UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

VINICIUS PINTO DE CARVALHO

Kant's derivation of the Formula of Universal Law in the *Groundwork*: a defense

A derivação kantiana da Fórmula da Lei Universal na *Fundamentação*: uma defesa

CAMPINAS
2021

VINICIUS PINTO DE CARVALHO

Kant's derivation of the Formula of Universal Law in the *Groundwork*: a defense

A derivação kantiana da Fórmula da Lei Universal na *Fundamentação*: uma defesa

Dissertation presented to the Institute of Philosophy and Human Sciences of the University of Campinas in partial fulfillment of the requirements for the Degree of Master, in the area of Philosophy.

Dissertação apresentada ao Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre, na área de Filosofia.

Supervisor/Orientador: Profa. Dra. Monique Hulshof

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO VINICIUS PINTO DE CARVALHO, E ORIENTADA PELA PROFA. DRA. MONIQUE HULSHOF.

CAMPINAS

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Filosofia e Ciências Humanas
Cecília Maria Jorge Nicolau - CRB 8/3387

C253k

Carvalho, Vinicius, 1996-
Kant's derivation of the Formula of Universal Law in the *Groundwork* : a defense / Vinicius Pinto de Carvalho. – Campinas, SP : [s.n.], 2021.

Orientador: Monique Hulshof.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas.

1. Kant, Immanuel, 1724-1804. 2. Princípios morais. 3. Ética. I. Hulshof, Monique, 1982-. II. Universidade Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: A derivação kantiana da Fórmula da Lei Universal na *Fundamentação* : uma defesa
Palavras-chave em inglês:
Moral principles
Ethics
Área de concentração: Filosofia
Titulação: Mestre em Filosofia
Banca examinadora:
Monique Hulshof [Orientador]
Joel Thiago Klein
Robinson dos Santos
Data de defesa: 26-08-2021
Programa de Pós-Graduação: Filosofia

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-3033-8104
- Currículo Lattes do autor: http://lattes.cnpq.br/8814565157435989

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

A Comissão Julgadora dos trabalhos de Defesa de Dissertação de Mestrado, composta pelos Professores Doutores a seguir descritos, em sessão pública realizada em 26/08/2021, considerou o candidato Vinicius Pinto de Carvalho aprovado.

Profa. Dra. Monique Hulshof
Prof. Dr. Joel Thiago Klein
Prof. Dr. Robinson dos Santos

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertações/Teses e na Coordenadoria do Programa de Pós-Graduação em Filosofia do Instituto de Filosofia e Ciências Humanas.

To my parents, for their love

Acknowledgements

This work was written during a pandemic. I do not mean this as an excuse for any mistakes and shortcomings; rather, I wish to give some dimension to how much grateful I am to the people who have helped me write a dissertation in Philosophy, during what was perhaps the most challenging time of their lives.

I could not start without thanking my parents, Alda and João, and my brother Guilherme. They are the kindest and most supportive people I know, and I am incredibly lucky to have them as family.

I especially want to thank my supervisor, Monique Hulshof, who has helped and encouraged me ever since we met. I am most grateful for the attention she has given me in these exceptional last two years.

I am grateful for the comments and criticisms of Robinson dos Santos and Joel Thiago Klein. Their suggestions made this work much better.

Jesse Zampieri Peternella, Pedro Gallina Ferreira, and especially Nicole Martinazzo have helped me from the very beginning of this project. I am thankful for their attention and support.

For the close friendship, I want to thank Thainá Coltro Demartini and Filipe Martone de Faria.

Finally, I want to thank the São Paulo Research Foundation – FAPESP (grant 2019/21992-8) for funding this research. Moreover, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

"[...] *the first attempt at a free choice, which, as the first one, probably did not turn out in conformity to expectation. Now the harm might have been as you like, yet about this it opened the human being's eyes (Genesis 3: 7). He discovered in himself a faculty of choosing for himself a way of living and not being bound to a single one, as other animals are. Yet upon the momentary delight that this marked superiority might have awakened in him, anxiety and fright must have followed right away, concerning how he, who still did not know the hidden properties and remote effects of any thing, should deal with this newly discovered faculty. He stood, as it were, on the brink of an abyss;"*

Immanuel Kant, *Conjectural beginning of human history*

ABSTRACT: One of Kant's aims in the *Groundwork of the Metaphysics of Morals* is to identify the supreme principle of morality. This research analyzes the arguments from the first and second sections, which identify the so-called Formula of Universal Law (FUL) – "act only according to that maxim through which you can at the same time will that it become a universal law" – as one formulation of such principle. According to many interpreters, these arguments contain a "gap", for Kant would have concluded the FUL after considering it equivalent to the principle that one should always conform one's maxims to universal law. Nonetheless, critics argue that these principles differ significantly because they confer different moral statuses to the same maxim. Thus, Kant would not have adequately justified the FUL. This dissertation reconstructs the derivation of the FUL in the first section (in GMS 402) and the second section (in GMS 420-1) of the *Groundwork*. I argue that Kant's concept of autonomy (or self-legislation), introduced (even if tacitly) in the first section and completely spelled out in the second, warrants his move from the idea of conformity to universal law to the FUL, such that there is no gap in the derivations of this formula.

KEY-WORDS: Kant, Immanuel; Moral Principles; Ethics

RESUMO: Um dos objetivos de Kant na *Fundamentação da Metafísica dos Costumes* é identificar o princípio supremo da moralidade. Esta pesquisa analisa os argumentos da primeira e segunda seções que identificam a chamada Fórmula da Lei Universal (FLU) – "age apenas segundo a máxima pela qual possas ao mesmo tempo querer que ela se torne uma lei universal" – como uma das formulações de tal princípio. Segundo muitos intérpretes, esses argumentos contêm uma "lacuna", pois Kant teria concluído a FLU após considerá-la equivalente ao princípio de que sempre se deve conformar as máximas a leis universais. No entanto, os críticos argumentam que esses princípios diferem significativamente, uma vez que conferem status morais diferentes a uma mesma máxima. Assim, Kant não teria justificado adequadamente a FLU. Esta dissertação reconstrói a derivação da FLU da primeira seção (em GMS 402) e da segunda seção (em GMS 420-1) da *Fundamentação*. Argumento que o conceito de autonomia (ou autolegislação) de Kant, introduzido (mesmo que tacitamente) na primeira seção e completamente explicitado na segunda, justifica a sua passagem da ideia de conformidade a leis universais para a FLU, tal que não há lacuna nas derivações desta fórmula.

PALAVRAS-CHAVE: Kant, Immanuel; Princípios Morais; Ética

Note on translations

Except for the *Critique of Pure Reason*, referenced with the A and B pagination standard, all references to Kant's works follow the abbreviation and pagination scheme of *Kant's gesammelte Schriften: herausgegeben von der Deutschen (formerly Königlichen Preuißischen) Akademie der Wissenschaften, Berlin: Walter de Gruyter, 1902-*. Translations used were:

Anth	*Anthropology from a pragmatic point of view*. Translated by Robert B. Louden in *Anthropology, History and Education (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Günter Zöller and Robert B. Louden. Cambridge: Cambridge University Press, 2007.
GMS	*Groundwork of the Metaphysics of Morals.* A German-English Edition. Edited and Translated by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press, 2011.
KpV	*Critique of practical reason*. Translated by Mary Gregor in *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.
KrV	*Critique of pure reason*. Translated and edited by Paul Guyer and Allen Wood. *The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press, 1998.
MS	*Metaphysics of morals.* Translated by Mary Gregor in *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.
V-Mo/Mron	*Moral Mrongrovius*. Morality according to Prof. Kant: Mrongovius's second set of lecture notes (selections), translated by Peter Heath in *Lectures on Ethics (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Peter Heath and J. B. Schneewind. Cambridge: Cambridge University Press, 1997.
R	*Religion within the boundaries of mere reason*. Translated by George Di Giovani in *Religion and Rational Theology (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Allen Wood and George Di Giovani. Cambridge: Cambridge University Press, 1996.
V-MS/ Vigil	*Moral Vigilantius*. Kant on the metaphysics of morals: Vigilantius's lecture notes, translated by Peter Heath in *Lectures on Ethics (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Peter Heath and J. B. Schneewind. Cambridge: Cambridge University Press, 1997.

Table of Contents

Introduction

In the *Lectures on Ethics*, Kant is reported to have said that "[t]here is no Science so filled with tautologies as ethics" (V-Mo/Collins, AA 27: 266). By a 'tautology' in this context he means moral principles incapable of providing moral guidance, that is, principles the application of which cannot show the agent which course of action is morally permissible, required, worthy and so on. Kant says that this deficiency of moral principles failing to be properly action-guiding can be found as far back as in Aristotle, but his main target in the *Lectures* is Baumgarten. Kant was required, after all, to use the *Initia philosophiae practicae prima acromatice* as the textbook on his courses on ethics, and in that work, Baumgarten (2020, p. 53-6) presents four moral principles:

1. *Fac bonum et omitte mallum* (do the good and abstain from evil);
2. *Quaere perfectionem, quantum potes* (seek perfection as much as you can);
3. *Vive convenienter naturae* (live according to nature, as much as you can);
4. *Ama optimum quantum potes* (love the best as much as you can)

Although Kant had to teach using Baumgarten's book, he did not hesitate to criticize it. He points to Baumgarten's lack of clarity in properly distinguishing hypothetical from categorical imperatives, and he says that Baumgarten failed to consider the distinctive kind of volition of a will that is not affected by sensible affections, which has serious implication to his understanding of what is *obligation* (*Verbindigkeit*) and *necessitation* (*Nöthigung*). Kant's most important point of criticism, nevertheless, concerns these four principles listed above. Bluntly, he says that "none of these statements are principles of morality" (V-Mo/Collins, AA 27: 266). But why not?

The reason for such a fierce condemnation is that Kant thinks that all of these principles cannot properly guide action. *Fac bonum et omitte mallum* is his favorite target because it fails in two aspects. First, it is a *vague principle*. We predicate 'good' out of many actions, in the sense that "the good can be good in a variety of ways for any given purpose" (V-Mo/Collins, AA 27: 264). In other words, actions can be said *good* in many different respects: we might say that an action is good meaning that it is *instrumentally* or *technically* good (as in "taking a plane is a good way to get to Paris, better than by boat"), or that it is good in a prudential way (as in "it would be good to have all my taxes paid"). What this means is that not

every action that is good is *morally* good. Since this principle cannot show us how to distinguish these different kinds of goodness of actions, it is inadequate for the purpose of moral guidance. Second, and more importantly, this principle is a *tautological principle*, and Kant says that "a tautological rule is one which, when called upon to decide a question, gives an empty answer" (V-Mo/Collins, AA 27: 264). Appropriately, this principle commands that we do what is good, which is trivial: because it is a **moral** principle, it will of course command an action deemed good. After all, it would be absurd if a moral principle positively commanded any action that was not morally good.[1] The problem is that this principle commands that we "do the good", but gives us no way to distinguish *what* is good. And if we cannot distinguish which actions are good, then it is a useless principle. As Kant says, "it tells us nothing of what is good, saying merely that I should do what I should do" (V-Mo/Collins, AA 27: 265). To illustrate, we can imagine another case of a trivial and tautological principle. Suppose you ask someone "how can I be a good person?", to which one replies: "It is not hard to be a good person. *All you have to do are good deeds*!". This principle is as tautological as Baumgarten's. We want moral principles that let us judge, to a certain extent, *which* actions are morally good. That is why a moral principle must have at least some action-guiding capacity.[2]

Although Kant was aware of the problems with developing a theory centered around a tautological principle, his own theory was later criticized for having the same problem he saw in Baumgarten's. Many commentators claim that Kant's arguments for the Formula of Universal Law (henceforth FUL), both in *Groundwork* I and II, contain a serious argumentative flaw.[3] In fact, they argue that the moral principle Kant derives from his arguments is not the FUL, but rather a tautological principle.

---

[1] Kant sees a close relation between moral prescriptions and the concept of *good* (GMS, AA 04: 413). This will be discussed in chapter two.

[2] Baumgarten's second principle, although not a "complete tautology", is also rejected by Kant. Kant argues that perfection is something different from moral goodness: perfection "is the completeness of the man in regard to his powers, capacity and readiness to carry out all the ends he may have" (V-Mo/Collins, AA 27: 265). But moral goodness is about the correct use of these powers and capacities. In a tone similar to that of the beginning of GMS I, Kant says that moral goodness consists of the "perfection of the will, not the capacities" (V-Mo/Collins, AA 27: 266). Meaning that a being's perfection is conditioned to the goodness of his will, the latter being the only unconditioned thing. Moreover, Baumgarten's third principle is rejected because it is not a moral, but a prudential principle. The natural laws to which we should live in accordance to concern the satisfaction of sensible desires and aim, ultimately, to happiness. And it is still a tautological principle of prudence, because it does not allow us to judge which satisfied desires would lead to happiness. At last, the fourth principle is quickly dismissed with Kant saying that "there are two ways of loving anything: from inclination and from principle. Thus a rascal loves the good on principle, but the bad from inclination" (V-Mo/Collins, AA 27: 266). Accordingly, the principle fails to distinguish this difference. Kant also talks about Baumgarten's principles in V-MS/Vigil, AA 27: 517-8. For more on his account of Baumgarten's moral philosophy in the *Lectures on Ethics*, see Bacin (2015).

[3] See Aune (1979), Wood (1990), Schönecker and Wood (2002), Allison (1991) and Hill (2000).

Now, in *Groundwork* I Kant analyzes the concept of duty, and after a series of distinctions and arguments he arrives at the conclusion that duties are actions commanded by a law. This intermediate conclusion then leads to the cryptic paragraph where he derives the FUL:

> But what kind of law can it possibly be, the representation of which – even without regard for the effect expected from it – must determine the will for it to be called good absolutely and without limitation? Since I have robbed the will of all impulses that could arise for it from following some particular law, nothing remains but as such the universal conformity of actions with law, which alone is to serve the will as its principle, i.e. I ought never to proceed except in a way that I could also will that my maxim should become a universal law. Here, then, mere conformity with law as such (not founded on any law determined with a view to certain actions) is what serves the will as its principle, and must so serve it if duty is not to be as such an empty delusion and a chimerical concept; common human reason in its practical judging is actually in perfect agreement with this, and always has the envisaged principle before its eyes. (GMS, AA 04: 402, orig. emphasis)

Setting aside for now a detailed exegesis of this paragraph (which will be the task of chapter one), its core argument is the following. Since his preceding analysis has shown that practical laws with a material content (e.g. an end to be attained, an inclination to be satisfied) cannot yield morally worthy actions, Kant concludes that actions with moral worth are determined by the pure form of a law, because a law's own form is the only thing that "remains" once he has "robbed the will of all impulses that could arise for it from following some particular law". Consequently, Kant affirms that it is "the universal conformity of actions with law, [which] alone is to serve the will as its principles". Bruce Aune, who first questioned the legitimacy of the derivation of the FUL, sees Kant arguing for a principle of a principle of the following sort:

(Principle of Conformity - PC[4]): Act on maxims that conform to universal law

This principle is supposed to capture the idea that "mere conformity with law as such (not founded on any law determined with a view to certain actions) is what serves the will as its principle". According to Aune, however, Kant's mistake was to reformulate PC into the principle below:

(**Formula of Universal Law - FUL**): I ought never to proceed except in a way that I could also will that my maxim should become a universal law

With the use of an "i.e." (*d. i, das ist*) in that passage quoted above, Kant seems to consider these two principles – PC and the FUL – to be equivalent, introducing the latter without

[4] I am calling "Principle of Conformity" what Aune (1979, p. 29) calls principle "L", and what Wood (1990, p. 78) calls "the CI [Categorical Imperative]".

further explanation. But Aune argues that these two principles cannot be equivalent because they differ in what he calls *practical import*: the precept to "act in conformity to universal laws" (PC) does not allow one to distinguish what laws are there to be followed.[5] This is a tautological principle just like Baumgarten's, for when applying it one cannot decide which actions are *licit* [*erlaubt*] and which are *ilicit* [*unerlaubt*]. It is uninformative, in the sense that it says that one should conform one's maxims to universal laws, but it gives one no means to identify which laws are there. How is one supposed to know whether there is a law that commands, for instance, "help those who need assistance, to the best of your ability"? For all that matters, "*do not* help those who need assistance" could be a law, and with what Kant has provided so far, there is no way of knowing it. In short, from the realization that duties are commanded by laws with no material content, all this principle says is that one's actions should be in conformity to these laws, *without revealing which laws there are*. As Hill (2000, p. 39) observes: "[…] the problem is this. Even though 'conform to universal law' is a commitment of conscientious agents, it is only a very minimal requirement, telling one very little about how to go about in moral deliberation."

