s philosophy. In the next section, I will show that Rawls’s partial assimilation of Kant’s philosophy is informed by the two questionable liberal and constructivist beliefs mentioned above. First, Rawls classifies Kant’s philosophy as an example of a *comprehensive doctrine* all too quickly and prematurely. Second, not unlike O’Neill, Rawls tends to adopt a *theoretical* (rather than *practical*) way of thinking in his interpretation of Kant’s practical philosophy. It results in a series of misreadings and renders a complete assimilation of Kant’s philosophical insights impossible.

4. **Rawls’s Verdict against Kant’s Philosophy**

Kant, as an Enlightenment thinker, and with his critical philosophy built upon the keystone of the concept of freedom (see CPrR 5:4), was not consistently regarded as one among the most important thinkers in the liberal tradition along
with Hobbes, Locke, Rousseau and Mill, until Rawls’s adaptation of a moral conception of the person in his Kantian constructivism. However, in Rawls’s own words when he explains the idea of Kantian constructivism: “the adjective ‘Kantian’ expresses analogy and not identity; it means roughly that a doctrine sufficiently resembles Kant’s in enough fundamental respects so that it is far closer to his view than to the other traditional moral conceptions that are appropriate for use as benchmarks of comparison” (KC 517). In other words, as is well-known, the adaptation of Kant’s philosophy in Rawls is only partial.

The adaptation of Kant is merely partial because, in Rawls’s view, Kant’s moral doctrine is an example of a comprehensive moral doctrine from which his political conception of justice should be independent. Comprehensive doctrine is an important concept in Rawls’s project and it is first introduced in JFPM when he tries to argue that his moral conception of the person is in effect a political conception:

…the conception of the person is a moral conception, one that begins from our everyday conception of persons as the basic units of thought, deliberation and responsibility, and adapted to a political conception of justice and not to a comprehensive moral doctrine. It is in effect a political conception of the person, and given the aims of justice as fairness, a conception of citizens (JFPM, 232).

In contrast to this moral, yet political, conception of the person in his own theory, Rawls goes on to assert that a conception of the person in a comprehensive moral doctrine is “a moral ideal to govern all of life”, the pursuit of which would be “incompatible with other conceptions of the good, with forms of personal,
moral, and religious life consistent with justice and which, therefore, have a proper place in a democratic society” (JFPM, 245). This assertion is in turn derived from his claims regarding the essential nature of reasonable comprehensive doctrines, which can be summarized as follows:

1. Reasonable comprehensive doctrines are those comprehensive doctrines reasonable persons affirm (PL 59).
2. Reasonable comprehensive doctrines are incompatible with each other (PL xvi-xviii).
3. A reasonable doctrine is an exercise of theoretical reason: “It covers the major religious, philosophical, and moral aspects of human life in a more or less consistent and coherent manner. It organizes and characterizes recognized values so that they are compatible with one another and express an intelligible view of the world.” (PL 59)
4. It is also an exercise of practical reason: It “[singles] out which values to count as especially significant and how to balance them when they conflict.” (PL 59)
5. It “is not necessarily fixed and unchanging, it normally belongs to, or draws upon, a tradition of thought and doctrine. Although stable over time, and not subject to sudden and unexplained changes, it tends to evolve slowly in the light of what, from its point of view, it sees as good and sufficient reasons.” (PL 59)

---

77 For the purpose of this chapter, I will only consider the concept of reasonable comprehensive doctrine, to which Kant’s moral philosophy belongs in Rawls’s view.
In Rawls, the above definition of reasonable comprehensive doctrines is in turn meant to include Kant’s philosophy and Kant’s philosophy is therefore associated with the fact of oppression:

…a continuing shared understanding on one comprehensive religious, philosophical, or moral doctrine can be maintained only by the oppressive use of state power. If we think of political society as a community united in affirming one and the same comprehensive doctrine, then the oppressive use of state power is necessary for political community. In the society of the Middle Ages, more or less united in affirming the Catholic faith, the Inquisition was not an accident; its suppression of heresy was needed to preserve that shared religious belief. The same holds, I believe, for any reasonable comprehensive philosophical and moral doctrine, whether religious or nonreligious. A society united on a reasonable form of utilitarianism or on the reasonable liberalisms of Kant or Mill, would likewise require the sanctions of state power to remain so. Call this “the fact of oppression.”

(PL 37)

With a view to developing a political conception of justice, Rawls thereby classifies all non-political, religious, philosophical, or moral doctrines as reasonable comprehensive doctrines and denies the possibility of uniting a society on these doctrines without the oppressive use of state power. It is clearly impossible to consider the plausibility of Rawls’s claim in every case, but at least in Kant’s case, Rawls’s verdict is highly dubious. Notwithstanding all the thoughtful elaborations quoted above, it is still not easy to see why and how
Rawls came to the conclusion that a society united on Kant’s philosophy would require “the sanctions of state power to remain so”. In Kant’s own view, it is not only unnecessary but also impossible to unite a society on his moral philosophy through the coercive force of the state. In fact, there is probably nothing more foreign to Kant than the idea of coercing others to be moral, for it is clearly impossible to coerce others to respect the moral law, let alone to make this respect the incentive of their maxims (see CPrR 5:83). There is also nothing like the Catholic faith in Kant’s philosophy which requires the affirmation of everyone in order to unite a society. In Kant’s philosophy, a people can only be united under a general legislative will (MM 6:318), which is nothing but the united will of the people (MM 6:313). Nowhere in Kant’s writings does he claim that a people need to affirm his philosophy in order to be united. Therefore, Kant would unhesitatingly reject the suggestion that a society be united on his philosophy by the sanctions of state power, for it is not only impossible but also absurd in his view.