By contrast, the FUL can guide moral deliberation in a much more significant way. Indeed, right after first exposing it, Kant applies it to show the immorality of making a false promise. And further on he gives great emphasis to the FUL's practical import. What is more, the FUL is not only supposed to enable us to identify which courses of actions are morally permissible and which are impermissible, but it is also supposed to assess an action's goodness: actions are good when done for the sake of this principle (i.e., from duty). Kant says:

> I do not, therefore, need any wide-ranging acuteness to see what I have to do for my willing to be morally good [*damit mein Wollen sittlich gut sey*]. Inexperienced with regard to the course of the world, incapable of bracing myself for whatever might come to pass in it, I just ask myself: can you also will that your maxim become a universal law? If not, then it must be rejected [*frage ich mich nur: Kannst du auch wollen, daß deine Maxime ein allgemeines Gesetz werde? Wo nicht, so ist sie verwerflich*] […] (GMS, AA 04: 403)

Consequently, for Aune the derivation has the following problem (which he labelled the *derivation gap problem*): on the one hand, Kant has a principle (PC) that follows from his analysis of the concept of duty; however, it is a tautological, uninformative principle. On the other hand, he has a principle that is properly action-guiding (the FUL), but whose

[5] See Aune (1979, p. 34) and Kerstein (2002, p. 7). To be clear, the "derivation gap problem" should not be taken as one version of the famous "empty formalism" charge made by Hegel. Hegel claimed that the FUL could not identify impermissible maxims unless the agent held substantive values. On the other hand, those who attribute a gap to Kant's derivation usually accept that the FUL is a proper action-guiding principle, yet deny that Kant validly inferred it from his arguments.

connection to PC and the preceding analysis of the concept of duty is unclear, at best, because Kant connects the two principles with a dubious "i.e.". Ultimately, if one accepts Aune's interpretations, the impression that Kant's analysis results in a moral principle that is just as tautological as Baumgarten's *fac bonum omitte mallum*. Facing this shortcoming, it seems that Kant surreptitiously comes up with another principle, one properly action-guiding, and presents it as a reformulation of the first one with the use of a cryptic "i.e.". As it appears, there is a *gap* between the practically innocuous PC and the action-guiding FUL, such that the move from one to the other is unjustified.

One way out of this problem would be to concede to Aune that the derivation in *Groundwork* I is in fact flawed, and try to find a sound argument in the derivation in *Groundwork* II. After all, since in the first section Kant is taking common moral cognition as a starting point and operating a *transition* to a properly philosophical moral cognition, it could be argued that the derivation in the second section is much more refined than the one in the first, because now Kant has already given more precise definitions to many concepts, such as the definition of *Wille* (GMS, AA 04: 412), the distinction between perfect and imperfect wills (GMS, AA 04: 412-3), and a more detailed characterization on the nature of the moral command with the distinction between hypothetical and categorical imperatives (GMS, AA 04: 414).

Notwithstanding, this strategy would be unsuccessful because Kant's derivation of the FUL in *Groundwork* II is also accused of containing the same argumentative gap between a properly justified yet tautological principle and the FUL. Briefly, since the relevant paragraph will be analyzed in the second chapter of this dissertation, the point is that in the derivation of the second section Kant arrives at the conclusion that in an action based on a categorical imperative "nothing is left but the universality of a law as such". From this, he immediately jumps to the FUL without much explanation. Aune sees here the same problem he raised concerning the first derivation: it looks as if Kant's analysis allows him to justify a moral principle that is in fact tautological, because it commands action in conformity to universal laws without making it possible for one to know *what laws there are*. With its application one cannot know, for example, whether a principle to "always prioritize my self-benefit" is a law or not, or if there is a law of beneficence. To make it informative and action-guiding, Kant reformulates it in terms of the FUL. The issue again is that this reformulation seems like an *ad-hoc* solution to PC's lack of practical import.

Now, with respect to Aune's analysis of the derivations, I believe that there is not a gap, and consequently not a fallacious equivalence, between the uninformative and tautological Principle of Conformity (PC) to the action-guiding Formula of Universal Law (FUL) simply

because Kant *does not* embrace PC. As Allison (2011, p. 138) correctly pointed out, when Kant talks about the principle of "conformity to universal law" he uses the indicative mood, whereas only when he presents the FUL he uses the imperative. On the one hand, this indicates that one should understood the former principle *descriptively*, in the sense that Kant is describing the nature of a will that is good: a good will is such that it conforms its maxims to universal law. On the other hand, only the FUL should be understood *prescriptively*, in the sense that it is the principle that one should adopt in order for one's will to be good, such that an agent conforms her maxims to universal law (thus having a *good* will) *when* she follows the FUL. The problem with Aune's analysis, thus, is to see a supposed action-guiding principle where there is only a descriptive one. It is to read "conformity to universal law as such" as "*Conform* [imperatival] to universal law as such".

What is more, we see the same thing in the derivation in the second section. When Kant says that when one acts on a categorical imperative "nothing is left but the universality of a law as such" he is making a conceptual claim about the very concept of a categorical imperative. In other words, he is saying that in actions based on categorical imperatives, there is no other motive left (in the sense of a *Bewegungsgrund*) but the universality of the law as such. Only then does he introduce the FUL, which, because it is in the imperative mood, should be understood as the prescriptive principle that agent's ought to employ. Just like in the first section, what we have is a move between a description of action on a categorical imperative to a principle the application of which (i.e., a formula) yields action performed on a categorical imperative; not a move between two different prescriptive principles like Aune's reading suggests.

Though the problem with the derivations of the FUL is not that Kant moves unwarrantedly from a vacuous to an action-guiding moral principle, this does not mean that there are no issues with it. Commentators are right to point out that Kant's arguments are extremely cryptic, and that the way he moves from a description of the nature of a good will (in the first section with an obscure "i.e."), and from a claim about the concept of action based on categorical imperative (in the second), to the FUL is abstruse and in very much need of explanation. To put it plainly, the problem is: why is it that the FUL is a test for the conformity of maxims to universal laws? In other words, why is it that one must be able *to will* that a maxim become a universal law to know that it conforms to universal law? Why not only be able to *conceive* it as universal law (Rickless, 2004, p. 571)? What makes Kant's derivations difficult to accept at face value is that, as Allen Wood says, "No reason has been suggested why the test of conformity to universal law should be what *I* can or do will" (1999, p. 48), and that it is just

not obvious, as Kant appears to think, that we "can equate 'maxims which accord with universal laws' with 'maxims which can themselves be willed as universal laws" (1999, p. 106-7).[6]

In light of this, the aim of this dissertation is to reconstruct Kant's arguments that lead to the derivation of the FUL, and to put forward an interpretation which vindicates his move from the idea to conformity to universal law to this formula. The first chapter will cover Kant's derivation in *Groundwork* I, and two interpretative theses will be argued for. First, I am going to defend that (i) even though Kant fully introduces the concept of autonomy only later in *Groundwork* II, he is already arguing in the first section with the assumption that moral laws are self-legislated laws. With recourse to Kant's analysis of the concept of duty, I will show that it is the concept of respect for law that provides textual support for this thesis. Accordingly, with the second thesis I will argue that (ii) when Kant says that "mere conformity to universal law as such […] is what serves the will as its principle" he means that agent's maxims must conform to universal law. This comes to saying that maxims must have the *form* of a law, which, in turn, is only possible through self-legislation. Combined with the first thesis and with a clarification of the concept of a *formula*, this will show how Kant can cash out "the mere conformity to universal law" principle in terms of the FUL.

In the second chapter I will tackle the derivation in *Groundwork* II. My strategy consists in showing that Kant continues to argue with the assumption that moral laws are self-legislated laws, and we will find evidence for this in his account of categorical imperatives. Kant insists that only categorical imperatives can express moral laws because only this kind of imperative expresses a universally and unconditionally valid command. With recourse to his distinction between heteronomy and autonomy and his discussion of previous moral theories, I will show that categorical imperatives can only have this kind of binding force if the laws they express are self-legislated. Consequently, this will explain how Kant can turn the idea that "nothing is left but the universality of a law as such, with which the maxim of action ought to conform" (GMS, AA 04: 421) into a moral principle, the application of which allow agents to check whether their practical principle (their maxim) conforms to universal law – that is, the FUL.

---

[6] See also Kerstein (2002, p. 87): "Does Kant not here take an illicit step from the notion that, by virtue of its very concept, a categorical imperative commands conformity to law to the further notion that it commands that you act only on maxims that you can at the same time will to become universal laws?"; and Engstrom (2009, p. 5-6, orig. emphasis): "Kant appears to slide *from* the merely formal and undisputed principle that a rational being should conform its will to any universal law it can recognize as valid for it as such […] *to* the substantive and questionable principle that such a being should act only on maxims it can will as laws holding universally, for all rational agents. […] In short, Kant's argument trades on an ambiguity in its talk of a maxim's conformity to 'the universality of a law in general', confusing willing in accordance with a law with willing a law."

Chapter 1: The Derivation of the FUL in *Groundwork* I

1.1 - Introduction

In the preface to the *Groundwork*, Kant argues that our moral experience evinces that moral judgments are grounded on an *a priori* principle. He says that when we think of morality, we inevitably think about a "common idea of duty and of moral laws" (GMS, AA 04: 389). This means that whenever we reflect on ordinary moral prescriptions, such as "do not kill" or "do not lie", we have the conscience that these are absolute obligations, that is, that they do not allow for exceptions. We also have the conscience that these are universal obligations, in the sense that every person is bound by it. Consequently, this indicates that our obligations and duties must necessarily be grounded on a principle discoverable *a priori*, for we know that "[e]xperience never gives its judgment true or strict but only assumed and comparative universality (through induction)" (KrV, B4). Therefore, whoever wants to discover the exact nature of the ground of our moral obligations must search for this principle not "in the nature of the human being, or in the circumstances of the world in which he is placed, but *a priori* solely in concepts of pure reason" (GMS, AA 04: 389).

Besides this philosophical need to discover the *a priori* ground of moral obligations, there are also more concrete reasons that motivate this search. On the one hand, since we have the notion that our duties are absolute and universal, there is a great danger in grounding them on experience. We face the risk of *mistakenly* taking some action as obligatory – for instance, because it is something that everyone does, or because it is something that a powerful authority commands us to do – and so concluding that everyone must act in such a way. This leaves open the possibility of general moral corruption, since there is not a secure and independent principle to which one may recur to establish what is and what is not duty. Experience gives a kind of conformity to morality that is only very "contingent and precarious", and this again justifies why "the moral law […] is to be sought nowhere else than in a pure philosophy" (GMS, AA 04: 390). On the other hand, we need a robust and clear rational justification for our moral obligations because, for Kant, we are naturally inclined to violating them. Since "reason issues its prescriptions unrelentingly, yet without promising anything to the inclinations", but also because the claims of inclinations are so "vehement and yet seem so reasonable", for Kant we have

> A propensity to rationalize [*vernünfteln*] against those strict laws of duty, and to cast doubt on their validity, or at least their purity and strictness and, where possible, to

> make them better suited to our wishes and inclinations, i.e. fundamentally to corrupt them and deprive them of their dignity […] (GMS, AA 04: 405).

Therefore, "the identification and corroboration of the *supreme principle of morality*", that is, the exact principle on which our obligations are grounded, is precisely Kant's goal in the *Groundwork*. To accomplish the first task, he proposes to proceed "analytically from common moral cognition to the determination of its supreme principle" (GMS, AA 04: 392), and this is where we enter the territory of the first section, entitled "*Transition from common to philosophical moral rational cognition*". Right at the outset, we have an important interpretative question. After all, what does Kant mean by saying that he has adopted a method that arrives analytically from common moral rational cognition to the supreme principle of morality? Does 'analytically' here refers to the analytic method Kant talks about in the first *Critique* and adopts in the *Prolegomena*? Or does Kant want to say that he is going to identify the supreme principle of morality by means of a series of conceptual *analyses*?

On the one hand, Paton (1947) is the commentator that most clearly defends that the analytic method of the *Groundwork* is like the one present in the *Prolegomena*. He reminds us that with the analytic method we start off with a set of cognitions and look for their conditions of possibility. In the case of the *Prolegomena*, Kant searches for the conditions of possibility of synthetic *a priori* judgments of mathematics and (pure) physics. On Paton's reading of the *Groundwork*, Kant begins with ordinary moral judgments – for instance, judgments that have predicates such as "good", "bad", "evil", "just", "licit", "worthy", and so on – which we believe to have an *a priori* element given the necessity and universality they contain, and look for that which the attribution of these properties to these judgments. In other words: for Paton, Kant is searching for the moral principle that validates moral judgments that contain the *a priori* elements of universality and necessity. On the other hand, commentators such as Allison (2011, p. 33-5) and Wood and Schönecker (2014, p. 19-20) doubt that Kant's talk of proceeding "analytically" refers to the analytic method of the *Prolegomena*. For them, by "analytically" all that Kant means is that he will proceed in the search for the supreme principle of morality by means of a series of conceptual analyses.

I believe that these interpretations are compatible. First, like Patton, I do see Kant searching for conditions of possibility in a method that is akin to that used in the *Prolegomena*. Yet while Paton thinks that Kant's starting point is the set of our ordinary moral judgments in general, I think it is more precise to say that he is actually looking for the conditions of possibility of just one judgment, namely the opening sentence of the *Groundwork*: "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be taken

to be good without limitation, except a GOOD WILL" (GMS, AA 04: 393, orig. capitalization). The crucial point to note is that Kant makes a regressive investigation in *Groundwork* I and II, thus employing the analytic method, *with the use of conceptual analyses*. As we shall see, in order to vindicate that first sentence Kant obviously needs to elucidate the notion of *good will*, which common moral cognition takes to be the only thing that is unconditionally valuable. And to do so, he undertakes an analysis (*Zergliederung*) of the concept of *duty,* which then advances to an analysis of the concept of *respect for law*, and so on, up to the point where he derives the supreme principle of morality, that is, the first formulation of the FUL. What is more, he says that only if the FUL is the supreme principle of morality (or its primary formula) does the concept of duty have any meaning. Accordingly, in one sense, it appears that the FUL is the condition of possibility of our having duties at all and, consequently, of having a good will.[7] Moreover, we see a similar pattern in *Groundwork* II. There Kant begins with an analysis of the concept of a rational being as such, which progresses to an analysis of the concept of a categorical imperative, given the supposition that (some) rational beings are bound by categorical imperatives. Then, after a series of arguments, Kant identifies the principle of the autonomy of the will, which, he claims, is the only principle that justifies imperatives having categorical form. Again, we see a path from what is conditioned to its conditions of possibility.[8]

That being said, in the first section of the *Groundwork* Kant identifies the supreme principle of morality with the following formula, known as the first variant of the Formula of Universal Law (FUL): "I ought never to proceed except in such a way *that I could also will that my maxim should become a universal law*" (GMS, AA 04: 402, orig. emphasis). In this chapter, I will advance a reading according to which Kant's derivation of the FUL is not a fallacy. I will be arguing contrary to the view discussed in the introduction, namely that there would a gap in Kant's arguments to the effect that the FUL would be invalidly inferred as an equivalent principle to the "mere conformity to universal law" requirement. My strategy will be the following: first, I will reconstruct Kant's account of the good will and its relation to the concept of duty. This is going to allow us to focus on one of Kant's examples of morally worthy action, which, as we shall see, is action done from duty. Second, the analysis of these cases of morally worthy actions will reveal three fundamental propositions about the concept of duty. A detailed explanation of these propositions will allow us to gain a better insight into many important

[7] To be clear, what I am claiming here is that the FUL must be the primary formula of the supreme principle of morality *in order for us, human beings,* to have a good will. As we shall see, for Kant there can be other kinds of beings who have a good will even though they are not subject to any duties, nor bound by any prescriptive principle like the FUL.

[8] Here I am following Allison (2011, p. 35-6).

concepts such as maxim, inclination and, especially, respect for law. And it is in the concept of respect for law that I will find textual evidence for my first interpretative thesis: that moral laws are self-legislated, and that Kant is already arguing with this assumption even though he will completely introduce the concept of autonomy only later in *Groundwork* II. Third, and with this first thesis established, I will bring attention to the famous paragraph of the derivation (GMS, AA 04: 402). The knowledge that the laws to which our maxims should conform are self-legislated laws, together with a clarification of the notion of a *formula*, will enable us to comprehend how Kant can cash out the "mere conformity to universal law" precept in terms of the FUL, without this move being unjustified.

1.2 - Duty, respect for law, and autonomy

As previously discussed, the first section of the *Groundwork* proceeds analytically from common moral cognition up to the identification of its supreme moral principle. Kant's strategy is to use our pre-philosophical conception of morality in order to reveal the moral principle that we already utilize, even if tacitly, to "distinguish what is good, what is evil, what conforms with duty or is contrary to it" (GMS, AA 04: 404). And he starts this task with the concept of a *good will*, which is the only thing that can "be taken to be good without limitation" and whose value outweighs any natural trait – such as wit, judgment, or "resolve and persistency of intent"– or any "gifts of fortune" – such as riches, power and even health and well-being – (GMS, AA 04: 393). Kant's point here is that whatever personal qualities, social abilities, or material goods a person might have, nothing guarantees that these will not be used in morally inacceptable ways *unless* this person is guided by principles of a good will. He observes that "the cold blood of a scoundrel makes him not only far more dangerous, but also immediately more loathsome in our eyes than he would have been taken to be without it" (GMS, AA 04: 394)".