Rawls’s verdict on Kant’s philosophy thus requires at least a careful reexamination before we can determine once and for all that a society united on Kant’s philosophy requires the oppressive use of state power, for it is questionable whether Kant fits into Rawls’s description of comprehensive doctrines as readily as Rawls seems to suggest. Rawls’s assertion is even more dubious if we consider the conclusion of chapter two, which is that his procedural interpretation of Kant’s notion of autonomy is significantly problematic. The credibility of his verdict

---

78 In fact, Kant goes so far as to warn that philosophers can easily confuse their judgment “by a mass of considerations foreign and irrelevant to the matter and deflect it from the straight course” and that it would be more advisable to leave moral matters to the better judgment of common reason of average people (see Gr 4:404).
could also be subject to further challenges due to the overwhelming emphasis he places on Kant’s moral writings and the complete absence of Kant’s political philosophy both in his interpretation and his partial assimilation of Kant.\textsuperscript{79} Rawls’s claim is therefore far from being conclusive on this matter.

Owing to the popularity of Rawls’s theory, his verdict on Kant is generally accepted by most, if not all, modern liberal thinkers. As a result, we have an interesting state of affairs in Rawls’s theory and in modern political philosophy in general. The growing interest in Kant’s moral philosophy, owing to a large extent to Rawls’s partial adaptation of Kant in his theory of justice, is restrained by the unequivocal denial of its potential foundational role in liberalism.\textsuperscript{80} Following the lead of Rawls, virtually all liberals are reluctant towards a thorough assimilation of Kant’s philosophical doctrine in their theories, for although the keystone of his critical philosophy is the concept of freedom, which they gladly applaud, the concept is at the same time tied to his critical philosophy as a whole. Fragments of Kant’s critical philosophy are therefore often recast in liberal theories out of their original contexts in Kant to support various liberal claims.\textsuperscript{81}

\textsuperscript{79} Rawls’s partial assimilation of Kant in his Kantian constructivism is mainly from Kant’s moral writings, especially \textit{Groundwork of the Metaphysics of Morals} and \textit{Critique of Practical Reason}. Rawls never engages seriously with the appropriate counterparts in Kant, i.e. Kant’s various political writings, especially the \textit{Doctrine of Right} (see Höffe, 2002, 228). On the very few occasions where Rawls refers to these writings, the references are of very minor importance (see e.g. TJ, 11, note 5).

\textsuperscript{80} For a more detailed account of the reception of Kant’s philosophy in mainstream liberalism, please read \textit{Kant and Modern Political Philosophy} (Flikschuh 2000), especially its introduction (1-11).

\textsuperscript{81} For example, in defense of individual freedom, it is customary to quote Kant’s famous saying: “… to treat himself and all others never merely as means but always at the same time as ends in themselves” (Gr 4:433). See, for example, \textit{Equality and Partiality} (Nagel 1991, 159) and “The Limits of Aristotelian Ethics” (Larmore 1992, 193-194). In defense of the freedom of the use of reason, it is convenient to quote Kant’s saying on the freedom of critique: “The very existence of reason depends upon this freedom, which has no dictatorial authority, but whose claim is never
This chapter does not seek to engage directly with Rawls’s verdict on Kant. However, the suspicions outlined above should at least allow one to argue that the negative verdict given by Rawls is not conclusive enough to eliminate entirely the possibility of a verdict to the contrary and to repudiate the new interpretive approach introduced in the rest of this chapter. The above discussion also makes it clear that most modern liberal political philosophers are not ready to accept the usual interpretation of Kant’s philosophy starting from the moral and ending with the political. The current state of affair thus necessitates the following interpretive approach to Kant’s practical philosophy starting from the political and ending with the moral. In the rest of this chapter, I will outline the alternative interpretive approach to Kant’s political philosophy for the reference of modern liberal thinkers, or political philosophers in general. I will start with the interpretive starting point of the new approach, namely, a political conception of the human being in Kant, in the following section.

5. A Political Conception of the Human Being in Kant

As I stated at the beginning, the aim of this chapter is to introduce an alternative interpretive approach to Kant’s practical philosophy that starts with the political and ends with the moral. The new approach will therefore start with a merely political conception of the human being, leaving aside the problem of normativity in the beginning, and see if it will lead us to Kant’s moral philosophy and its normative implications in the end when it is absolutely necessary. A peculiar feature of this starting conception of the person is therefore that Kant’s anything more than the agreement of free citizens, each of whom must be able to express his reservations, indeed even his veto, without holding back” (A738-739/B766-767). For instance, it is a keystone of O’Neill’s arguments in her Constructions of Reason (O’Neill 1989), see 15, 37 and 57.
account of *personality* be left aside for a moment in the beginning. In other words, the idea of the moral law (and the respect that is inseparable from it) (see Rel 6:28) will not be considered at the beginning of the interpretive approach. The starting point of the interpretive approach is thus a *hypothetical* situation in which the lawgiving capability of practical reason expounded by Kant is not invoked.

The task of formulating such a starting conception of the person is more straightforward than one might expect, for it is readily available in exactly the same place where Kant introduces the notion of *personality*. In a section titled “Concerning The Original Predisposition to Good in Human Nature” in his *Religion within the Boundaries of Mere Reason*, Kant introduces three kinds of *predisposition* in human nature, including the predisposition to *animality*, the predisposition to *humanity* and the predisposition to *personality* (see Rel 6:26). In his view, the three predispositions\(^{82}\) constitute “the constituent parts required for a [human being] as well as the forms of their combination that make for such a being” (Rel 6:28). Or in other words, “they belong to the possibility of human nature” (Rel 6:28). In Kant’s original account, a complete conception of the person thus consists of all the three predispositions mentioned above.

Among the three predispositions designated by Kant, the predisposition to *personality* represents “the susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (see Rel 6:26-28). It is the only one “rooted in reason practical of itself, i.e. in reason legislating unconditionally” (Rel

---

\(^{82}\) As Palmquist rightly notes, “Kant’s use of ‘predisposition’ sometimes in the singular and sometimes in the plural is the source of considerable confusion, some commentators assuming that he proposes three (or more) entirely distinct predispositions” (Palmquist 2015, 64). I will not take issue with this problem here in this chapter, but I share Palmquist’s view that the three kinds of predisposition should be seen as “classifications of our overall predisposition”, which should be in the singular (Palmquist 2015, 64).
6:28) and is therefore the only one among the three to be left aside in the merely political conception of the human being required for the alternative interpretive approach. In the absence of personality, the starting conception of the person required therefore consists of the remaining two predispositions, namely, the predisposition to the animality and the predisposition to the humanity of the human being.\(^{83}\)

In Kant’s own words, the predisposition to animality “may be brought under the general title of physical or merely mechanical self-love, i.e. a love for which reason is not required” (Rel 6:26). It entails the love for self-preservation, for the propagation of the species and for community with other human beings, i.e. the social drive (see Rel 6:26). On the other hand, the predisposition to humanity “can be brought under the general title of a self-love which is physical and yet involves comparison (for which reason is required); that is, only in comparison with others does one judge oneself happy or unhappy” (Rel 6:27). This predisposition, unlike the predisposition to animality, “is rooted in a reason which is indeed practical, but only as subservient to other incentives” (Rel 6:28).

The combination of these two predispositions depicts a conception of the human being that is highly political in nature. The predisposition to animality does not require reason and yet includes in it our love not only for self-preservation but also for the preservation and propagation of the whole species. The predisposition thus locates our care for the whole species in our instinct instead of our rationality. The inclusion of our social drive in the predisposition to animality is especially remarkable, for it implies, as Palmquist

\(^{83}\) Kant also introduces three different grades of a natural propensity to evil in human nature, but since they presuppose morality (see Rel 6:29-30), they are not considered in my starting conception of the person here.
rightly notes, that our love “for forming social bonds with other members of our species” does not require our *rational capacity* and has to do with our *instinct* as human beings (see Palmquist 2015, 66). The predisposition to *humanity* is even more *political* in nature, for it points to an original predisposition in human nature to judge one’s own happiness only in comparison with others’ and therefore unveils an *original* or even *instinctive* demand of everyone to be at least as fortunate as others (see Palmquist 2015, 68).

The instinctive love for self-preservation, the instinctive love for propagation of the species, the instinctive love for community with others and the instinctive demand to be at least as fortunate as others together constitute a *primitive political conception of the human being* which can serve as the starting point of the new interpretive approach introduced in this chapter. However, they originate from only *two* of the *three* kinds of original predisposition identified by Kant (in the absence of *personality*) and therefore by themselves do not constitute a *complete* account of our original predisposition to *good*. In other words, in Kant’s view, they belong to the *constituent parts* required for a human being (see Rel 6:28) and yet are not *sufficient* to represent the *whole* original predisposition to good in human nature. 84

This clarification is especially important if we consider the fact that, according to Kant, human beings can indeed *use* these two predispositions *inappropriately* (see Rel 6:28). In fact, as soon as he introduces the two predispositions in *Religion*, Kant promptly warns in the same passage that *all sorts of vices* can be *grafted* on them, although the vices do not of themselves

84 I share Palmquist’s interpretation that in Kant’s view “[o]ur original predisposition to good is not a ‘contingent’ possibility; it is not one that we could do without and still be human persons…” (Palmquist 2015, 71).
issue from the predispositions as a root (see Rel 6:26-27). Specifically, Kant
warns that on the predisposition to animality can be grafted the vices of the
savagery of nature and the bestial vices of gluttony, lust and wild lawlessness (in
relation to other human beings). (Rel 6:27) Similarly, with regard to the
predisposition to humanity, Kant warns in a long passage that:

Out of this self-love [in the predisposition to humanity] originates the
inclination to gain worth in the opinion of others, originally, of course, merely equal worth: not allowing anyone superiority over oneself, bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others. - Upon this, namely, upon jealousy and rivalry, can be grafted the greatest vices of secret or open hostility to all whom we consider alien to us. These vices, however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure; for nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture. Hence the vices that are grafted upon this inclination can also be named vices of culture, and in their extreme degree of malignancy (where they are simply the idea of a maximum of evil that surpasses humanity), e.g. in envy, ingratitude, joy in others’ misfortunes, etc., they are called diabolical vices. (Rel 6:27)

I am quoting this long passage in its entirety here for it concisely summarizes
the various vices that are highly relevant in the new interpretive approach. Not unlike the vices grafted on the predisposition to animality, the vices mentioned in the passage are also only grafted on the predisposition to humanity. The role of the two predispositions in the occasion of the vices is nothing more than being the origin of the corresponding inclinations from which the vices arise. Although the vices would cease to exist in the absence of the two predispositions and the inclinations associated with them, the very idea of a human being and the whole idea of a predisposition to good in human nature would also be unimaginable in the absence of the same. Kant therefore finds it incorrect to attribute the occasion of the vices to the two predispositions or even the inclinations associated with them, for it is unmistakably clear that the two predispositions and the inclinations can be used appropriately (or else Kant would not warn that they can be used inappropriately) (see Rel 6:28).