As the only thing that is unconditionally good, the good will then serves as a limiting condition on the goodness of everything else, including happiness. The idea at this point is that an impartial spectator would disapprove of someone who is happy but shows no concern as to whether her actions are morally right or wrong. Having a good will, therefore, "appears to constitute the indispensable condition even of the worthiness to be happy." (GMS, AA 04: 393). And this marks an important difference between the value of happiness and that of a good will. Although both are, for Kant, good in themselves – meaning that we want a good will and happiness not because of some further end or object they could bring about, but because

they are good things for having *per se* –, only the good will is good in every occasion and in every circumstance, whereas there are cases in which happiness is undeserved or morally inappropriate.[9]

Thus, if a good will has absolute, unconditional worth, it means that its goodness is not determined as a function of what it brings about as consequences of its action, or the appropriateness of its intentions. A good will is "good just by its willing, i.e. in itself" (GMS, AA 04, 394). In order to understand why a good will is good just by its very acts of volition, we must first turn our attention to the concept of *duty*, because in beings whose will is subject to "certain subjective limitations and hindrances" (GMS, AA 04: 397), a good will manifests itself through dutiful actions. Accordingly, Kant gives four examples of such actions, but he distinguishes three possible sources of motivation. The agent may perform a dutiful action while being motivated by the recognition that the action is morally right, and in such cases the action is done *from duty* (*aus Pflicht*). The agent may comply with her duty because she enjoys performing the dutiful action, and so it is done in *conformity with duty* (*pflichtgemäßig*) but from immediate inclination. And finally, the agent may do what is morally right because doing so furthers some of her interests, and thus we have an action that is in conformity with duty, yet performed from mediate inclination.

Given these three ways to comply with duty, Kant is interested in knowing which kind of motivation is that of a person with a good will. He quickly dismisses the third and last kind, because doing what is morally right *just because* it is beneficial to the agent is clearly not worthy of the high-esteem associated with moral worth.[10] His example depicts a shopkeeper who charges everyone the same price, even if the customer is a child (GMS, AA 04: 397). Yet he does this not out of "principles of honesty", but rather out of prudential reasons: he is afraid that if people discover that he overcharges inexperienced customers, they will discredit him and run to his competitors. So, this merchant does comply with the duty to be honest regarding the price of his products, but his action has no moral worth. If he had no competitors, he would not mind overcharging unexperienced customers. It follows that morally worthy actions are clearly not motivated by a mediate inclination.

The difficult question is whether there is some difference to moral worth, let us say, in a

[9] See V-Mo/Mron II, AA 29: 607: "Even happiness in the bad will is nothing good. If a happy man does not have a good will, he laughs at the unfortunate, and does nothing to help. The good will is good without restriction, because everywhere it is itself the restriction."

[10] Different metaethical readings about moral worth and value in the Kantian literature can have significant implications for how one understands other aspects of Kant's practical philosophy. Nonetheless, since my discussion of moral worth here is brief and intended only to pave the way for the three propositions on duty, I have tried remain neutral, without committing me to realist or antirealist views. For more on this, see Rauscher (2018).

man helping someone because he feels sympathy or pity (cases of immediate inclination), or because he just recognizes that helping is a duty (a case of action purely from duty). Kant presents us with three different situations (GMS, AA 04: 398). In the first, a man helps others not with some "motivating ground of vanity, or self-interest", but because he is "so attuned to compassion", that is, he helps not because he will take some benefit from it, but because he has a natural proclivity to generosity; he has an immediate inclination to beneficence. As we shall see below in detail, to have an inclination means to have a habitual feeling of pleasure or displeasure associated with the representation of an object of choice that the agent can bring about, and to take this feeling as a *motive* (*Bewegungsgrund*) to act and achieve the end. Hence, to have an inclination to beneficence means that the agent takes the satisfaction he gets out of helping – "the inner gratification in spreading joy" (GMS, AA 04: 398) – as the motivation to help.

Insomuch as this kind of motivation to help surely makes the action, which is conform to duty, worthy of "praise and encouragement", it does not give it that high esteem (*Hochschätzung*) accorded to actions with moral worth. This is because "the maxim lacks moral content [*sittlicher Gehalt*], namely to do such action not from inclination, but *from duty*" (GMS, AA 04: 398). In a second situation, Kant supposes that this same man has temporarily lost his natural tendency to be beneficent: his mind is "beclouded by his own grief". Yet if in these conditions he nevertheless still helps someone, then *now* his maxim would have genuine moral worth. In the third and last situation, Kant goes further still. Instead of having *temporarily lost* that natural stimulation to compassion, he supposes that that man has never had any sign of sympathy; indeed, he has a cold "indifference to the suffering of others". In this case, since *ex hypothesi* no inclinations speak in favor of beneficent action, the only motive for him to be beneficent is the recognition of *the duty* to be beneficent. In this case, the action is done not from inclination, but from duty, and this makes it morally worthy.

What can we learn from this example? First, we need to distinguish the motive from the purpose or end of the action. Maxims, as we will see in more detail, are the principles agents act on. As such, they express the agent's conception of what she is doing (the end she is trying to achieve, through which action, in what circumstances) for the sake of what (her motive, that which gives her a reason to pursue the end). In Kant's example of beneficence, the agent's purpose is always the same: to help those in need. What changes is the motive. He can say to himself "I will help those in need since that is the right thing to do (that is my duty, etc.)" or "I will help those in need since it brings me satisfaction and joy". Kant's point, as we shall see below, is that both maxims are in conformity to duty, yet only the former brings moral worth

to the action. Second, we should not conclude from Kant's example that for an action to have moral worth the agent must have no inclinations in support of it, let alone that the agent must overcome competing inclinations. It must be kept in mind here that in the example Kant is talking about *the same man*: first he has an immediate inclination towards beneficence; then, this tendency to help is temporarily unavailable (his mind is "beclouded by sorrow"); finally, he has no favorable inclination at all, having even contrary ones. As Herman (1993, p. 18, orig. emphasis) puts it: "Of *him* it is then said: only when the inclination to help others is not available does *his* helping action have moral worth". We should not generalize the situation faced by this *specific man* and conclude that only when there is no concurrent inclination does the action have moral worth. Indeed, Kant argues from what has been called the method of isolation: since for him only one kind of motive can confer moral worth to an action, he stipulates a situation – e.g. help is demanded from the agent – and at each step removes one possible motive (self-interest, immediate inclination such as compassion, and so on) for action while keeping the same purpose, up to the point where only *that* motive relevant for moral worth is present.[11]

But why is it that only actions done from duty have moral worth? Kant's argument for this thesis concerns the fact that whenever someone acts out of inclination, the person is not responding to a *moral* concern: she is just exhibiting a natural, unreflected, and accidental response related to her propensity to feeling pleasure or displeasure concerning her actions. This natural response can in many times lead to a morally right action, but there is no guarantee that it will always do. In fact, there is always the possibility that it might lead to an *immoral* action. Here is a good example, adapted from Herman (1993, p. 5), to show the contingent relation between inclinations and morally good action: suppose Daniel is walking down the street at night, passing in front of the Pinacoteca's backdoor. He sees someone with a hood on struggling to load a heavy statue into a van. As someone who has an immediate inclination to beneficence and who takes it as a motive to act, Daniel of course offers to help the person load the statue. Now, it does not require much reflection to realize that he just helped a thief stealing an artwork. Appropriately, this shows very clearly that "the class of actions that follows from the inclinations to help others is not a subset of the class of right or dutiful actions".[12] Reliance

[11] See Paton (1947, p. 47): "Hence in order to be quite certain that we are judging that value of actions done for the sake of duty, he asks us to remove the immediate inclination and assess the value of action in its absence. Thus, for example, he asserts that when we have through grief lost the immediate inclination to live and desire nothing so much as death, there still remains the duty of preserving our own life; and that in doing so, not from inclination, but for the sake of duty, the action has for the first time genuine and unique moral worth".

[12] I adapted this example from Herman (1993, p. 5). See also Aune (1979, p. 11): "Kant would certainly want to say that a principle of benevolence could not *itself* confer moral value on actions because benevolent actions are not good without qualification. As he no doubt sees it, there are countless possible conditions in which benevolence

on inclinations as the only source of motivation does not guarantee that the action will always be morally right. And even when inclinations do lead an agent to act morally, they do so in a very circumstantial way. Kant's central point, thus, is that only the motive of duty can give moral worth to an action because all other motives give only a "very contingent and precarious" (*sehr zufällig und mißlich*) (GMS, AA 04: 390) motive to do what morality requires.[13] And this brings us to is his famous missing first proportion on duty:

> **First proposition**: human actions have moral worth only when done from duty.[14]

Kant's beneficence example also shows that the moral worth of an action is determined not by its consequences, nor by virtue of the agent's purposes. After all, in the three situations, the purpose (to help someone in need) and the consequence (someone in need is helped) are the same, but there is still a difference in moral worth. According to the second proposition, the moral worth of an action is determined by the agent's practical principle, that is, her *maxim*:

> **Second proposition:** "An action from duty has its moral worth *not in the purpose* that is to be attained by it, but in the maxim according to which it is resolved upon, and thus it does not depend on the actuality of the object of the action, but merely on the *principle* of *willing* according to which – regardless of any object of the desiderative faculty – the action is done." (GMS, AA 04: 399-400, orig, emphasis)

---

is morally objectionable–for example, in fighting just wars or in punishing heinous crimes–and it is not at all evident that this view is untenable. One can, of course, insist that benevolence is generally good, or good nearly all the time, but insisting on this point is compatible with acknowledging that any principle of benevolence is subject to exceptions or qualifications. In fact, it is compatible with claiming that a benevolent action is morally wrong when it conflicts with one's moral duty."

[13] See also KpV, 05: 81: "[…] moral worth, must be place solely in this: that the action takes place from duty, for the sake of the law alone."

[14] As Rickeless (2004, p. 555) has made clear, the standard reconstruction "actions have moral worth only when done from duty" needs to be amended to the version used above, for actions of a being with a perfect will do have moral worth, yet are not done from duty. I favor the (amended) traditional version of the first proposition mainly for three reasons: (1) Schönecker and Wood's (2002, p. 60) alternative proposal, according to which the first proposition is "Eine Handlung aus Pflicht ist eine Handlung aus Achtung fürs Gesetz", is closely related to their "no-inclination" reading of moral worth. Since I do not think this reading about moral worth is correct, I reject their proposal. (2) Allison's (2011, p. 125, "A good will under human conditions is one whose maxims have moral content") and Timmermann's (2007, p. 26, "An action that coincides with duty has moral worth if and only if its maxim produces it by necessity, even without or contrary to inclination") proposals are closely connected to a specific interpretation about the identity of the good will. Very briefly, they defend that a good will is to be identified with having a good *Denkungsart* (or *Gesinnung*). Although plausible, I have reservations about this reading, for it seems that it conflates good will and virtue. A criticism of this sort can be found in Herman (1993, pp. 12-3), but see also Ameriks (2003, ch. 7). (3) Since addressing these points would be out of the scope of this dissertation, I prefer the (amended) standard reading – as in Wood (1999) and Potter (1975) –, for I take that the cogency of the interpretation about the derivation that is put forward here is not affected by choosing any of the alternatives for the first proposition. For discussions of the different proposals for the first proposition, see Allison (2011, p. 122-6) and Horn, Mieth and Scarano (2007, p. 185-7).

Looking back at Kant's example, we see that when the man acts out of immediate inclination "*the maxim lacks moral content*, namely to do such actions not from inclination, but *from duty*" (GMS, AA 04: 398, emphasis added). This suggests there is something *about the maxim* that makes an action morally worthy, and this warrants a more detailed consideration of this concept which is at the heart of Kant's theory of action.[15] He offers two definitions of maxims in the *Groundwork*. In the first section, he says:

> A *maxim* is the subjective principle of willing; the objective principle (i.e., the one that would also subjectively serve all rational beings as the practical principle if reason had complete control over the desiderative faculty) is the practical *law*. (GMS, AA 04: 401n)

And in the second, with some more detail:

> A *maxim* is the subjective principle for acting, and must be distinguished from the *objective principle*, namely the practical law. The former contains the practical rule that reason determines in conformity with the conditions of the subject (quite often his ignorance, or his inclinations), and is thus the principle according to which the subject *acts*; but the law is the objective principle, valid for every rational being, and the principle according to which it *ought to act*, i.e. an imperative. (GMS, AA 04: 420-1n)

The first thing to note is the sharp contrast Kant makes between maxims and laws. In the first definition, by calling the former subjective and the latter objective practical principles, and saying that if *reason* had complete control of the desiderative faculty (more on this faculty below), Kant implies that there is a normative relation between them. Agent acts on maxims, and if they were fully rational, their maxims would in some sense be in conformity to laws. This becomes clearer when Kant says in the second definition that maxims are the principles according to which agents *act*, whereas laws are the principles according to which they *ought to act*. Moreover, this relation of conformity also provides a way to evaluate the rationality of maxims: the closer they are to laws, the more rational it is to adopt them. (In the second section we will see that maxims are judged rational or reasonable according to two different standards, namely two kinds of imperatives: instrumental and moral).

The sharp contrast between maxims and laws also shows that a maxim, *qua* subjective principle*,* is freely chosen by the agent, and that it embodies a description of the action *from the point of view of the agent*. It reveals what circumstances she thought she was

[15] I am following here Kitcher's (2004, 565-6) suggestion: "Philosophically, Kant's position is motivationalist. The moral character of an action does not depend on its purpose or end, but on its relation to motivation. Given his position, it should be easier to follow his reasoning by highlighting the motivational aspect of maxims."

in, what action she thought was appropriate in such circumstances, for the sake of what purpose, and for what reasons.[16] Schematically, maxims have been described with the following structure: *In C, I to do A for purpose P, because of G*.[17] Where (C) specifies a circumstance, (A) specifies an action, and (P) specifies the purpose, or the end that is aimed for, in doing A in C, for the sake of motive (G). Now, Kant's examples of maxims are inconsistent as to their level of generality, such that apparently none seems to perfectly match this schema. Sometimes he speaks of maxims as if they were general policies of conduct, general rules of live for the long-haul: "to let no insult pass unavenged" (KpV, AA 05: 19). Sometimes he does not even spell out the maxim of the agent, only suggesting his "attitude" or "way of thinking" (*Denkungsart*).[18] I nonetheless adopt this schema because it captures the structure of any maxim *when fully spelled out*. For example, one of Kant's most paradigmatic maxims – and maybe the one that is most precisely formulated – is "when I believe myself to be in need of money I shall borrow money, and promise to repay it, even though I know that it will never happen" (GMS, AA 04: 422). This maxim contains only a description of a circumstance (financial need), and of an act (making a false promise), but the purpose or end of making a false promise is obvious: to get out of a financially dire situation. Moreover, Kant says that this is a principle of "self-love, of one's own benefit", which reveals the agent's motive: he wants to make a false promise *because that will be in his benefit*. It was the same with the maxim of the prudent shopkeeper, for instance. He charged the right price "not from principles of honesty" but because that would benefit him. That is his motivation in charging the right price, and so that must be part of his maxim. Had he charged the right price thinking that "this is the honest thing to do", his motive would be different, and so his maxim.[19]

What I am interested now is just in this motivational component of the maxim – the (G) component –, because it is this part of if that can give it moral content, depending on what kind of motivation the agent adopts. In fact, for Kant all motivation can be reduced to either inclination (or a set of inclinations) or duty (in the form of respect for law). We have briefly

[16] As Herman (2016, p. 32, orig. emphasis) says: "The maxim presents the action that I will to do in a form that makes the structure of my willing explicit. The maxim shows *how* particular features of the situation are reasons for acting by presenting them in a rule or principle that dictates doing what I did in just these sorts of circumstances, and, explicitly or implicitly, for what purpose."

[17] I take this schema from Kitcher (2004). Herman (2016) offers a similar one.

[18] As in the beneficence example of the second section (GMS, AA 04: 423). I discuss this example in the next chapter.

[19] Here is a good example from Kant about maxims that have the same purpose yet are indeed significantly different because of their motive: "The *maxim* of greedy avarice (prodigality) is to get and maintain all the means to a good living *with the intention of enjoyment.* – The maxim of *miserly* avarice, on the other hand, is to acquire as well as to maintain all the means to good living, but *with no intention of enjoyment* (i.e., in such a way that one's end is only possession, not enjoyment)" (MS, AA 06: 432).

characterized inclinations above, saying that they are a habitual response to the feeling of pleasure and displeasure that comes with the representation of an object as a possible object of choice. But to fully understand what inclinations are, and what is the feeling of respect for law, we need to take a brief step back into the fundaments of Kant's theory of rational action.

The starting point for this is the *Begehrungsvermögen,* that is, the "desiderative faculty" or the "faculty of desire". In the *Groundwork*, Kant is very brief when treating this faculty. He says in the second proposition on duty that the faculty of desire has "objects" (GMS, AA 04: 400), and in a footnote in the second section he remarks that the "dependence of the desiderative faculty on sensations is called inclination, and this therefore always proves a *need*" (GMS, AA 04: 413, orig. emphasis).[20] In the *Critique of Practical Reason*, Kant defines this faculty, which belongs to all beings (and not only rational beings), as "a being's *faculty to be by means of its representations the cause of the reality of the objects of these representations*" (KpV, AA 05: 09, orig. emphasis).[21] Moreover, he acknowledges the lack of clarity in his previous treatment of this concept, yet reveals one of his reasons for that: we must be careful not to offer a definition of "faculty of desire" that leaves no room for the possibility of its being determined in a non-sensible way. Appropriately, the faculty of desire is nothing else than the capacity belonging to every living being to represent an object, desire it, and by virtue of this desire, make the represented object existent or effective. The talk of "representing an object" might suggest that reason is necessarily involved in this task. Yet this is not the case, for this is not a strictly rational capacity: just as we, rational beings, can represent or imagine the delight of having a certain dish, desire it, and then make it "effective", so to speak, by going to a restaurant or cooking it ourselves, so a lion can see a prey, desire it as an object to be effected, and try to catch it.[22]

In fact, desiring or aversing an object is always connected to a feeling of pleasure or displeasure, respectively. In the *Metaphysics of Morals*, Kant says that "*pleasure* or *displeasure*, susceptibility to which is called *feeling*, is always connected with desire or

---

[20] "die Abhängigkeit des Begehrungsvermögens von Empfindungen heißt Neigung, und diese beweistet also jederzeit ein Bedürfnis".