Without going too deep into the relationship between the vices, the inclinations and the two predispositions, we only need to note (for the purpose of the political conception of the human being) that these two predispositions make possible certain vices that must be addressed in a society and that these vices arise from the misuse of the two predispositions by the subjects. In other words, even before we consider the predisposition to personality (which alone cannot be misused among the three), human beings cannot legitimately attribute the occasion of the vices to their human nature but only to themselves. After the section on the predispositions to good in Religion, Kant therefore introduces the idea of the propensity to evil in human nature, which he asserts can be thought of “as brought by the human being upon himself” (Rel 6:29). For the purpose of this chapter, we do not need to bother with the distinction between the predisposition to good and the propensity to evil in human nature. All we need to know in the
political conception of the human being required for this chapter is that it consists of a predisposed love for community with other human beings, a predisposed demand to be at least as fortunate as other human beings and an inevitable tendency to misuse our predispositions, which results in an unjust desire to acquire superiority for oneself over others (see Rel 6:27). The political implications of the combination of these features in the political conception of the human being can be represented in a single concept in Kant, namely, the unsociable sociability of human beings. In the following section, I will illustrate why the unsociable sociability of human beings leads to the concept of public will in Kant’s philosophy.

6. Lawless Freedom, the Public Will and the Principle of Right

In *Idea for a Universal History*, Kant defines the unsociable sociability of human beings as follows:

…their propensity to enter into society, which, however, is combined with a thoroughgoing resistance that constantly threatens to break up this society. The predisposition for this obviously lies in human nature. The human being has an inclination to become socialized, since in such a condition he feels himself as more a human being, i.e. feels the development of his natural predispositions. But he also has a great propensity to individualize (isolate) himself, because he simultaneously encounters in himself the unsociable property of willing to direct everything so as to get his own way, and hence expects resistance everywhere because he knows of himself that he is inclined on his side toward resistance against others. (IUH 8:20-21)
Kant’s account of the unsociable sociability of human beings is a vivid depiction of the inevitable antagonism in society as a result of the misuse of the two predispositions discussed in the last section. The instinctive love for community and the instinctive demand to be at least as fortunate as others coexist with a constant fear of others attaining a hateful superiority over us (see Rel 6:27). The latter fear cannot be driven away by the love for community with others, and the love for community cannot be eradicated by the fear either. Kant therefore asserts that unsociable sociability lies in human nature (IUH 8:20) and claims that our fellows in our society are someone we cannot stand, but also cannot leave alone (IUH 8:21). An inevitable antagonism in society thus arises, for out of this constant fear of being inferior to others grows the unjust desire to direct everything so as to get one’s own way (IUH 8:21), either as a preventive measure to ensure one’s own safety or as a fulfillment of the unrestrained love for superiority over others.

An important characteristic of the unsociable sociability of human beings is its inherent contradictory nature, for one’s desire to be the master of others clearly contradicts the same desire of others. In view of this, in Idea for a Universal History, Kant goes on to argue that this amounts to a misuse of freedom:

… the human being is an animal which, when it lives among others of its species, has need of a master. For he certainly misuses his freedom in regard to others of his kind; and although as a rational creature he wishes a law that sets limits to the freedom of all, his selfish animal inclination still misleads him into excepting himself from it where he may. (IUH 8:23)
The misuse of freedom is a recurring subject in Kant, and it appears under different names in different writings, namely, “barbaric freedom” (IUH 8:26), “reciprocally conflicting freedom” (CPJ 5:432), “[to look merely at himself] in the use of his freedom” (CBHH 8:115), “savage (lawless) freedom” (PP 8:357), “wild, lawless freedom” (MM 6:316) and “unconditional freedom” (Anthro 7:327). What is common in all these depictions of a state of lawless freedom is that, in the absence of any opposing power, the abuse of such freedom inevitably leads to a state of constant antagonism in which the lawless freedom of each unavoidably collides with the same freedom of others. In Kant, such a state of constant antagonism is called a state of nature and it is not very different from the one depicted by Hobbes. In Towards Perpetual Peace, Kant specifically contrasts the state of nature, which is a condition of war, with a condition of peace:

A condition of peace among men living near one another is not a state of nature (status naturalis), which is much rather a condition of war, that is, it involves the constant threat of an outbreak of hostilities even if this does not always occur. A condition of peace must therefore be established; for suspension of hostilities is not yet assurance of peace, and unless such assurance is afforded one neighbor by another (as can happen only in a lawful condition), the former, who has called upon the latter for it, can treat him as an enemy. (PP 8:438-439)

Hostilities stem from a restrained claim for unconditional freedom when a human being treats other beings merely as a means to his or her own end, and in so doing tries to be master over other beings who by nature are equal to him or her (see Anthro 7:327). Since the constant threat of an outbreak of hostilities
arises from the *unsocial sociability* of human beings, which is part of *human nature*, the threat is there before the actual experience of hostility. The condition of war, or the state of nature, is thus a *rational idea*, i.e. an idea that we learn not from experience but from human nature. In order to leave this condition of war, it is thus necessary to enter a *public lawful condition*, which must also be a *rational idea*, in opposition to the former:

It is not experience from which we learn of the maxim of violence in human beings and of their malevolent tendency to attack one another before external legislation endowed with power appears, thus it is not some deed that makes coercion through public law necessary. On the contrary, however well disposed and law-abiding human beings might be, it still lies a priori in the rational idea of such a condition (one that is not rightful) that before a public lawful condition is established individual human beings, peoples and states can never be secure against violence from one another, since each has its own right to do what seems right and good to it and not to be dependent upon another’s opinion about this. (MM 6:312)