[21] "Das Begehrungsvermögen ist das Vermögen desselben, durch seine Vorstellungen Ursache von der Wirklichkeit der Gegenstände dieser Vorstellungen zu sein."

[22] The difference between the rational agent and the lion (or any non-rational animal, for what matters) lies in the fact that the latter have what Kant calls *arbitrium brutum*, such that its faculty of choice "cannot be determined other than through sensible impulses". The former, however, have an *arbitrium liberum*, which means that its faculty of choice can be determined by sensible impulses, but not necessarily is; it can also be determined to action quite independently of sensible impulses, "through motives that can only be represented in reason" (KrV, A802/B830). As Allison (1990, p. 38-9) and Guyer (2017) argue, the distinction between these two kinds of *arbitria* later developed into Kant's Incorporation Thesis (for more on this see footnotes 32 and 33 below).

aversion" (MS, AA 06: 211, orig. emphasis). It is fundamental to see, however, that the feeling of pleasure or displeasure is not the cause of every desire and aversion. In other words, some of our desires and aversion have a non-sensible cause. And this is where Kant's aforementioned precaution when defining the faculty of desire pays off, for if he had affirmed that all desires depend on some affection or feeling, as the *cause* of that very desire for the object, then there would be no possibility for practical reason to be pure. As Engstrom rightfully observes, that would be to introduce a throughout-going empiricism by definitional fiat.[23] Kant makes this point very clear when he says that

> It is easily seen that the question whether pleasure must always be put at the basis of the faculty of desire or whether under certain conditions pleasure only follows upon its determination, is left undecided by this exposition […] Such a precaution – namely, not to anticipate one's judgments by definitions ventured before complete analysis of the concept, which is often achieved very late – is to be highly recommended throughout philosophy, and yet is often neglected. (KpV, AA 05: 10)

Now, it is precisely when the feeling of pleasure is the cause of the desire for an object that we can talk about inclinations. We have seen right above that in the *Groundwork* Kant defines an inclination as the "dependency of the faculty of desire on sensations" (GMS, AA 04: 413). In the *Metaphysics of Morals* we can appreciate with more precision the relation between inclinations and the faculty of desire. Kant says:

> As for practical pleasure [i.e., pleasure related to desiring something, in opposition to merely contemplating it, as in *taste*] that determination of the faculty of desire which is caused and therefore necessarily *preceded* by such a pleasure is called *desire* in the narrow sense; habitual desire is called *inclination.* (MS, AA 06: 212) [24]

In other words, the prospect of enjoying an object causes a feeling of pleasure.[25] And when this feeling is the cause of a recurrent or habitual desire for the object, that is, when this feeling of pleasure is what regularly prompts the agent into making that desired object something existent and real (*wirklich*), then we can say that the agent has an inclination for that

---

[23] See Engstrom (2009, p. 27): "Kant emphasizes that the faculty of desire should not be defined in a way that would build into its very concept the supposition that a feeling of pleasure must lie at the basis of its determination, so that desiring something always depends in the end on this capacity's being or having been affected by some pleasing experience. To do so would be in effect to rule out from the start, by definitional fiat, the possibility of conceiving of the will as practical reason rather than merely as rational desire. It would be to build into the very concept of desire the Humean supposition that desire cannot arise from reason, even though it may be directed by it. A thoroughgoing empiricism would be introduced, by stipulation, into practical philosophy."

[24] "Was aber die praktische Lust betrifft, so wird die Bestimmung des Begehrungsvermögens, vor welcher diese Lust als Ursache nothwendig vorhergehen muß, im engen Verstande Begierde, die habituelle Begierde aber Neigung heißen."

[25] As in KpV, AA 05: 20: "However dissimilar representations of objects may be […] the feeling of pleasure by which alone they properly constitute the determining ground of the will (the agreeableness, the gratification expected from the object, which impels activity to produce it) […]". See also Willaschek (1992, p. 61).

object.[26] And very importantly, by "object" (*Gegenstand*) of the faculty of desire Kant understands whatever thing we might desire, be it simple substance like a book, an activity (going on a hunt, playing a game) and, primarily, a state of affair in general (KpV, AA 05: 23).[27] In GMS I, for instance, Kant describes agents who have inclinations towards very different kinds of objects, namely inclinations towards preserving their own lives, towards honor, philanthropy, and happiness.

Nevertheless, to say that inclinations (as habitual desires) *determine* the faculty of desire demands careful qualifications. This means that we have no control over what objects our sensible affections will make us desire; however, it does not mean that we have no control over how our desires influence or determine our will.[28] Better put, it does not mean that we are causally determined to act in function of our sensible desires. Inclinations are *incentives*, which compel us into adopting certain maxims: to do something – which is specified in the maxim – in order to suffice the need (*Bedürfnis*) for an object that is the material of the faculty of desire. But this need does not causally determine us into acting in order to satisfy the inclination. We have the capacity to *choose* on which maxims we are going to act on, and we choose to satisfy an inclination by incorporating it into our maxim *as an incentive* (*Triebfeder*).[29]

Appropriately, we can now come back to the discussion of Kant's three propositions on duty to see what feature of the maxim gives it moral content. I said that what was important for my purposes was the motivational component of the maxim (G), so now consider again the

---

[26] See also Anth, AA 08: 251.

[27] See Allison (2011, p. 262): "By 'object of volition' Kant understands an objective or possible state of affairs that is to be brought about intentionally through the agency of the subject, because it is viewed by the subject as in some sense good."

[28] See Herman (2016, p. 7): "Our desires may prompt us to seek an end (and we are not free with respect to the desires we have), but as rational beings we do not automatically have an end–we are not automatically moved to act–as the result of having a (strongest) desire". This is rooted in Kant's distinction between *arbitrium brutum* and *arbitrium liberum*: "**Freedom in the practical sense** is the independence of the power of choice from **necessitation** by impulses of sensibility. For a power of choice is **sensible** insofar as it is **pathologically affected** (through moving-cause of sensibility); it is called an **animal** power of choice (*arbitrium brutum*) if it can be **pathologically necessitated.** The human power of choice [*Willkür*] is indeed an *arbitrium sensitivum*, yet not *brutum* but *liberum*, because sensibility does not render its action necessary, but in the human being there is a faculty of determining oneself from oneself, independently of necessitation by sensible impulses." (KrV, A 534/B 562, orig. bold).

[29] An incentive is the "subjective ground of desiring". It impels, but does not *determine* the agent to action. One must *choose* a specific maxim – either one with a formal or with a material incentive – as a motivating ground (*Bewegungsgrund*), and only then the incentive becomes a motive to act on (see GMS, AA 04: 427). For more on this, see Willaschek (1992, p. 73): "Kant spricht sehr anschaulich von der 'Triebfedern' des Handelns. Die Neigung etwa sind solche Triebfedern, denn sie motivieren dazu, auf eine bestimmte Weise zu handeln. Für ein Wesen wie den Menschen, dessen Willkür 'vernünftige bestimmbar' ist, sind Neigungen jedoch nur *Prima-facie*-Motive, denn Menschen geben nicht jeder Neigung nach, sondern entscheiden sich aufgrund vernünftiger Überlegungen". Accordingly, this feature of Kant's theory of agency is known in the literature as the "Incorporation Thesis", most famously discussed by Allison (1990, p. 38-9).

beneficence example of GMS I. It seems reasonable to say that the maxims of helping from immediate inclination and helping from duty are the same *except* for the motivational component. They relate the same circumstances ("people need help") and they propose the same end ("to provide help"), yet their ground of adoption are different motives. In the former case, the agent helps with the incentive of inclination, which means his maxim is a subjective *material* practical principle. In the latter, he helps with the incentive provided by the recognition that it is his duty to help, and so his maxim is a subjective *formal* practical principle. This means that it is precisely in virtue of what incentive is incorporated into the maxim that its moral content is established.[30] Consequently, the moral worth of an action depends on the incentive the agent choses to act on, by incorporating it to his maxim as a *motive* (*Bewegungsgrund*) to act. As Herman (1993, p. 11) nicely puts it:

> The key to understanding Kant is in the idea that moral worth does not turn on the presence or absence of inclination supporting an action, but on its inclusion in the agent's maxim *as* a determining ground of action: as a motive. Kantian motives are neither desires nor causes. An agent's motives reflect his *reasons* for acting. An agent may also find reasons in his passions, principles, or practical interests. All of these, in themselves, are 'incentives' (*Triebfedern*), not motives, to action. It is a mark of a rational agent that incentives determine the will only as they are taken up into an agent's maxim. Indeed, it is only when an agent has a maxim that we can talk about his motive.

Moreover, in the paragraph explaining the second proposition Kant says that the "will stands halfway between its **a priori** principle, which is formal, and its **a posteriori** incentive, which is material, as it were at a crossroads" (GMS, AA 04: 400, orig. emphasis). And since the examples have shown that inclinations do not confer moral worth to actions, it follows by elimination[31] that when an action is done from duty it is done with a maxim that has a *formal incentive.*[32] Better said, it is done from "a formal principle of willing as such … as every material principle has been taken away from it" (GMS, AA 04: 400). Hence, the introduction of the concept of *respect for the law* (*Achtung fürs Gesetz*) in the third proposition on duty comes precisely to explain what is the incentive incorporated in maxims with moral content.

---

[30] See Timmermann (2007, p. 29n32): "'Moral content' and 'moral worth': it is the moral *content* of the maxim on which we act that makes the action morally *good*. Moral content of a maxim is equated with the commitment to do the morally correct action from duty."

[31] See Allison (2011, p. 126): "In his explanation of this proposition, Kant makes it clear that he is adopting his usual strategy of arguing by elimination and that he bases this elimination on the results of the preceding argument."

[32] See Timmermann (2007, p. 180): "Reverence for the law (*Achtung*) – the desire to do what is morally required (IV 440.6-7) – and inclination (*Neigung*) are competing *Triebfedern*."

Third proposition: "*Duty is the necessity of an action from respect for the law*" (GMS, AA 04: 400, orig. emphasis).

Now, ever since the preface Kant has been saying that he is working with a given conception of morality according to which morality concerns duties that bind us with absolute necessity, in the sense that they are obligation from which we cannot excuse ourselves from complying. And for this to be the case, these obligations must be commanded by moral laws, which must have an *a priori* ground. Still, the existence of moral laws that bind with necessity does not imply that agents under their binding scope will in fact comply with them. Finite rational beings, like human beings, act on maxims, which are the "principle[s] according to which the subject *acts*". Laws, however, as objective practical principles, are the principles "according to which it [i.e. the rational being] *ought to act*" (GMS, AA 04: 421).[33] This means that there is always a relation of *conformity* between maxims and laws: the principle the agent adopts may or may not be the principle she ought to adopt in that situation. As Kant says in one of the *Lectures on Ethics* (V-Mo/Mron II, AA 29: 603): "Objective principia are laws, and differ from subjective principles, or from maxims from which I act. Objective principia are those in which morality consists, and subjective, those whereby I attain to morality".[34]

Yet, as we saw right above, one distinctive property of maxims is that they contain the agent's motivation for doing a certain action: they incorporate an incentive. And as we noted, it is precisely this motivational component of the maxim that determines the moral worth of the action. Therefore, when it comes to moral worth, it is not enough that the agent adopts a maxim that conforms to a law; she must adopt it *because* it conforms to a law or, better said, "only on the condition that [it] conform[s]" (Allison, 2011, p. 140). With the examples of *Groundwork* I, Kant wanted (among other things) to show the relevance of the distinction between acting from duty and acting in conformity with duty, since it is this distinction which reveals that inclinations are morally inadequate incentives.

The third proposition on duty shows that the incentive of morally worthy actions is *respect for the law*. Doubtless, the mention of a *feeling* as the determining motivational factor in morally worthy actions might be puzzling to the readers of the *Groundwork*. After all, is it

[33] Kant also clearly distinguishes maxims from laws at the beginning of the second *Critique*: "Practical principles are propositions that contain a general determination of the will, having under it several practical rules. They are subjective, or *maxims*, when the condition is regarded by the subject as holding only for his will; but they are objective, or practical *laws*, when the condition is cognized as objective, that is, as holding for the will of every rational being." (KpV, AA 05: 19).

[34] See also O'Neill (2015, p. 122): "Laws and [objective] principles are *types*, which might be adopted or rejected as maxims by one or many or all agents, at some or many or all times. They set out the content of possible maxims, just as propositions set out the content of possible beliefs."

not a foundation for a *metaphysics* of morals? How can Kant (or anyone, for what matters) know anything *a priori* about a feeling, something notoriously related to sensibility? What is the difference between the feeling of respect for law and the inclinations? Would it not be just another inclination? And for what is more: if to act in a truly morally worthy way we need to be influenced by a feeling, how can Kant's thesis that there are moral laws not fall to the ground? For he says that if a law "relies in the least part on empirical grounds, perhaps just for a motivating ground [*Bewegungsgrund*]", then it is not actually a law, but rather a "prescription that is founded on principles of mere experience" (GMS, AA 04: 389). It is just a practical rule (*practische Regel*), without any valid pretension to universality or necessity.

Naturally, Kant is well aware of these difficulties. His strategy thus consists in claiming that the feeling of respect for law is a *sui generis* feeling. It is not "*received* by influence" but rather "*self-wrought* by a rational concept" (GMS, AA 04: 401, orig. emphasis). This means that differently from the feelings of pleasure and displeasure, the feeling of respect for law does not have a sensible origin. It is not caused by any sensible affection, and this makes it "specifically different from all feeling of the former kind, which come down to inclination or fear" (GMS, AA 04: 401). Kant continues in the following terms:

> What I recognize immediately as a law for myself I recognize with respect, which signifies merely the consciousness of the *subordination* of my will to a law, without mediation of other influences on my sense. The immediate determination of the will by the law and the consciousness of this is called *respect*, so that it is viewed as the *effect* of the law on the subject and not as its *cause*. (GMS, AA 04: 401)

Importantly, the feeling of respect for the law is the *effect* of the determination of the will by the law, and not the *cause* of this determination. Let us make this clearer by explaining what it would be like if the feeling of respect were the cause of the determination of the will. Suppose an agent faces a situation where she must choose between two different courses of action. One of them is favored by an inclination – by the agent judging that doing so would be convenient or advantageous –, and the other one is favored by her moral judging – her sense of "what is right". For the sake of the argument, we can imagine that the value of the inclination (as an incentive for action) conflicts with the feeling of respect for law, which is attached to the agent's moral sense: there is a struggle between these two sources of motivation. Now, because this feeling of respect "infringes on my self-love" (GMS, AA 04: 401), it outweighs the inclination, such that the agent then endorses the moral law as a principle of action and acts correspondently.

Certainly, this reading would be mistaken. Kant makes it clear that his feeling of respect for law is not like the moral feelings of the sentimentalists. We do not *accept* the

authority of the moral law – that is, abide by its prescriptions – because we have a feeling that impels us to do so: "There is here no *antecedent* feeling in the subject that would be attuned to morality" and "for this no special kind of feeling need be assumed, under the name of a practical or moral feeling preceding the moral law and serving as its basis" (KpV, AA 05: 75). And it is clear that it has to be like this, because if it were not, there would be no moral *laws* properly; rather, there would be only moral rules. A moral feeling that is pathologically affected, that is, caused by means of sensibility, is not guaranteed to always outweigh inclinations, let alone do so in every agent: yet duty is the *necessity* of an action from respect for law. Therefore, the feeling of respect "does not serve for appraising actions and certainly not for grounding the objective moral law itself, but only as an incentive to make this law its maxim" (KpV, AA 05: 76).

Thus, it is not that we feel respect and then accept the moral law and resolve to follow it. What happens is the reverse: the law determines the will "without mediation of other influences on my sense" (GMS, AA 04: 401). When Kant says that "what I recognize *immediately* as a law for myself I recognize with respect" (emphasis added), he is saying that we instantly recognize the law as binding, without mediation of anything. The recognition of a principle as a law is so authoritative that it immediately determines the agents will to comply with it and induces, or arouses, a certain kind of feeling. This feeling is *reverence* or *respect for the law*.[35]

But how exactly does the determination of the will by a law elicits this feeling? What is more, how is it that we can be conscious of a feeling *a priori*? It is difficult to find answers to these question in Kant's account of the feeling of respect in the *Groundwork*. To offer a more detailed explanation, while still paying attention to the differences between the accounts of these two works, we will bring in the explanation Kant gives in the third chapter of the *Critique of Practical Reason,* entitled "On the incentives of pure practical reason".[36] There Kant says that our nature as sensible beings is such that the faculty of desire is always affected by *sensible* desires.[37] And not only are we always under the influence of these desires (and

[35] See Paton (1947. p. 65): "On Kant's view we feel reverence because we recognize that the law is binding upon our wills. The great error of the moral sense school is to suppose that the law is binding because we feel reverence. No feeling can be the basis of a binding moral law, but the moral law may be the ground of a specific moral emotion."

[36] The main difference between the two accounts concerns the fact that in the *Groundwork* Kant opposes respect and self-love, whereas in the *Critique of Practical Reason* he offers a more detailed moral psychology, opposing respect to self-love and to self-conceit. On the differences between the two accounts, see Allison (2011, p. 131n20).

[37] In fact, having a faculty of desire that is affected by sensible desires is a defining property, for Kant, of every living being: "The *faculty of desire* is the faculty to be, by means of one's representations, the cause of the objects of these representations. The faculty of a being to act in accordance with its representations is called *life*" (MS, AA 06: 211).

when they are habitual, of inclinations) but we also have the tendency of taking the demands of our sensible nature as "primarily and originally valid" (KpV, AA 05: 74). Taking all our inclinations as a set, we have what Kant calls *self-love*, which can be described as "the propensity to make oneself as having subjective determining grounds of choice into the objective determining ground of the will in general" (KpV, AA 05: 74). Notwithstanding, we know that when it comes to acting morally what must serve the will as its determining ground are practical laws, and that they must determine the will without mediation of the senses. This creates a problem for *self-love*, since that propensity in prioritizing the inclinations – which are essentially subjective – conflicts with the objective, universal, and necessary character of obligations established by practical laws. After all, practical laws determine "not only without the cooperation of sensible impulses but even with rejection of all of them and with infringement upon all inclinations insofar as they could be opposed to that law" (KpV, AA 05: 72).

Appropriately, as the will is determined by a practical law, self-love's propensity in prioritizing inclinations to the law's expense is severely repealed. The determination of the will by the law "outweighs" any inclination, "or at least excludes it entirely from calculations when we make a choice" (GMS, AA 04: 400). It is in this sense that the feeling of respect "is actually the representation of a worth that *infringes* [*Abbruch tut*] on my self-love" (GMS, AA 04: 401, emphasis added) and that it "by thwarting all our inclinations produce[s] a feeling that can be called pain". (KpV, AA 05: 73). Knowing that all inclinations under the heading of the concept of self-love have their value as motives [*Bewegungsgründe*] necessarily diminished by the feeling of respect for law, we can then cognize one of the effects exercised by this feeling without having to appeal to experience. This is Kant's explanation of how we can cognize *a priori* the workings of the feeling of respect. This feeling does not affect us through sensibility, but manifests itself when the moral law necessarily causes a certain effect when it excludes the preponderance of inclinations at the moment of choice. Once we understand that the feeling of respect for law is the result of the will being determined by the law, and that this determination runs counter to inclinations, we do not actually need to feel it in order to be able to know what kind of effect it will have on us.

But the feeling of respect with its property of outweighing the force of inclinations as motives is the negative side of the determination of the will by the law. There is also a positive one, since this determining law is also "the form of an intellectual causality" (KpV, AA 05: 73) and the feeling of respect is "self-wrought [*selbstgewirkt*] by a rational concept" (GMS, AA 04: 401). Moreover, it is this positive side that better explains why exactly the

determination of the will by the law elicits a feeling of *respect*, and not some other feeling or emotion. Why do we respect the law but not one or more objects of desire, or the inclinations themselves? To this Kant says

> For the object as the effect of the action I have in mind I can indeed have *inclination*, but not *respect*, precisely because it is merely an effect and not activity of a will. Likewise, I cannot have respect for inclination as such, whether it is mine or that of another; I can at most in the first case approve of it, in the second at times love it myself, i.e. view it as favourable to my own advantage. (GMS, AA 04: 400)

The reason why we have *respect* for the law is because the law is not an object produced by some action, but rather *an activity of a will*. Now, in a crucial yet underappreciated sentence, Kant introduces in GMS I one of the most fundamental tenets of his moral philosophy. The laws for which we feel respect are an activity of the will in the sense that they are *self-legislated*: "The *object* of respect is therefore solely the law, the one that we impose upon *ourselves* and yet as in itself necessary." (GMS 401n, orig. emphasis). Anticipating an argument to be developed in the final parts of GMS II, Kant claims that it is precisely because they are self-legislated that these laws are unconditionally binding. We are bound by them despite our non-moral interests and discretionary ends: "as a law we subject to it, without consulting self-love; *as imposed upon us by ourselves, it is yet a consequence of our own will* […]" (GMS, AA 04: 401n, emphasis added).

Here we have the textual evidence that supports the first thesis: that in GMS I, Kant is already arguing with the assumption that moral laws are self-legislated. It is because moral laws are self-legislated that he says they are *an activity of the will* and, correspondently, that is why we feel respect for them. That "reason extracts from me immediate respect" for a maxim that can "fit as principle into a possible universal legislation" is something that every agent experiences, although they "do not yet see on what it is founded" (GMS AA, 04: 403). As we will come to learn later, it is founded on "the idea of the dignity of a rational being that obeys no law other than that which at the same time it itself gives" (GMS, AA 04: 434) because it is "[o]ur own will, in so far as it would act only under the condition of a possible universal legislation through its maxims […] is the actual object of respect" (GMS, AA 04: 440).[38] Kant is talking about autonomy here, and in the next section we shall see how this introduction of autonomy in *Groundwork* I, albeit tacit, is paramount to the derivation of the FUL.

Before we get to the next section, though, let us summarize Kant's arguments in GMS I so far. He has been arguing ever since the preface that we take our moral obligations to

[38] See Alison (1990, p. 126): "Accordingly, what we strictly speaking respect, on this view, is the autonomy of pure practical reason in ourselves and, by extension, in all rational agents". See also Velleman (2005a).

be commanded by moral laws. And practical principles are moral laws when they are *self-legislated*. Indeed, our attitude of feeling respect for moral laws is the mark of the recognition of their distinct kind of normativity. We do not show respect for the objects of the faculty of desire, let alone for our inclinations. We may very well like and love our inclinations, since we enjoy their objects. In fact, inclinations too have a normative character, since we are compelled to "remedy the need" (GMS, AA 04: 413) they create, by adopting *material* practical principles. Our attitude towards this kind of practical principle, however, does not exhibit that *high-esteem* [*Hochschätzung*] we show for principles that originate from our own will. When agents engage in practical deliberation, when they think about what to do, the mere "*representation of the law* in itself" (GMS, AA 04: 401, orig. emphasis) outweighs and "excludes [the inclinations] entirely from calculations when we make a choice" (GMS, AA 04: 400).[39] It is exactly this deliberative priority that shows the greater normative force of moral laws compared to material practical principles, and which explains why we have *reverence* or *respect* for them.

Of course, this greater normative force does not guarantee that the agent will abide by the moral laws, that she will necessarily adopt a maxim that conforms to a law. That would only be the case if "reason had complete control over the desiderative faculty" (GMS, AA 04: 400).[40] But when the agent does conform her maxim to the law, and does so motivated by the conscience that she ought to, **then** her action has moral worth. Kant ends the exposition of the third proposition on duty saying that "an action from duty is to separate off entirely the influence of inclination, and with it every object of the will; thus nothing remains for the will that could determine it except, objectively, the *law* and, subjectively, *pure respect* for this practical law" (GMS, AA 04: 400).

1.3 - The derivation of the FUL

With the three propositions on duty, Kant's analysis of common moral cognition in GMS I has established that: (1) to choose to act from duty is the kind of volition proper of a

[39] This allows us to take two conclusions: (1) perfect rational beings (e.g. God) cannot feel respect for the law. Indeed, they are not subject to the influences of incentives in general. Since these beings are not sensibly affected by the feeling of pleasure and displeasure, they do not have inclinations. Since they do not have inclinations, they do have nothing that hinders the determination of the will by the law: "Gott hat keine Achtung fürs Gesetz, denn er hat keine Neigung, die die Achtung einschränken soll" (AA 27: 1326) and as Klemme (2017, p. 76) observes: "Ein reines Vernunftwesen befolg das Gesetz ja zwanglos mit Notwendigkeit"; (2) Non-rational animals also do not have any incentives. To have an incentive means to have the capacity to decide whether to choose it as a motivating ground (*Bewegungsgrund*), and this only takes place in beings "whose reason does not by its nature necessarily conform with the objective law." For more on this see Allison (1990, p. 122).

[40] "wenn Vernunft volle Gewalt über das Begehrungsvermögen hätte."

good will under "limitations and hindrances", and only actions that are done from duty have moral worth; (2) moral worth depends not on the quality of the agent's intention nor on the consequences of the action, but on the kind of practical principle on which she acts; and (3) that this principle acted on is the representation of a practical law that determines the will immediately and without acquiescence to inclination. This preponderance over inclination, in turn, elicits a feeling of respect for the law, which the agent then endorses as a subjective motivating ground to act. Accordingly, the derivation of the FUL in a rather cryptic paragraph is put in motion by this focal question: what kind of law is this, and what does it command? Kant says:

> But what kind of law can it possibly be, the representation of which – even without regard for the effect expected from it – must determine the will for it to be called good absolutely and without limitation? Since I have robbed the will of all impulses that could arise for it from following some particular law, nothing remains but as such the universal conformity of actions with law, which alone is to serve the will as its principle, i.e. I ought never to proceed except in a way *that I could also will that my maxim should become a universal law*. Here, then, mere conformity with law as such (not founded on any law determined with a view to certain actions) is what serves the will as its principle, and must so serve it if duty is not to be as such an empty delusion and a chimerical concept; common human reason in its practical judging is actually in perfect agreement with this, and always has the envisaged principle before its eyes. (GMS, AA 04: 402, orig. emphasis)[41]

Two qualifications are in order. The first is the fact that Kant uses 'law' here to refer both to objective formal practical principles – which command an action without regard to the objects to be attained by it – and objective material practical principles – which command an action precisely in virtue of the objects to be secured by it. These are respectively referred to as categorical and hypothetical imperatives in GMS II. A possible source of confusion here may come from the fact that later Kant denies that the latter have the status of 'laws' since they are not unconditionally binding.[42] The second qualification concerns the fact that Kant moves swiftly between talk of substantive objective principles and practical laws to *the* moral principle (or *the* Moral Law), that is, the supreme principle (*oberste Princip*) on the basis of which agents

[41] "Was kann aber wohl für ein Gesetz sein, dessen Vorstellung, auch ohne auf die daraus erwartete Wirkung Rücksicht zu nehmen, den Willen bestimmen muß, damit dieser schlechterdings und ohne Einschränkung gut heißen könne? Da ich den Willen aller Antriebe beraubet habe, die ihm aus der Befolgung irgend eines Gesetzes entspringen könnten, so bleibt nichts als die allgemeine Gesetzmäßigkeit der Handlungen überhaupt übrig, welche allein dem Willen zum Princip dienen soll, d. i. ich soll niemals anders verfahren, als so, *daß ich auch wollen könne, meine Maxime solle ein allgemeines Gesetz werden.* Hier ist nun die bloße Gesetzmäßigkeit überhaupt (ohne irgend ein auf gewisse Handlungen bestimmtes Gesetz zum Grunde zu legen) das, was dem Willen zum Princip dient, und ihm auch dazu dienen muß, wenn Pflicht nicht überall ein leerer Wahn und chimärischer Begriff sein soll."

[42] Only categorical imperatives are properly practical *laws*, for only they carry "with it that necessity which we demand for a law" (GMS, AA, 04: 420). Hypothetical imperatives only serve as *precepts* of the will, not as laws: "[Laws] … must be categorical; otherwise they are not laws." (KpV, AA, 05: 20).

assess those substantive principles. As Reath (2006, p. 107) points out, Kant's reasoning here is that the concept of a practical law can yield the supreme practical law.

With these qualifications in mind, let us see how that happens. In the first sentence, Kant is questioning what kind of practical principle must a good will follow: is it an objective *material* practical principle or an objective *formal* practical principle? The answer comes with an argument by elimination in the second sentence. As we have seen, an agent acts on a material practical principle when what interests her is the object of some inclination, for the action specified by the principle will satisfy the "need" (*Bedürfnis*) created by the inclination with the attainment of the object.[43] However, the examples of actions done from duty and the corresponding three propositions have established precisely that actions done from inclination have no moral worth, even if in conformity with duty. Kant makes this point by saying that he has "robbed the will of all impulses that could arise for it from following some particular law". By "some *particular* law" (*irgend eines Gesetzes*, my emphasis) he means an objective yet material practical principle, a principle the agent adopts by virtue of its "hoped-for" (*verhoffte*) effects (GMS, AA 04: 401).

This means that an agent with a good will does not adopt a practical principle on account of the material ends that could be attained by it; she does not adopt "any law determined with a view to certain actions" (*irgend ein auf gewisse Handlungen bestimmtes Gesetz*) because the motivation behind this kind of action is not morally worthy.[44] When it comes to a good will, the practical principle is chosen not because the action is good for something else, but because it is itself good.[45] We should recall that a good will is "good just by its willing, i.e. in itself" (*allein durch das Wollen, d. i. an sich, gut*) (GMS, AA 04: 394), and that "an action done from duty is to separate off entirely the influence of inclination, and with it every object of the will" (GMS, AA 04: 400). Therefore, if what is fundamental for the adoption of a practical principle by a good will is not its *matter*, that is, its objects as effects of the action, Kant reckons that it

[43] See GMS, AA 04: 413n (orig. emphasis): "The dependence of the desiderative faculty on sensations is called inclination, and this therefore always proves a *need*."

[44] See Timmermann (2007, p. 44): "… the preceding analysis of the concept of duty has shown that morally good action is precisely *not* the action that is performed for the sake of some particular end that one intends to promote, realize or bring about. If so, all laws that apply only on condition that one intends to realize some such end must be excluded from our list of candidates. They cannot be the law that inspires reverence and motivates morally good action."

[45] Kant reaffirms this in the second section (GMS, AA 04: 413n): "We saw in the first section: that in an action from duty one must pay attention not to the interest in the object, but merely to that in the action itself and in its principle in reason (the law)."

can only be its *form*,[46] namely, its lawfulness: "thus nothing remains for the will that could determine it except, objectively, the *law* and, subjectively, *pure respect* for this practical law" (GMS, AA 04: 400). Appropriately, he concludes in the derivation paragraph that "nothing remains but as such the universal conformity of actions with law, *which alone is to serve the will as its principle*" (emphasis added).[47]

Recall that it is at this point that critics, especially Aune, argue that Kant's derivation goes astray. Kant has concluded that the agent with a good will chooses a principle not because of its material prospects, but only by virtue of its lawfulness. Accordingly, on Aune's reading, when Kant says that "mere conformity with law as such … is what serves the will as its principle", he is saying that a good will acts on the following principle: "Act on maxims that conform to universal law" (PC). In other words, Aune sees Kant embracing PC as a prescriptive, action-guiding principle. The problem is that Kant's gloss with that "i.e." to the effect that the FUL is equivalent to PC is considered unwarranted, since he is replacing without argument an uninformative action-guiding principle (PC) for an informative yet unjustified one (Wood, 1999, pp. 47-9; Aune, pp. 28-34; Hill, 2000). There appears to be a gap between PC and the FUL which Kant's arguments cannot fulfill.

Now, one of the commentators that have tried to defend Kant's derivation of the FUL from this charge of gapiness is Samuel Kerstein, in his *Kant's Search for the Supreme Principle of Morality*, with what he calls "the criterial reading" (2002, p. 73). According to Kerstein, Kant's strategy in *Groundwork* I and II is twofold: first, Kant develops criteria that any putative supreme principle of morality must fulfill; second, he argues that the FUL does fulfil all the criteria, and that no other candidate for supreme principle of morality – such as a hedonistic or perfectionist principle – satisfies the requirements.

The first criterion Kerstein identifies is that the supreme principle of morality must be such that the agent has a good will only when she is determined to act by this principle. In other words, a will is a *good will* only when this principle determines its actions. This also implies, in connection to Kant's first proposition on duty, that actions have moral worth only when done *because* they conform to this principle, that is, only when done from duty. Second, and in connection to Kant's second proposition, the supreme principle of morality must be such that actions have moral worth not by virtue of their effects and consequences, but only by virtue

[46] Consider again GMS, AA 04: 400, where Kant says that the will is determined by a "formal principle of willing". See also Potter (1975, p. 78).
[47] "so bleibt nichts als die allgemeine Gesetzmäßigkeit der Handlungen überhaupt übrig, welche allein dem Willen zum Princip dienen soll."

of the agent's maxim (GMS, AA 04: 399-400). Third, the supreme principle of morality must be such that just the representation of it as a law (i.e. as a binding principle) suffices to motivate the agent. Otherwise said, when conforming to this principle agents act out of respect for a law (GMS, AA 04: 400-1). Lastly, Kerstein finds a fourth criterion in the second section: the supreme principle of morality must be such that it ought to be possible to derive from it the duties that accord with our common moral knowledge (GMS, AA 04: 421). In other words, this supreme principle must be in coherence with ordinary human reason in moral affairs.