This is how Kant argues for the necessity of *coercion through public law* in a *public lawful condition*. The following points are noteworthy in the passage. First, the *necessity* is hereby linked with our preference for a condition of peace over a condition of war. In other words, if one prefers the condition of war, the necessity

---

85 In fact, Kant goes so far as to argue that, in the state of nature, a human being “already wrongs me just by being near me… by which he constantly threatens me; and I can coerce him either to enter with me into a condition of being under civil laws or to leave my neighborhood” (PP 8:439).
could not be established. However, such a preference is considered impossible, for the condition of peace is essential for self-preservation. Second, Kant adds in the passage that apart from the malevolent tendency in the unsociable sociability of human beings, violence may also arise from well-disposed human beings when they want to pursue what seems right and good to them. In other words, violence may result not only from the tendency to seek to be master over other beings but also from conflicting judgments as to what is good to oneself and to others. Third, Kant also reiterates in the passage that we do not learn of the condition of war or the condition of peace from experience but from the a priori concept of human nature, in which we find the predisposition to animality, the predisposition to humanity and the tendency to misuse them. The necessity of coercion through public law and a public lawful condition is thus also established a priori.

Apart from the points above, the following implications of the idea of a public lawful condition should also be noted. First of all, as its name suggests, the idea of a public lawful condition is one that is governed by laws. This is an accurate description of the idea because the condition is one in which the lawless or inappropriate use of freedom (or the two predispositions in the political conception of the human being) is opposed by another power in the transition from the condition of war to the condition of peace. Since the opposing power against the lawless or inappropriate use of freedom can only deal with the external threats it brings about in the condition of war but not the underlying misuse of freedom by the subjects (for it is impossible to require a human being by external force to use his or her freedom appropriately), the means invoked in the public lawful condition can be nothing but coercion through public law (which dictates how external threats should be contained externally and lawfully).

Several implications thus follow from the last one. The public lawful
condition is a condition in which the freedom of each is limited by coercion through public law to the extent that the external threats posed to each other by their lawless use of freedom is contained. In other words, the idea of a public lawful condition invokes a peculiar concept of external lawgiving that does not (and indeed cannot) require the appropriate internal use of freedom on the part of every individual but requires merely that their freedom can coexist externally. In Kant, this peculiar concept of lawgiving belongs with the idea of a general legislative will or a public will, under which the peaceful coexistence of the freedom of all is possible. Put differently, in a public lawful condition, the public law is seen as the act of the public will (TP 8:294), with an aim to bring about the “limitation of freedom of each to the condition of its harmony with the freedom of everyone insofar as this is possible in accordance with a universal law” (TP 8:290).

The following principle of right therefore serves as a universal principle in the concept of public will and its act of external lawgiving:

Any action is right if it can coexist with everyone’s freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone’s freedom in accordance with a universal law. (MM 6:231)

A public law is thus considered just and as the act of the public will only

---

86 The term “general legislative will” can be found in MM 6:318 and MM 6:320. The term “public will” can be found in TP 8:294, TP 8:303 and PP 8:352. Kant also uses the following terms to convey the same idea: the “collective will” (MM 6:342), the “common will” (MM 6:257, 6:271, 6:273; TP 8:291; PP 8:371), the “common and public will” (TP 8:297), the “general will” (MM 6:259, 6:269, 6:294, 6:326, 6:329, 6:342; TP 8:292, 8:302, 8:304; PP 8:351, 8:352, 8:366, 8:378, 8:383), “the general (united) will” (TP 8:295) and “the united will” (MM 6:274, 6:302, 6:313, 6:314, 6:328, 6:338, 6:339; TP 8:297; PP 8:371).
when it is given in conformity with the principle of right (see TP 8:297). Coercion in accordance with a just public law is thereby a hindering of all the hindrances to freedom in accordance with universal laws (see MM 6:231), with an aim to attain a public lawful condition or a rightful condition (see TP 8:293), in which the external relations of human beings are governed by the principle of right. The concept of public will and the principle of right therefore together constitute the limiting condition of politics (see PP 8:372), i.e. the condition for the peaceful coexistence of the external freedom of all. The principle of right also brings with it an idea of the external lawfulness of freedom (see CBHH 8:118), away from which one’s use of freedom cannot possibly coexist with others. From this Kant infers that freedom in accordance with the principle of right is the only innate right:

Freedom (independence from being constrained by another’s choice),
insofar as it can coexist with the freedom of every other in accordance with a universal law, is the only original right belonging to every man by virtue of his humanity. (MM 6:237)

This section introduced how the political conception of the human being in the last section leads to the idea of a limiting condition of politics in Kant, represented primarily by the concept of public will and the principle of right. Admittedly, in Kant’s political writings, the implications of the limiting condition of politics, i.e. the principle of right, is further considered in its application to three distinct forms of politics, namely the right of citizens of a state (of individuals within a people), the right of nations (of states in relation to one another) and cosmopolitan right (see MM 6:311; PP 8:349). However, in order to
stay focused on the transition from the political to the moral in Kant, I will not
discuss the details involved here in this chapter. Instead, in order to sketch the
complete outline of this alternative interpretive approach to Kant’s political
philosophy, I will continue to examine why and how the limiting condition of
politics necessarily leads to the concept of morality in Kant’s moral philosophy.

7. From the Limiting Condition of Politics to Morality

In essence, the idea of a limiting condition of politics, i.e. the concept of right
discussed in the last section, arises from the unavoidable propensity in human
nature to misuse the predisposition to animality and the predisposition to humanity,
which, when considered specifically in the context of politics, amounts to a
misuse of freedom. However, the derivation in the last section merely established
the theoretical necessity of the limiting condition of politics. Nowhere in the
above discussion have we shown that the limiting condition is also binding on the
will of every human being. In other words, its practical necessity is not yet
vindicated. The distinction between theoretical necessity and practical necessity is
crucial because the absence of the latter gives rise to the problem that Kant
himself highlights in his *Idea for a Universal History with a Cosmopolitan Aim*:

For he certainly misuses his freedom in regard to others of his kind; and
although as a rational creature he wishes a law that sets limits to the
freedom of all, his selfish animal inclination still misleads him into
excepting himself from it where he may. (IUH 8:23)

The theoretical necessity of the limiting condition of politics can thus only
account for the desire of a human being for the limitation of the freedom of
everyone else. It does not thereby automatically account for the willingness of a human being to give up his or her own unlimited freedom in the public lawful condition. This conclusion is not surprising if we reconsider the nature of a public lawful condition carefully. As I argued in the last section, the fundamental problem the idea of a public lawful condition sets out to solve is the constant threats in the condition of war. The concept of external lawgiving in a public lawful condition does not (and indeed cannot) require the appropriate internal use of freedom on the part of every individual but requires merely that their freedom can coexist externally. Coercion through public law in a public lawful condition is therefore meant to be an opposing power against all external hindrances to freedom in accordance with universal laws (see MM 6:231) but not an opposing power in human nature against the internal inevitable tendency to misuse our predispositions (to animality and to humanity), which is what is required to account for a subjective willingness of human beings to give up their own unlimited freedom in a public lawful condition.

A direct consequence of the absence of such subjective willingness is this: even given the recognition of the theoretical necessity of a public lawful condition, the unsociable sociability in human nature, i.e. the selfish animal inclination, still misleads one to will the condition of peace and one's own unlimited freedom at the same time.\(^87\) Now since the former, i.e. the condition of peace, is essential for

\(^{87}\) My view can be contrasted with Shell’s view here. Shell rightly notes that “Kant breaks with an earlier liberal tradition, which tended to ground entrance into the civil compact (though not allegiance to it) on rational self-interest alone” (Shell 2009, 236). However, Shell goes on to identify the moral necessity of entering into civil condition in the necessity of the actualization of the concept of right (see Shell 2009, 236). My view expressed in the main text is that, in the absence of a subjective willingness to restrain our abuse of freedom, our selfish animal inclination would lead us to put others, but not ourselves, under the obligation to enter into a civil condition for the actualization of the concept of right. In other words, the necessity identified by Shell is still
self-preservation (by leaving the condition of war), the establishment of a public lawful condition is soluble even in the presence of the latter tendency to undermine it. This state of affairs is arguably what Kant tries to convey in his depiction of the establishment of a nation of devils in Toward Perpetual Peace:

The problem of establishing a state, no matter how hard it may sound, is soluble even for a nation of devils (if only they have understanding) and goes like this: “Given a multitude of rational beings all of whom need universal laws for their preservation but each of whom is inclined covertly to exempt himself from them, so to order this multitude and establish their constitution that, although in their private dispositions they strive against one another, these yet so check one another that in their public conduct the result is the same as if they had no such evil dispositions.” (PP 8:366)

In the absence of a subjective willingness to give up one’s unlimited freedom, what Kant depicts in the passage about the establishment of a nation of devils would be all we can expect from human beings and we would have to admit that human beings are devils (in the sense described in the passage) by nature. We

---

88 In Kant’s view, a people is essentially forced to submit to the constraint of public laws either by internal discord or external threat of war (see PP 8.365-366), which would threaten their self-preservation. With regard to international relations, Kant spells out the conditions for international peace in detail in PP. Rawls also discusses the conditions for internal peace in The Laws of Peoples. However, as I stated near the end of the introduction of this dissertation, since the subject matter here is the roles of the moral and the political in the philosophies of Kant and Rawls, I will not compare the substantial contents of their international peace theories in this dissertation.

89 For human beings to be evil by nature, reason would have to be evil by nature. In Kant’s view, however, such an assertion contains too much to provide a ground of moral evil in human being. Kant explicitly states that human beings are not, and cannot be, diabolical (see Rel 6:35).
would have to admit also that the public laws in the nation of devils cannot be seen as the act of a public will (TP 8:294) or a united will of the people, for the idea of a united will in a nation of devils is not thinkable due to the reciprocally conflicting nature of the particular wills of the devils. The idea of a public will also becomes nothing more than an empty thought in theory and the principle of right would be nothing more than a mere modus vivendi necessary only for the purpose of self-preservation.

While in Kant’s view this state of affairs is by itself sufficient for the establishment of a state, we should not disregard his equally important claim right before the passage on the nation of devils that a public lawful condition is “the most difficult one to establish and even more to maintain” (PP 8:366). In the nation of devils, the peaceful coexistence of the external freedom of all on the surface is constantly threatened by the underlying evil disposition of everyone in the nation. Since the primary concern of the devils is never the establishing or the maintaining of the state but only their self-preservation (an end to which the condition of peace serves merely as a means), it is understandably easier to maintain a public lawful condition when the threat (either from within or without the state) to their self-preservation is imminent. However, once they have gained considerable power in comparison to its neighboring states and potential internal discord is contained to a manageable scale (as is true in many developed countries nowadays), the primary cause for the devils to give up their own unlimited freedom in the public lawful condition also recedes in strength. This in turn unavoidably adds to the difficulty to maintain a public lawful condition in the state amid the heightening desire for freedom from the constraint of public laws.