Kerstein's reading is attractive for many reasons: first, it clearly shows how important Kant's analysis of the concepts of good will and duty – and specifically the three propositions on duty – are for the derivation of the FUL. As we shall see, one of the keys to understanding the derivation of the FUL in is found in these propositions, which Aune and Wood do not take in due consideration in their readings. Second, Kerstein clearly shows how Kant can dismiss any consequentialist principle as the supreme principle of morality, since any such principle fails to comply with the second criteria. This is an important achievement, considering that Kant's and Kantian views on morality are frequently assessed in comparison to consequentialist ones. Third, Kerstein's interpretation reinforces the importance of common moral reason in Kant's search for the supreme principle of morality, a feature that can be easily overlooked.

Kerstein's *criterial reading*, however, also faces important difficulties. First, we may question if Kant really is arguing for the supreme principle of morality in the way Kerstein suggests that he does. Nowhere does Kant say, neither suggests, that he is listing *criteria* that the supreme principle of morality must satisfy. Perhaps by itself this is not enough to discredit Kerstein's reading, but it points in that direction for the following reason: Kerstein's criterial reading makes it hard to understand the relation between *Groundwork* I and *Groundwork* II. After all, why does Kant leave the fourth criterion the second section? Does it not make the account of the first section incomplete and, therefore, unconvincing? Moreover, if Kant has three criteria already in *Groundwork* I, what is the point of beginning a whole new argument for the FUL in *Groundwork* II? More importantly, why does Kant think he needs the concept of a *rational being as such* (*vernünftiges Wesen überhaupt*) and a distinct presentation of the "practical rational faculty" (*practische Vernunftsvermögen*) (GMS, AA 04: 412) in order to explicate the concept of duty, if pretty much all the criteria have been found in *Groundwork* I? To criticisms of this kind, Kerstein replies that we find the same criteria of *Groundwork* I in Kant's derivation of the FUL in *Groundwork* II, and ultimately that his *criterial reading* is the

only alternative to the traditional interpretation which finds a gap between an uninformative principle and the FUL in both sections.

I now wish to argue for an interpretation which does not find a gap in Kant's arguments while also not adopting the view that Kant argues by way of listing criteria for the supreme principle of morality. At the same time, I believe this interpretation can better account for the differences between *Groundwork* I and II. With more detail, my view is that Kant needs the account of the first section to show that the supreme principle of morality is part of our common conception of morality. By giving a clear formulation of the principle we already, however obscurely, employ in our moral judging, he believes our commitment to morality – to the attribution of a higher value to a will which acts from duty – and our ability to make correct moral judgments will be buttressed against a "natural dialect" between the claims of duty and the inclinations. Moreover, on my reading Kant needs *Groundwork* II to justify the derivation of the FUL because it is with the concept of a rational being as such that, through a series of analysis, he arrives at the concept of autonomy. And it is this latter concept, I will now argue, that is the key for understanding that Kant moves not from PC to the FUL – since he does not hold PC as Aune reads it –, but from the idea that a good will acts out of conformity to universal law as such to a formula which can determine which maxims conform to universal law.

To see why Kant's derivation of the FUL is in fact justified, we must pay attention to two things. First, we need to remember that the FUL is, after all, a *formula*. Just like in mathematics, a formula in moral philosophy is for Kant a statement "which determines quite precisely what is to be done to solve a problem" (KpV, AA 05: 09). In a footnote in the preface to the *Critique of Practical Reason*, Kant remarks that a reviewer of the *Groundwork* thought it a deficiency of the work the fact that it did not provide a new principle of morality, but rather only a new formula. But how could it have been otherwise? After all, since "ought implies can", "It should actually have been possible to presume all along that acquaintance with what it is incumbent upon everyone, and hence also to know, would be the affair of every human being, even the commonest" (GMS, AA 04: 404). Accordingly, it is clear that in the *Groundwork* Kant is trying to give a formula that could be used "to distinguish what is good, what is evil, what conforms with duty or is contrary to it" (GMS, AA 04: 404). And more importantly, since what matters is "the mere conformity to law as such", the FUL, *qua* formula, specifies a procedure the application of which allows agents to satisfy this requirement.[48]

[48] In the next chapter I suggest that the problem agents face is that there are imperatives of practical reason which demand that agent's subjective principles (maxims) be in conformity to objective principles (laws). Accordingly,

Second: how exactly does a maxim conform to universal law? As we saw before, there is clearly a relation of conformity between maxims and practical laws, since the former are the principles on which the agent acts, and the latter the principles according to which the agent ought to act. What remains to be understood is how this conformity takes place and how we can identify it. To see how a maxim conforms to universal law, we need to remember what a practical law is: a practical principle that determines the will by virtue of its mere form (GMS, AA 04: 400). Yet what is the form of a law? For Kant, a practical law has two essential formal properties: it binds unconditionally (absolutely) and universally. We find evidence for this already in the preface when he says that

> a law, if it is to hold morally, i.e. as the ground of an obligation, must carry with it *absolute necessity*; that the command: thou shalt not lie, *does not just hold for human beings only*, as if other rational beings did not have to heed it; and so with all remaining moral laws" (GMS, AA 04: 389, emphasis added).[49]

Appropriately, he also says in the *Vigilantius* lectures that "laws of duty […] must be necessarily and universally valid" (V-MS/Vigil, AA 29: 499). Hence, a maxim conforms to universal law when its universal counterpart has the form of a law, namely when its universal counterpart can be universally and necessarily binding.[50] Now, here is one fundamental question: what makes a practical principle a moral *law*? What makes it universally and necessarily binding? We know that a *material* practical principle is always adopted because it satisfies some interest or desire for an object that the agent has. But this means that the binding force of a material practical principle is always conditional and particular: it depends on whether that specific agent has the corresponding interest or desire for the object. If the agent gives up wanting the object, the principle is no longer binding. And whoever does not have that interest or desire will also not be bound by the principle. As we will see in more detail in the next chapter, Kant calls these kinds of practical principles *heteronomous*, because what connects the will of the agent to the principle is not the will just by itself. It is rather some object of desire suggested by an inclination or some by representation of reason.[51]

---

the FUL serves as a formula because maxims that pass it (maxims that can be willed as universal laws) are shown to be in conformity to practical laws.

[49] See also the beginning of the second section, where Kant says that "the law" pertaining to "the concept of morality […] is so extensive in its significance that it must hold not merely for human beings but for all *rational beings as* such, not merely under contingent conditions and with exceptions, but with *absolute necessity;*" (GMS, AA 04: 408, orig. emphasis).

[50] See Reath (2010) and Engstrom (2009, pp. 132-4).

[51] In fact, Kant says that *all* moral theories that precede his own are heteronomous. His preferred (and easier) targets are hedonistic theories, because they make moral obligations conditional on the feeling of pleasure or displeasure: we ought to do that which brings about pleasure (happiness) and keeps displeasure (pain) away. Though he acknowledges that, by a necessity of nature, we all have happiness as an end, this end is twofold

Accordingly, here comes the crucial step for the interpretation I am proposing here: only self-legislation can make a principle universally and absolutely binding. For if the principle is not self-legislated, if it does not "spring" from the agent's own will, then it is adopted because it accords with some pathological interest (such as happiness or pleasure) or some exteriorly imposed representation of reason (such as the idea of perfection). This makes it only conditionally binding, however. The conclusion we can draw, thus, is that only autonomous willing gives a practical principle the form of a law. In other words, only principles originating from the agent's own will have the form of a practical law. Indeed, this is the moment to emphasize the importance of the first thesis that was defended: once we know that Kant is already arguing with the assumption that moral laws are self-legislated and that only self-legislation can make a principle universally and unconditionally binding, then it comes as no surprise that we find this argument when unpacking the paragraph of the derivation.

We can now appreciate how the FUL works as a formula to determine whether a maxim conforms to universal law. By demanding that the agent also be able *to will* her maxim at the same time as universal law, the FUL identifies if the maxim can be regarded as self-legislated through its universal counterpart. If it can, then it has the form of a law and, therefore, conforms to universal law. In other words, the idea that actions of a good will are in conformity to universal law can be cashed out in a moral principle such as the FUL just because the universal laws to which maxims ought to conform are self-legislated laws. And this is due to the fact that only self-legislation can make a practical principle a proper practical *law*, universally and absolutely binding.[52]

Now, this will be even clearer once we get to Kant's application of the FUL in *Groundwork* II, where he explains with more detail what happens when a maxim *cannot* be willed as a universal law. Notwithstanding, the example he provides right after the derivation of the FUL in the first section is also instructive. The question is: "may I not, when I am in trouble, make a promise with the intention not to keep it?" (GMS, AA 04: 402). Kant points out that although false promises can often be advantageous in the short-term, it might be more

indeterminate. First because we are not sure about what will make us happy; second, because people have different conceptions about what they think will make them happy. This makes practical principles that have the feeling of pleasure or displeasure as their determining ground unable to bind absolutely and universally.

[52] See Engstrom (2009, pp. 151-2): "Thus, in employing this formula one relies not only on the idea that a law to which all are subject, but also on the thought of oneself as a being who, while *subject* to that law, is also *willing* it. Indeed, in considering whether one can will one's maxim as a universal law, it is *through* conceiving of oneself as willing the law that one conceives of oneself as subject to it. This is the very thing one must be able to do in order to be a self-legislating subject, for such a subject stands only under practical laws, which according to their very idea are laws to which one cannot conceive oneself as subject except through conceiving of oneself as willing them". See also Sensen (2013, p. 272).

prudent not to make them. After all, the loss of trust from others finding out about it can be far more detrimental than what one tried to avoid in the first place by making the false promise. What is important, however, is that to make a truthful promise from a principle of honesty, that is, from duty, is quite different from making a truthful promise from fear of bad consequences. Kant says that in the former case "the concept of the action in itself already contains a law for me, whereas in the second I must first look around elsewhere to see what effects on me this might involve" (GMS, AA 04: 402). This suggests that in acting from duty one acts according to a law that makes itself present as soon as the action is considered. It is a form of autonomy, or self-legislation, since the agent does not need to "look elsewhere" to find a reason not to make a false promise, such as a the fear for bad future consequences.

In any case, this is something that I will deal with in more detail in the next chapter, where I will put this interpretation to the test when dealing with the derivation of the FUL in *Groundwork* II. For now, it is important to remind that the first section ends with Kant's claim that his analysis is in complete agreement with common moral cognition. In fact, with the analysis of the concept of a good will, we saw that a good will is made visible in beings with impure will when they act *from duty*. And the corresponding analysis of the concept of duty led us to the moral principle that common human reason "always actually has before its eyes and uses as the standard of its judging" (GMS, AA 04: 403), although it does not think of it in such an abstract form. As Guyer says:

> In Section I, Kant argues that a genuine even if less than entirely explicit understanding of the fundamental principle of morality is reflected in our common conceptions of good will and duty and in the moral judgments that we make about particular cases of human action, especially when those cases are presented to us in ways that do not immediately involve our own interests. From our common conception of good will and duty and from such particular cases, a clear formulation of the genuine principle of morality can be extracted. (2000, p. 212)

So this is what it means for a "willing to be morally good" (GMS, AA 04: 403): when we act on a maxim that can be part of universal legislation, we do it because we are conscious of the necessity of acting in this way because the action is good in itself, and not by virtue of the hopped-for effects or consequences: "A good will is good not because of what it effects, or accomplishes, not because of its fitness to attain some intended end, but good just b its willing, i.e. in itself" (GMS, AA 04: 394). This necessity is followed by the consciousness of a worth whose estimation "far outweighs any worth of what is extolled by inclination", that is, the worth of actions done for pure respect for the practical law. And this "constitutes duty, to which every other motivating ground must give away, because it is the condition of a will good *in itself*, whose worth surpasses everything" (GMS, AA 04: 403).

1.4 - Objection: autonomy in *Groundwork* I?

As I have argued throughout this chapter, the key to understading Kant's derivation of the FUL in *Groundwork* I is the acknowledgement that he is arguing with the assumption that moral laws are self-legislated laws. And I have provided textual evidence that I believe supports this thesis. Notwithstanding, Kerstein has criticized readings that find the notion of self-legislation in the first section, on grounds that autonomy "is not mentioned once in the derivation in *Groundwork* I. Its deployment belongs to the second section of that work".[53]

Now, in this quote Kerstein is criticizing Korsgaard's account of the derivation of the FUL in *Groundwork* I. Korsgaard argues from the start that respect for law is the recognition of a distinctive kind of normativity, owed only to principles that bind with unconditional force: "*Nothing* except my own will can make a law normative *for me* […] Autonomy is the only possible source of intrinsic normativity, and so of obligation" she remarks" (1996, p. 65, orig. emphasis). This finds support in Kant saying that we cannot respect the objects of inclinations, nor inclinations themselves, because a moral law has a kind of deliberative priority: it "outweighs" and "excludes" inclinations and their objects entirely "when we make a choice" (GMS, AA 04: 400). Moreover, this is because we think of objects of desire and inclinations as something that we have a passive receptivity to, whereas a moral law is a product of our nature as rational agents, as a token of the "activity of a will" (GMS, AA 04: 400).

In all fairness to Kerstein, it is true that Korsgaard provides no textual evidence for the claim that autonomy is present in *Groundwork* I. In this sense, I have dealt with this by recurring to the relevant passages from the footnote on the feeling of respect. There, Kant says very clearly: "The *object* of respect is therefore solely the law, the one that we impose upon *ourselves* and yet as in itself necessary" (GMS, AA 04: 401n, orig. emphasis) and that "as a law we subject to it, without consulting self-love; *as imposed upon us by ourselves, it is yet a consequence of our own will* […]" (GMS, AA 04: 401n, emphasis added). Sure enough, the terms 'autonomy' or 'self-legislation' are not mentioned at all in the first section of the *Groundwork*. But I believe it to be clear that the idea of self-legislation is present in Kant's account of the feeling of respect, as I have interpreted it in the first chapter.

The second kind of objection to the view I am defending here acknowledges that Kant introduces the idea of autonomy in the first section, but affirms that this introduction is

[53] Kerstein (2002, p. 206n10). For the same criticism, see Gaut and Kerstein (1999, p. 24): "Korsgaard in her interpretation appears to be trying to exploit some of the considerations on which an account of autonomy might draw to make the derivation work. But autonomy is not mentioned once in the derivation in *Groundwork* 1. Its deployment belongs to the second section of that work."

unwarranted. It claims that since the idea of autonomy is only completely explicated in the second section, Kant violated the methodological tenet of *Groundwork* I – according to which it relies solely on common moral knowledge – with his tacit introduction of the idea of self-legislation in the footnote on the feeling of respect. In commenting the failure of the derivation of the FUL, Wood says:

> Of course, once we are familiar with the later argument of the *Groundwork*, we recognize this as an introduction of the Kantian idea of autonomy (which was also hinted at in the footnote discussing respect at G 4: 401). That idea, however, is grounded on philosophical arguments we cannot evaluate at this point because we will not even get to them until well into the Second Section. Here Kant is supposed to be appealing only to ordinary rational moral cognition, which is obviously insufficient to get what he wants. (1999, pp. 48-9)

Wood's complaint is that Kant could not have used autonomy in *Groundwork* I because autonomy is not present in common moral cognition. To this objection, I offer two replies. First, it is not all that clear that the first section relies *exclusively* on the analysis of common rational cognition. The section is a "transition *from* common *to* philosophical moral cognition" (our emphasis), and so it is expected to have *some* level of philosophical moral cognition. For instance, Kant's discussion of the will "as at a crossroads" (GMS, AA 04: 400) seems to mark this transition, for there he introduces important technical terms such as *Wille, Triebfedern, Begehrungsvermögen*, and the distinction between practical and formal principles (see Rickless 2004, p. 571n17). Similarly, Paton (1947, p. 69) suggests that the transition from common to philosophical moral cognition occurs exactly at the derivation (GMS, AA 04: 402).

Second, even if we grant that the first section has no proper philosophical moral cognition and relies exclusively on common moral cognition, this does not imply that because the concept of autonomy is only fully explained and developed in *Groundwork* II, it could not be used in *Groundwork* I. Many important concepts and ideas used or hinted at in the first section receive a more cautious treatment in the second: maxim, law, practical reason, as well as the concepts mentioned right above. The same happens with autonomy, in virtue of Kant's use of the analytic (regressive) method in the first two sections: his strategy is to unpack and clarify what ordinary moral cognition already presupposes to be part of a certain *conception* of morality, namely, one that is about duty, and universally and absolutely binding laws: "*everyone must admit* that a law, if it is to hold morally, i.e. as the ground of an obligation, must carry with it absolute necessity; that the command: thou shalt not lie, does not just hold for human beings only […]" (GMS, AA 04: 389, emphasis added). The fact that the central concepts that form this conception of morality receive a more detailed treatment as the analysis progresses does not mean that they are not present from the start. Indeed, referring to *Groundwork* I, Kant

says in the beginning of the second section that "[i]t is clear from what has been said that *all moral concepts* have their seat and origin completely a priori in reason, and *indeed in the commonest human reason* […]" (GMS, AA 04: 411, emphasis added). Since autonomy clearly is a moral concept (perhaps the *most* central moral concept for Kant in the *Groundwork*), it naturally is part of the "commonest human reason".