Kant’s depiction of the nation of devils therefore introduces only the minimum requirement for the establishment of a state. In the absence of a
subjective willingness on the part of each individual to submit to the constraint of public laws, the sufficient condition for the more difficult task of maintaining a public lawful condition is nonetheless absent. As I argued near the beginning of this section, such subjective willingness is explicable only in the presence of an opposing power in human nature against the internal inevitable tendency to misuse our predispositions (to animality and to humanity). This opposing power can thus be nothing but a predisposition in human nature to the appropriate use of the above two predispositions. Or in other words, a predisposition to the lawful use of freedom (in opposition to the propensity to the lawless use of freedom) must be present in human nature in order to account for the ground over which the maintaining of the condition of peace is explicable. This requirement essentially leads us back to the predisposition to personality we left out in the beginning of the interpretive approach, for this predisposition “alone is rooted in reason practical of itself, i.e. in reason legislating unconditionally” in the faculty of desire and the exercise of the power of choice (see Rel 6:28). In essence, therefore, the opposing power against our unjust desire for unlimited freedom is our “susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (Rel 6:27) and this is the only intrinsic power in human nature we can count on in the pursuit of perpetual peace for the whole human species.

The maintaining of the condition of peace thus requires the predisposition to personality in human nature. In other words, Kant’s political philosophy requires his moral philosophy in the end, even if we interpret it starting with a merely political conception of the human being. The significance of this finding, however, is not thereby only an interpretive one. The whole argumentation in this interpretive approach also suggests that we should not merely attend to the easier task of establishing a superficial condition of peace but also to the more difficult
task of perpetual peace, for in the former we might be ignorant of what is required from within ourselves to maintain a sustainable and thoroughgoing condition of peace. Kant’s message here is unmistakably clear: if we fail to recognize and cultivate the subjective ground within ourselves for the establishment and maintaining of a public lawful condition, “a good constitution can come into being only by blind chance” (MM 6:372). It is only when the members of a people uphold the supremacy of the principle of right, i.e. the limiting condition of politics, from within themselves and for its own sake, can they entrust themselves with the duty to be “their own authors of their enduring welfare and at the same time that of others” (A809/B837). This ultimately requires the members of the people to look deeper than the public laws on the surface and to the intrinsic strength and motivation in the predisposition to personality they possess in themselves by nature as a human being.

8. Closing Remarks

This chapter began with a brief review of the conflict between Kant’s and Rawls’s approaches to political philosophy. While Kant might accuse Rawls of invoking an unvindicated moral conception of the person as a starting element in his theory of justice, Rawls maintains that a society united on Kant’s philosophy, which is an example of a reasonable comprehensive doctrine, requires the oppressive use of state power to remain so. I argued that although their conflict stems primarily from their different views on the role of the moral in the political, it is manifested in an apparently unresolvable conflict mainly because both of them attempt to have the moral (whether it be a full-fledged moral doctrine as in Kant or a moral conception of the person as in Rawls) serving both as a starting and a normative element in the political. This chapter therefore offers a third way
out of the deadlock by interpreting Kant’s practical philosophy starting from a merely political conception of the human being and see if this alternative interpretive approach will lead us to his moral philosophy in the end.90

In his Religion, Kant introduces the original predisposition to good in human nature, and his account turns out to be a valuable resource for the formulation of the starting political conception of the human being required for the interpretive approach. Among the three predispositions introduced by Kant, only the predisposition to personality is morality in its very essence. The interpretive approach therefore began with the other two predispositions, leaving aside the moral at the beginning of the interpretation. The political conception of the human being characterized by the two predispositions (and the inevitable tendency to misuse them) then leads us to the unsociable sociability of human beings and the tendency to the misuse of freedom. The tendency to the misuse of freedom leaves us in a state of constant antagonism, in which the lawless freedom of each unavoidably collides with the same freedom of others. The state of constant antagonism necessitates the transition to the condition of peace from the condition of war primarily for the purpose of self-preservation. From this arises the theoretical necessity of the idea of a public will, or general legislative will, in a public lawful condition, in which human beings must submit to the constraint of public laws in order to bring about the peaceful coexistence of their external freedom.

90 The interpretive approach from the political to the moral introduced in this chapter can be read alongside Palmquist’s “The Kingdom of God is at Hand!” (Palmquist 1994), in which he rightly points out that: “The ultimate end… will come about when there is no longer any distinction between empirical manifestations of religious and political systems and the pure moral reason to which they conform…” (434). The interpretive approach in this chapter leads to the same conclusion by showing that the political must conform to the moral in the end in order to be sustainable.
I went on to argue that although the *theoretical necessity* of a public lawful condition might be sufficient for its *establishment*, it cannot serve as the sufficient condition for the *maintaining* of a sustainable condition of peace, for it falls short of accounting for the *practical necessity* for every human being to give up their unlimited freedom in a public lawful condition. In order to *maintain* a condition of peace, I argued that we need an *opposing power in human nature* against our tendency to misuse our freedom. This leads to the *predisposition to personality* in the end, for this is exactly what is needed in our predisposition to account for the *practical necessity* of submitting ourselves to the constraint of public laws. This completes the alternative interpretive approach, for the political conception of the human being in Kant finally leads to his moral philosophy in the end.