Consequently, it is reasonable to hold that Kant already begins his analysis with the idea that moral laws are necessarily and universally binding by virtue of being self-legislated – since this is part of common moral cognition, even if not in a distinctively clear way. I concur with Reath (2006, p. 206), then, who says that "although Kant does not introduce this conception of autonomy until later in the argument of *Groundwork*, II, it is available to him as soon as he has the concept of a practical law".[54] Autonomy is, after all, found in common moral cognition.

In the next chapter, I will try to show that Kant also argues with the assumption that moral laws are self-legislated when he derives the FUL in *Groundwork* II. I believe it to be manifest in Kant's claim, *before* the derivation, that only categorical imperatives can express "a practical LAW, and that the others [i.e. hypothetical imperatives] can indeed one and all be called *principles* of the will, but not laws" (GMS, AA 04: 420, orig. capitalization). We shall see that this is because only self-legislation can account for "that necessity which we demand for a law" (GMS, AA 04: 420).

[54] See also Engstrom (2009, p. 150).

Chapter 2: The derivation of the FUL in *Groundwork* II

2.1 - Introduction

Kant makes it very clear in the preface to the *Groundwork* that the first part of his overall goal, that of identifying the supreme principle of morality, is to be carried out in *Groundwork* I and in *Groundwork* II. And this naturally leads to the question: if Kant has already identified the FUL as the supreme principle of morality in the first section (GMS, AA 04: 402), then why does he need to do it all over again in the second? What is the need for a whole new section if it going to identify as the supreme principle the same principle that has been found in the first?

Following Allison (Allison, 2011, pp. 23-29), I believe that Kant needs to offer a new argument for the FUL due to the modal status and the scope of this principle. According to common moral cognition – "which is worthy of great respect here [in *Groundwork* II]", as Kant emphasizes (GMS, AA 04: 412) – we think that moral laws are valid for all rational beings in an unconditional way: that they must necessarily do what duty commands, despite what they want with their desires and inclinations. But the fact that we conceive and experience moral demands in this way does not guarantee that moral laws actually have a universal scope and unconditional binding force. For all that matters, this could be based on just a peculiar feature of our human moral psychology, by virtue of which we are led into experiencing moral demands as backed up by universal and necessarily binding laws; however, that alone would not be enough to vindicate the claim that there are universally and necessarily binding moral laws and moral principles.

Moreover, we see at the beginning of the section that Kant also wants to offer an account of the supreme principle of morality that is completely apart from popular moral philosophy. Philosophers associated with this school endeavored to discover the principles of morality through empirical observation of human behavior. They tried to ground moral principles in a "mish-mash of gleaned observations and self-rationalizing principles" (GMS, AA 04: 409), having used "now the special function of human nature (but once in a while also the idea of a rational nature as such), now perfection, now happiness, here moral feeling, there fear of God, a bit of this and a bit of that" (GMS, AA 04: 410), and they did so in an appealing popular language, which made the theory more accessible to readers when compared to the dreary "subtleties" of a foundation for a metaphysics of morals (GMS, AA 04: 391-2). Popular moral philosophies, Kant says, are "useful for everyday chatter" but they "renounce all

thorough insight" (GMS, AA 04: 409) since they cannot account for a universally and necessarily binding principle.

For Kant, therefore, it is necessary to give the supreme principle of morality found in the first section a secure footing by deriving it from the analysis of an *a priori* concept, thus vindicating its scope and modality. He says:

> If one that, adds unless someone wants to refuse the concept of morality all truth and reference to some possible object, one cannot deny that its law is so extensive in its significance that it must hold not merely for human beings but for all *rational beings as such*, not merely under contingent conditions and with exceptions, but with *absolute necessity*; then it is clear that no experience can give occasion to infer even just the possibility of such apodictic laws. For by what right can we bring what is perhaps valid only under the contingent conditions of humanity into unlimited respect, as a universal prescription for every rational nature […] if they are merely empirical, and did not originate completely a priori from pure but practical reason? (GMS, AA 04: 408, orig. emphasis)

But which *a priori* concept should be analyzed? Kant's answer here is straightforward: if the supreme principle of morality (and the moral laws it prescribes) bind with absolute necessity every *rational* being – and not just *human* beings –, then it is reasonable to suppose that moral laws are binding by virtue of a common property that rational beings share. Thus, the concept to be analyzed is the "universal concept of a rational being as such" (*allgemeine Begriffe eines vernünftigen Wesens überhaupt*). Indeed, and more precisely, what needs to be analyzed, to be "trace[d] and distinctly present[ed]", is the "practical rational faculty [*practische Vernunftvermögen*] from its general rules of determination up to where there arises the concept of duty" (GMS, AA 04: 412). The property all rational beings share, by virtue of which they are subject to universally and absolutely binding moral laws, is that of having the faculty of practical reason.

In this chapter, I will reconstruct Kant's analysis of practical reason up to the point where he derives the FUL. We shall see that Kant identifies practical reason with a capacity to act from the representation of laws; however, these laws can determine the will in two different ways. Consequently, after distinguishing hypothetical from categorical imperatives, Kant derives the FUL from the mere concept of the latter. In other words, by analyzing what a categorical imperative is and what sort of actions it commands, Kant arrives at a formula the following of which allows agents to act on categorical imperatives – the imperatives that express moral commands. Just like in the first section, though, Kant moves from a principle that requires the mere conformity of maxims to universal law to the FUL in what appears to be an unjustified move. I will thus reconstruct the argument for the derivation and argue that it is not fallacious, since it rests on the assumption that moral laws are self-legislated laws, and this

explains how Kant can cash out the "mere conformity to universal law" requirement into the FUL. Moreover, I will show how this interpretation finds support in Kant's subsequent explanations of why there are certain maxims that we cannot will at the same time as universal laws.

2.2 - Practical reason and the will

Unsurprisingly, Kant begins the core argument of GMS II with an analysis of *practical reason*, or the *will*:

> Every thing in nature works according to laws. Only a rational being has the capacity to act *according to the representation* of laws, i.e. according to principles, or a *will*. Since *reason* is required for deriving actions from laws, the will is nothing other than practical reason. If reason determines the will without fail, then the actions of such a being that are recognized as objectively necessary are also subjectively necessary; i.e. the will is a capacity to choose *only that* which reason, independently of inclination, recognizes as practically necessary, i.e. as good. If, however, reason all by itself does not sufficiently determine the will, if it is also subject to subjective conditions (to certain incentives) that are not always in agreement with the objective ones; in a word, if the will does not *in itself* completely conform with reason (as is actually the case with human beings), then actions objectively recognized as necessary are subjectively contingent, and the determination of such a will, in conformity with objective laws, is *necessitation*; i.e. the relation of objective laws to a will of a rational being by grounds of reason, to which this will is not, however, according to its nature necessarily obedient. (GMS, AA 04: 412)

The first thing to note here is that Kant uses 'laws' in the first sentence with a different meaning than in the second sentence.[55] Everything in nature works according to laws of nature, in the sense that laws of nature are deterministic, "as laws according to which everything happens" (GMS, AA 04: 387). In this sense, even non-rational animals can be said to act on the "representation" of natural laws: when you throw a ball, the dog "knows" that it will make an elliptical path and fall, and he takes that into account when he runs for it. Kant's characterization of the will here, however, is supposed to capture what is distinctive of *rational agency*, for only rational beings have practical *reason*.[56] So 'laws' in the second sentence must refer to a different kind of laws other than natural laws.

Now, since reason in general is defined as the faculty for making syllogistic inferences (KrV A305/B362), practical reason is the faculty of making practical syllogisms, in deriving (*abtheilen*) action from the representation of laws. Yet Kant glosses acting on the representation of laws to acting on principles. And as we have seen, *maxims* are the practical

[55] See Willaschek (2006, p. 125), Timmermann (2007, p. 59-60) and Allison (2011, p. 151-3).
[56] Later, Kant defines the will as "a kind of causality of living beings *in so far as they are rational*" (GMS, AA 04: 446, emphasis added).

principles subjects act on. This suggests that 'laws' there in the second sentence actually refer to maxims. Yet considering the rest of the paragraph, this reading is untenable, for Kant is here offering a picture of rational agency that applies not only to human beings, but also to beings whose will is perfect, or holy. These beings are not subject to sensible influences, which means that they do not have inclinations[57] and, therefore, that they do not have maxims. Reason determines their will without fail: whenever they will an action, they necessarily will it in accordance with what a law prescribes; they necessarily and exclusively follow practical reason's dictates.[58] This in turn means that 'laws' in the second sentence cannot refer to maxims. Lastly, another possibility is that it might refer to moral laws. But this also cannot make sense of what comes next, for Kant says that the laws on whose representation we act are laws that necessitate and give rise, as we shall see, both to moral as well as to non-moral imperatives.

For these reasons, what remains is that 'laws' in the second sentence refers to objective practical principles *in general*, that is, to practical laws, the laws "according to which it [the rational being] *ought to act*" (GMS, AA 04: 421n). Practical laws, as we have seen, are the practical principles that every rational being would necessarily and unequivocally follow if "reason had complete control over the desiderative faculty" (GMS, AA 04: 400), which in turn means that practical laws prescribe *rational conduct*. Accordingly, beings who possess a perfect or holly will, beings who can "choose *only that* which reason, independently of inclination, recognizes as practically necessary, i.e. as good" (GMS, AA 04: 412, orig. emphasis) are *always* practically rational. In the case of human beings, however, who are "subject to subjective conditions (to certain incentives) that are not always in agreement with the objective ones", the will is objectively determined by practical laws but it can nevertheless fail to follow them: it is not by "its nature necessarily obedient" (GMS, AA 04: 412). Consequently, in these cases

[57] In GMS, AA 04: 413, Kant makes it clear that inclinations depend on the faculty of desire being sensibly affected. For the view that the holy will does have inclinations, see Willaschek (2006, p. 130) and the next footnote.
[58] In the first section (GMS, AA 04: 400n) Kant says that "if reason had complete control over the desiderative faculty" the objective principle (the law) would "also subjectively serve all rational beings as the practical principle". See GMS, AA 04: 413n and also KpV, AA 05: 79: "All three concepts, however – that of an *incentive*, of an *interest* and of a *maxim* – can only be applied to finite beings. For they all presuppose a limitation of the nature of a being, in that the subjective constitution of its choice does not of itself accord with the objective law of a practical reason […]". Willaschek (2006, p. 130ff), however, argues that this view must be incorrect because "A being that has no inclinations (and is also not subject to any other 'subjective conditions') has also, therefore, no maxims, and thus nothing to which it could apply the categorical imperative. Thus, even a being with a holy will needs inclinations (or other purely rational 'subjective conditions') in order to be able to act rationally". *Pace* Willaschek, we must recall that holy beings do not and *cannot* apply the categorical imperative because they are not subject to any kind of necessitation at all: "[…] no imperatives hold for the *divine* will and generally for a *holy* will: here the *ought* is out of place, because *willing* already of itself necessarily agrees with the law." (GMS, AA 04: 414, orig. emphasis).

practical laws not only determine but they *necessitate* the will: they *command* the agent to act in a certain way. Kant says:

> The representation of an objective principle in so far as it is necessitating for a will is called a command (of reason), and the formula of the command is called IMPERATIVE.
>
> All imperatives are expressed by an *ought*, and by this indicate the relation of an objective law of reason to a will that according to its subjective constitution is not necessarily determined by it (a necessitation). (GMS, AA 04: 413, orig. emphasis and capitalization)

Practical laws necessitate an imperfect being's will through a command (*Gebot*), and this command is issued as a formula (*Formel*) called imperative (*Imperativ*). As we have seen in the first chapter, a formula for Kant expresses a rule for the resolution of a problem.[59] Appropriately, if imperatives are formulae, this means that they dictate a certain rule or procedure the following of which allows an agent to solve a *practical* "problem". Otherwise said, it tells an agent what to do in order to satisfy some practical requirement. Since what is required by imperatives is that one conforms one's subjective practical principles to objective practical principles, we can conclude that imperatives express rules whose application allow agents to conform their maxims to laws.[60]

Moreover, all imperatives "say that to do or omit something would be good" (GMS, AA 04: 413), and what is good (*Gut*) is distinguished from what is merely agreeable (*Angenehm*). The former is objectively valid, in the sense that what is good is good for every rational being; meanwhile, the latter is subjectively valid, because it is "that which influences the will only by means of sensation […] which hold[s] only for the senses of this or that one" (GMS, AA 04: 413). Consequently, this means that all imperatives command an action because it is *rationally good*, and they command it "partly because the subject does not always know that it is good, partly because, even if he knew this, his maxims could still be opposed to the objective principles of practical reason" (GMS, AA 04: 414). Now, an important thing to note is that practical laws command actions that are not merely *morally* good, but also *instrumentally* good. They command morally good actions because we are subject to inclinations that usually lead us to actions that are contrary to duty, given that "[t]he human being feels within himself a powerful counterweight to all the commands of duty […] in his needs and inclinations […]" (GMS, AA 04: 405). Yet they also command instrumentally good actions because when we have an end given by inclination, our other inclinations might nonetheless stand in the way of

[59] See KpV, AA 05: 08n.
[60] See Timmermann (2007, p. 64).

our choosing and pursuing the best possible means to that end. To illustrate: if I want to go to Paris tomorrow and know that to do so I *must* buy a plane ticket, reason *commands* me buy it, because other inclinations or vices such as avarice and lazyness can get in the way and risk the attainment of the end that I have chosen to will. Of course, I can drop the end. But as long as I have it, I am rationally constrained to adopt the necessary means.

Indeed, Kant draws the famous distinction between two kinds of imperatives to capture the difference between laws commanding morally good actions and laws commanding instrumentally good actions:

> Because every practical law represents a possible action as good and hence, for a subject practically determinable by reason, as necessary, all imperatives are formulae for the determination of an actions necessary according to the principle of a will that is good in some way. Now, if the action would be good merely as a means to *something else*, the imperative is *hypothetical*; if the action is represented as good *in itself*, hence as a necessary in a will that in itself conforms to reason, as its principles, then it is *categorical*. (GMS, AA 04: 414)

At first sight, one might think that the distinction between hypothetical and categorical imperatives mirrors the distinction between hypothetical and categorical judgements that Kant makes in the *Critique of Pure Reason*. Roughly, this distinction is about "relations of thinking in judgments" (KrV A 73/B 98). On the one hand, categorical judgments are simple predications of the form "*x is A*", such as "The wall is blue". On the other hand, hypothetical judgments establish a relation of "ground to consequence", meaning they establish a relation of implication and are expressed in the form of *conditionals*. Kant's example is "If there is perfect justice, then obstinate evil will be punished", and he says that "[w]hether both of these propositions in themselves are true remains unsettled here. It is only the implication that is thought by means of this [hypothetical] judgment" (KrV A 73/B 98). Bringing this distinction to the *Groundwork* apparently makes sense, for Kant says that hypothetical imperatives represent something as a good means "to achieving something else that one wants" (GMS, AA 04: 414). This suggests that they are also expressed as conditionals of the form "Do *x* (means) to get y (end)". Besides, he says that categorical imperatives command an action represented as good *in itself*, with no need for any condition. This suggests that categorical imperatives have the form of categorical judgments such as "Do *x*".

Consider the following sentences, however:

(1) If you have made a promise, keep it!

(2) Buy that plane ticket to Paris!

Sentence (1) has the logical form of a conditional and so it is a hypothetical judgment, but for Kant it is obviously a categorical imperative. We ought to keep our promises

regardless of any conditions, such as if it would be beneficial or in our self-interest to keep them. Sentence (2) does not have the logical form of a conditional, so it seems to just pose a categorical (i.e. unconditional) command; still, for Kant is it is obviously a hypothetical imperative. The conclusion we can draw from this is that the distinction between hypothetical and categorical imperatives "has nothing to do with grammatical (or logical) form, since a moral imperative can be expressed in a hypothetical form and a non-moral one in a categorical form" (Allison, 2011, p. 156). It follows that the difference between hypothetical and categorical imperatives lies in *how* they command, and it is motivational in nature. Hypothetical imperatives command under *a condition*: you ought to buy that plane ticket to Paris *on the condition* that you want or that you need to go to Paris (and arrive quickly). This also means that they are binding just as long as the agent has the specific end. If the agent decides that she does not want to go Paris anymore, then she is no longer necessitated to buy the plane ticket.[61] But categorical imperatives command *unconditionally*: promises must be kept under all circumstances – setting aside possible conflict of duties –, irrespective of whether the agent wants to keep them. Consequently, this means that differently from hypothetical imperatives, there is no way to get rid of the necessitation imposed by a categorical imperative. They are unconditionally binding.

After arguing that there are two kinds of hypothetical imperatives – they are problematically practical principles when the end for which they specify an action is merely *possible* (*Möglich*); they are assertorically practical principles, instead, when the end is *actual* (or *existent*, *Wirklich*) – this is how Kant explains the difference in binding force of hypothetical and categorical imperatives:

> Willing according to these three kinds of principles is also clearly distinguished by the *unequal manner* in which they necessitate the will. Now, to make this noticeable they would, I believe, be labelled most suitably in their order by saying that they are either *rules* of skill, or *counsels* of prudence, or *commands* (*laws*) of morality. For only the *law* carries with it the concept of an *unconditional* and indeed objective and hence universally valid *necessity*, and commands are laws that must be obeyed, i.e. must be complied with even contrary to inclination. *Giving counsel* does indeed contain necessity, but it can hold only under a subjective contingent condition, if this or that human being counts this or that as belonging to his happiness, whereas the categorical imperative is limited by no condition, and as absolutely and yet practically necessary can quite properly be called a command. Imperatives of the first kind could also be called *technical* (belonging to art), the second *pragmatic* (to welfare), the third *moral* [*moralisch*] (belonging to free conduct as such, i.e. to morals [*Sitten*]). (GMS, AA 04: 416-7, orig. emphasis)

[61] Though if she has *promised* someone that she would go to Paris, then she is still necessitated to go to Paris (through a categorical imperative, in this case) and so to will the necessary means.