I believe the alternative interpretive approach introduced in this chapter will allow those who are more accustomed to the liberal way of thinking to re-evaluate, or re-appreciate, the value of Kant’s philosophy to their project. As is evident from the discussion in this chapter, a society united on Kant’s philosophy by no means requires the oppressive use of state power to remain so. In fact, quite to the contrary of what Rawls suggests in his verdict against Kant, Kant’s philosophy actually reveals to us how a society can remain united not by the oppressive use of state power but by a power intrinsic to human nature, that is, our susceptibility to respect for the moral law as human beings. Is this not what is needed in any political philosophy?
Conclusion

The four main chapters dealt with four distinctive issues with regard to the roles of the moral and the political in the philosophies of Kant and Rawls. They led to different findings that were already discussed in each chapter’s respective closing remark. Before I proceed to the conclusion that can be drawn by consolidating their findings, let me summarize the conclusions reached in the four chapters here:

1. The moral and the political are inseparable in Kant: The first chapter introduced a new interpretive approach to Kant’s practical philosophy with a view to offering a holistic and comprehensive account of Kant’s moral and political philosophy. The focus of the chapter is the notorious problem of the unity of Kant’s account of the moral law in Groundwork and his principle of right in Doctrine of Right. I argued for their inseparability by focusing on the concept of the limiting condition of the freedom of action. While the moral law is the limiting condition of the freedom of action in general, the principle of right is nothing but the limiting condition of the freedom of action in our external relations. In the interpretive account given in the chapter, I showed that the two necessarily imply each other. The moral and the political are therefore inseparable in Kant. Before the end of the chapter, I also argued for the distinction between a positive agreement and a negative agreement with humanity as an end in itself. While the latter corresponds to the negative duty of right, the former gives rise to a positive duty in politics that is
moral by nature. According to the FH, this positive duty requires that “everyone also tries, as far as he can, to further the ends of others” (Gr 4:430). This finding further enriches the connection between the moral and the political in Kant.

2. **The moral in Rawls is not defensible:** In the second chapter, I argued that despite Rawls’s method of avoidance, Rawls’s understanding of morality arguably serves a foundational role in his political theory of justice and thereby also amounts to a philosophical stance in its own right. In particular, his moral conception of the person and his procedural interpretation of Kant’s conception of autonomy together constitute a substantial account of morality that can be contrasted with Kant’s original account. In the subsequent comparison of the two accounts of morality, I argued that Rawls’s account is not defensible because he is not aware of the importance of the self-lawgiving capability of practical reason in Kant’s moral philosophy. In the absence of this capability of practical reason, Rawls’s practical reason can only be conditioned. As a result, he cannot possibly vindicate the practical necessity of his theory, because the condition itself is unavoidably controversial.

3. **The roles of the political in Rawls are questionable:** In the third chapter, I evaluated from a Kantian perspective the four roles of political philosophy proposed by Rawls. They are highly relevant to this research because they reflect in the most direct manner Rawls’s commitment towards a complete detachment of the political from the moral. In this chapter, the four roles were scrutinized and contrasted with Kant’s views on the respective subject matters. In the discussion of each of the four roles, I explained why Kant could not possibly agree with Rawls and
proposed an alternative to Rawls’s idea. It became clear by the end of the chapter that, in the absence of an accurate account of *morality*, even Rawls’s account of the roles of political philosophy would be indefensible.

4. **The political necessarily leads to the moral in Kant:** On top of the interpretive approach to Kant’s practical philosophy introduced in the first chapter, I introduced yet another interpretive approach to Kant in the opposite direction in the last chapter: an interpretation starting with a *political* conception of the human being in Kant and ending with his moral philosophy. The political conception consists of the predisposition to animality and the predisposition to humanity found in Kant’s *Religion*. From these two predispositions, I argued that the possibility of their misuse, which according to Kant is an inseparable possibility attached to them, gives rise to the need for the *establishment* of a public lawful condition. However, although the political conception of the human being alone is sufficient for its *establishment*, it cannot at the same time serve as the sufficient condition for the *maintaining* of a public lawful condition. The maintaining of a sustainable condition of peace among human beings requires another predisposition in human nature in opposition to the intrinsic tendency to the misuse of freedom. This requirement finally leads us back to Kant’s moral philosophy and the predisposition to *personality*. The advantage of this new interpretive approach is twofold. First, it effectively avoids Rawls’s charge that Kant’s political philosophy is *derived from* a single philosophical doctrine that might not be endorsed by everyone in a pluralistic society, for the new interpretive approach *starts with* a political conception of the
human being. Second, it does not thereby betray Kant’s original plan of having his moral philosophy serve as the ground of his political philosophy, for the new interpretive approach ends with his moral philosophy.

The findings of this dissertation therefore can be summarized by the following statements. (1) *The moral* and *the political* are inseparable in Kant’s philosophy. (2) *The moral* in Rawls is not defensible. (3) The roles of the *political* in Rawls are questionable. (4) *The political* necessarily leads to *the moral* in Kant’s philosophy. The overall implications of these findings can be further interpreted as follows. From (1) and (4), we learn that at least with regard to Kant’s philosophy, Rawls’s verdict that political philosophy should be done independently of moral philosophy is not correct. *The moral* and *the political* in Kant together constitute a practical philosophy that does not fall victim to Rawls’s charge against comprehensive doctrines. From (2) and (3), we learn of the various difficulties and limitations faced by Rawls both in *the moral* and in *the political* in the absence of a philosophy of *morality*. The combination of these two findings strongly suggests that, if Kant’s moral philosophy is a correct and accurate account of morality, his political philosophy might be the only viable political philosophy for human beings.
Bibliography


Reich, Klaus. *Kant and Rousseau*. Tübingen: Mohr-Siebeck, 1936.


CURRICULUM VITAE

Academic qualifications of the thesis author, Mr. Wong Sai Ming:

-Received the degree of Bachelor of Engineering (First Class Honours) from The Chinese University of Hong Kong, December 1997.

-Received the Master degree from New Asia Institute of Advanced Chinese Studies, July 2014.

December 2018