Accordingly, only categorical imperatives express *laws* because only this kind of imperative binds unconditionally and universally. The question Kant immediately turns to, then, is: how are all of these imperatives possible? He has explained that there is a difference to their biding force, but he is yet to explain *how* they bind. As he himself warns the reader, this question does not mean "how can we execute the actions commanded by imperatives?", as in "how do I buy a plane ticket to Paris?", to use again one of our examples. The problem Kant wants to deal with is understanding how the *necessitation* of imperatives is possible. He wants to know how come "if I want to go to Paris" then I *ought* also "to want to buy a plane ticket to Paris" and how come "If I made a promise" then I *ought* also "to keep it regardless of whether I want to". In other words, how can we be bound by these *oughts*?

Kant quickly explains how imperatives of skill are possible. He says:

> How an imperative of skill is possible probably requires no special discussion. Whoever wills the end also wills (in so far as reason has decisive influence on his actions) the indispensable means to it that is in his control. As far as willing is concerned, this proposition is analytic; for in willing of an object, as my effect, my causality is already thought, as an acting cause [*denn in dem Wollen eines Objectes, als meiner Wirkung, wird schon meiner Caußalität, als handelne Ursache ... gedacht*], i.e. the use of means, and the imperative already extracts the concept of actions necessary to this end from the concept of a willing of this end (synthetic propositions are certainly needed to determined the means themselves to an intended purpose, but they concern the ground for actualizing not the act of will, but the object). (GMS, AA 04: 417)

As Kant emphasizes, we need synthetic propositions to determine *what the means are* for actualizing (*wirklich zu machen*) an end. If I will to be in Paris tomorrow, that does not mean that I know analytically that I ought to call Air France. (We naturally need some empirical input when determining what are the best or the only means available, such as knowledge that planes can transport people between continents overnight). Instead, if I will to be in Paris I do analytically will "the indispensable means" (namely in this case buying a plane ticket) to go to Paris, as far as I am rational. Why? Well, first, because when an I will an end – and not merely wish for it[62] – I am under the constrains of rationality, for recall that the capacity to will something belongs exclusively to rational beings (GMS, AA 04: 412). Second, when I will an end, I conceive of it as something whose existence *I* cause: *my will* is thought of as an acting causality (*handelne Ursache*). It is *my will* what makes the end real, in a way that the end is regarded of as an *effect of my will* (*Als* ***meiner*** *Wirkung*). Third, if my will functions as an acting cause, it must do so through the "use of means". In other words, I realize an end by using or

[62] For Kant, wishing for something means that one has a desire for some end or object, but does not strive to bring it about – one is complacent. Willing something, in turn, means one is committed to using everything in one's power to bring about the desired end. For this distinction see MS, AA 06: 213.

adopting whatever indispensable means are available to me ("the indispensable means to it that *is in his control*", Kant says at 4: 417). There is *no other way* of realizing an end (as something that *my causality causes*) other than adopting the necessary means. It is for this reason that Kant says that from the idea of someone "willing an end" it follows analytically that as long as she is being rational she "wills the necessary means". Surely, this does not mean that whenever we will an end, we actually *do* will the necessary means. There is the possibility of failure since we are not always fully rational.[63] But it does mean that whenever we set an end and will it, the hypothetical imperative of skill binds us to choose the necessary means to its achievement.[64]

Before we turn to categorical imperatives, it is worth discussing Kant's talk of analyticity here. He says: "Whoever wills the end also wills (in so far as reason has decisive influence on his actions) the indispensable means to it that is in his control. As far as willing is concerned, this proposition is analytic" (GMS, AA 04: 417). In what sense is this proposition analytic? First, I believe it is analytic in the most immediate Kantian sense of the term: as which predicates are contained in a subject. To "will the means" is *part of the concept of* "willing an end". The "as far as willing is concern" proviso merely says that what is analytic is the *willing* of necessary ends, not that the ends are discovered analytic: I need empirical data to know that traveling by plane is a means to get to Paris overnight. Second, hypothetical imperatives are analytic in the sense that the action they enjoin is thought to be already contained in the will of the agent. When one wills *x*, one *already wills* indispensable means *y*.

What about the possibility of categorical imperatives? How are they possible? Are they also analytic? Kant warns us again, as he did at the beginning of the section, that we cannot discover this by an appeal to experience. As much morally worthy and self-disinterested an action might appear to be, we cannot really know what the agent's motivating ground was: whether it was from duty and respect for law that someone refrained from making a deceitful promise, or if it was from a "fear of embarrassment", hence from a condition, that might have "covertly influence[d] the will" (GMS, AA 04: 419). What is more, this cannot be solved by an appeal to the first-person. As much confident one can be that one did act *from duty*, that does not guarantee that there was not "actually a covert impulse of self-love under the mere pretence

[63] Fully rational beings necessarily will the ends to their actions because in them "reason has decisive control over the desiderative faculty" (GMS, AA 04: 400n). Since we human beings are not fully rational, it is possible that we will an end and fail to will the necessary means. That is why we are *commanded* to do so.

[64] Imperatives of prudence "would totally and entirely coincide with those of skill, and be equally analytic, if only it were so easy to provide a determinate concept of happiness" (GMS, AA 04: 417). The problem is that happiness is an indeterminate concept, meaning we do not exactly know what will make us happy. In other words, for Kant we cannot clearly conceive and end for imperatives of prudence, and without a clearly determinate end, we also cannot conceive (and will) the necessary means. For more on this see notes 74 and 75.

of that idea […] we can never, even by the most strenuous examination, get entirely behind our covert incentives" (GMS, AA 04: 407). All of this indicates that the problem with understanding how categorical imperatives are possible lies in the fact that they are *synthetic* practical propositions. Whereas when one wills an end one also analytically wills its necessary means (in so far as one is rational), in the case of categorical imperatives "the willing of an action" is connected immediately and *a priori* (without the need for a sensible desire) with "the concept of the will" without one willing any other action or end (GMS, AA 04: 420n). In other words, the willing of the action "do not lie" is connected to the will of the agent without any presupposed conditions such as "if you do not want to get caught". Categorical imperatives express synthetic propositions in the sense that the action is not something the agent necessarily wishes to do: it is not previously contained in his willing, so to speak. And this leads Kant to postpone the answer to the *How are categorical imperatives possible?* question to GMS III, since it will "require particular and arduous effort" (GMS, AA 04: 420).

For the moment, Kant wants to see "whether the mere concept of a categorical imperative may perhaps also furnish its formula, which contains the proposition that alone can be a categorical imperative" (GMS, AA 04: 420). This marks the beginning of the derivation of the FUL in the second section.

2.3 - The derivation of the FUL

Let us summarize what Kant has argued for in the second section until this point. First, we know that rational beings have a faculty for deriving actions from practical laws, that is, they have *practical reason* or a *will*. Second, we know that there are two kinds of rational beings when it comes to how practical reason operates: on the one hand, perfect rational beings are unaffected by inclinations, which means they necessarily abide by practical laws since nothing hinders them from doing so; on the other hand, imperfect rational beings like ourselves are *necessitated* by practical laws, since our maxims are not necessary in conformity to practical laws.

Furthermore, there are two kinds of necessitation, which get captured in the distinction between hypothetical and categorical imperatives. The former kind of imperatives commands an action on the condition that it is a necessary means to an end. The latter commands an action unconditionally. And it is precisely this kind of imperative that captures the distinct way in which moral laws determine the will. As something that common moral cognition already recognizes, moral laws bind universally and unconditionally: we think that

the prohibition on lying, for example, holds for everyone on every occasion (see GMS, AA 04: 389). Since only categorical imperatives bind universally and unconditionally, only this kind of imperative "may be called that OF MORALITY" (GMS, AA 04: 416, orig. capitalization) and only this kind of imperative expresses practical *laws* that can ground duties. As Reath (2006, p. 132) says:

> Kant is led to the concept of a practical law through the need to ground conclusions about duty. The claim that an agent has a duty to act in a certain way holds that there are reasons for action that apply with necessity (that apply independently of and limit the reasons given by the agent's desires) and universality (that anyone can regard as valid). Since conclusions about how one ought to act are derived from practical principles, claims about duty must be traced to a kind of practical principle that can ground their normative force.

Appropriately, since Kant wants to leave the question of the possibility of categorical imperatives to GMS III, he will therefore analyze the concept of a categorical imperative in order to see if it is possible to extract from it its *formula*. In other words, keeping in mind: first, that the goal of GMS II is to identify the supreme principle of morality and, second, that categorical imperatives are the imperatives of morality, Kant will try to extract the supreme principle of morality from the concept of a categorical imperative – just like he derived the principle of duty from the concept of duty in GMS I. He says:

> When I think of a *hypothetical* imperative as such I do not know in advance what it will contain, until I am given the condition. But when I think of a *categorical* imperative I know at once what it contains. For since besides the law the imperative contains only the necessity of the maxim to conform with this law, whereas the law contains no condition to which it was limited, nothing is left but the universality of a law as such, with which the maxim of the action ought to conform, and it is this conformity alone that the imperative actually represents as necessary.
>
> There is therefore only a single categorical imperative, and it is this: *act only according to that maxim through which you can at the same time will that it become a universal law*. (GMS, AA 04: 420-1, orig. emphasis)

In the first sentence, Kant explains that until an end or purpose is given (i.e. the "condition"), one cannot know what action a hypothetical imperative will enjoin. This is because hypothetical imperatives "represent the practical necessity of a possible action as a means to achieving *something else that one wants*" (GMS, AA 04: 414, emphasis added). All we know is that it will enjoin whatever is deemed necessary to achieve the end, but without the end, we do not (and cannot) know specifically what it will command. Kant thinks the situation is different with regard to categorical imperatives, since they represent "an action as objectively necessary by itself, without reference to *another* end" (*ohne Beziehung auf einen andern Zweck*, GMS, AA 04: 414, emphasis added). This means that, contrary to what Kant's contrast between hypothetical and categorical imperatives might apparently suggest, categorical imperatives do

command the attainment of some end or purpose.[65] The difference, as the "without reference to *another* end" proviso suggests, is that in a categorical imperative the end to be attained is judged to be practically good in itself. The agent wills to achieve it because it is in itself good, and not good for some other *further* end. Consider again the beneficence example from the first section. On the one hand, when the man helps from duty he wills the purpose of helping others without any other further end in mind, such as the benefits or pleasure that that could give him. On the other hand, when he helps from immediate inclination he values the action of helping others, for that furthers another one of his purposes, namely the inner gratification he feels when he helps. Kant makes this clear when he says that:

> In the first case the action interests me, in the second the object of the action (in so far as it is agreeable to me). We saw in the first section: that an action from duty must pay attention not to the interest in the object, but merely to that in the action itself and in its principle in reason (the law)." (GMS, AA 04: 414n)

So, maxims that can be expressed by categorical imperatives differ from maxims that can only be expressed by hypothetical imperatives not in that they do not have an end, but because the end (purpose) they contain makes no exclusive appeal to peculiar conditions of the subject. As such, it can be adopted by anyone, universally. Accordingly, continuing the exegesis of the derivation paragraph, we see in the third sentence that a categorical imperative necessarily requires the conformity of a maxim to a law. This is unsurprising, since *every* imperative requires the conformity of a maxim to a law: every imperative is a command of reason directed to a being with an imperfect will who does not necessarily conform its subjective principles (maxims) to objective principles (laws). Kant makes this point clear by adding a footnote in that page explaining that maxims are principles according to which subjects *act*, whereas laws are principles according to which subjects *ought* to act. (GMS, AA 04: 420n).

But to what kind of law should the maxim conform? Since it is a categorical imperative, it cannot be a law whose ground of adoption is the realization of some object or end that is of contingent interest to the agent. If it were, the binding would be conditional, giving rise only to a hypothetical imperative. It must, therefore, be a law which has no agent-dependent condition on its adoption. For example, a law commanding "If you made a promise, you ought to keep it, but only if you want to" relies on the agent's wants and desires. It has what Kant

[65] Categorical imperatives in fact require that we adopt some ends. First, it requires that we treat the rational being as an end in itself – not as some end to be achieved (*ein zu bewirkender Zweck*), but rather as an already existent end (*selbstständiger Zweck*), and so as something that must be preserved (see GMS, AA 04: 437). Second, it requires that we adopt the ends of virtue, e.g. beneficence, gratitude, sympathy, etc. (MS: 06: 452).

calls a *material* (i.e. related to the agent's interests[66]) condition, and it cannot be part of an unconditional command. A law commanding "If you made a promise, you ought to keep it" on the other hand, commands the keeping of a promise to *everyone* who has made one, regardless of their wants and desires. In this sense, it is *universally* valid. This is why Kant says that once "the law contains no condition to which it was limited, nothing is left but the universality of a law as such". In other words, when it comes to categorical imperatives, the law to which the maxim should conform has no *material* content (since that would prevent it from being unconditionally commanded), but only the form of a law, a *lawfulness,* which Kant equates with universality. This means that the maxim, to be in conformity to the law, must possess *the form of a law*: it must be a universally valid principle, a principle any rational agent, when in the relevant circumstances, could adopt. It is "this conformity alone that the imperative represents as necessary".[67] As Herman points out:

> This is what would be for my maxim to conform to an unconditioned principle of willing […] If the imperative commands that I act on a maxim that is rational for me on grounds valid for all rational beings, then it is in effect commanding that I act on a *universally valid maxim* – that the maxim of my action conform to 'law'. Not that my maxim should conform to *a* law, specified in advance, but that my maxim be *law-like*: not merely a subjective principle tailored to my interests, and so 'valid' only as I (or my other agent) have such interests […] Thus we see how the requirement emerges that a maxim have the form of a law, the form of universality, and thus what it is about categorical imperatives that leads Kant to say it requires of a maxim that it conform to the universality of a law as such and that this conformity alone is what is represented as necessary by the imperative. (2016, p. 169-70, orig. emphasis)

Let us recall that, according to the same critics of the derivation from GMS I, this is where the argument goes wrong. Kant has concluded that what any categorical imperative requires is that the agent's maxim be universally valid, that it be lawful: "the universality of a law as such, with which the maxim of the action ought to conform". In other words, all that agents must do is adopt maxims that conform to universal laws, a requirement that we expressed in the introduction in the following principle:

(Principle of Conformity - PC): Act on maxims that conform to universal law

Nonetheless, this requirement is uninformative. It says that we should act on maxims that conform to universal law, but it gives us no means to identify which laws there are. How are we supposed to know whether there is a law that commands, for instance, "help

[66] See GMS, AA 04: 427: "Practical principles are *formal* if they abstract from all subjective ends; they are *material* if they have these, and hence certain incentives, at their foundation."
[67] See Reath (2006, p. 90-1n): "It [the Categorical Imperative] states the requirement of acting from reasons that satisfy the criteria of necessity and universality, or are sufficient to justify one's actions fully to anyone. So the supreme practical law says: act from maxims that have the form of universal law."

those who need assistance, to the best of your ability"? For all that matters, "*do not* help those who need assistance" could be a law, and with what Kant has given us so far, there is no way of knowing it. Yet just like in the first section, Kant immediately proposes another principle as the actual formula of the categorical imperative. He says that there is "only a single categorical imperative, and it is this", such that one is *not only* required to act on maxims that conform to universal law, but one actually must act on maxims one can will as universal law at the same time:

> (**Formula of Universal Law - FUL**): act only according to that maxim through which you can at the same time will that it become a universal law

As critics have pointed out, the problem is that this reformulation of the requirement appears to be illegitimate. It seems to be a subtle move from a warranted yet uninformative principle to one that is properly action-guiding, yet unjustified. Why must we be able *to will* that our maxim at the same time as a universal law?

Now, as in the first section, the first step is to acknowledge that Kant does not hold PC as some commentators would want him to. The "universality of a law as such, with which the maxim of the action ought to conform" is a statement about the nature of action done on a categorical imperative, not the statement of an action-guiding principle. Consequently, the task is not to try to fill a gap between an uninformative principle (PC) and an informative one (the FUL). Rather, it is to understand how Kant can provide a principle, the application of which lets agents check whether their maxims conform to universal law, from the realization that in categorical imperatives all that matters is the universality of a law as such.

The first step is to understand how does a maxim conform to universal law. Now, since in categorical imperatives the law to which the maxim should conform has no material content, all there is left for the maxim to conform is to the law's form. Consequently, a maxim conforms to universal law if it can have the form of a law. This leads to next question: what is then the form of a law? In effect, consider the following passage:

> the categorical imperative alone expresses a practical LAW, and that the others can indeed one and all be called *principles* of the will, but not laws; since what is necessary to do merely for attaining a discretionary purpose can be regarded as in itself contingent, and we can always be rid of the prescription if we give up the purpose, whereas the unconditional command leaves the will no free discretion with regard to the opposite, and hence alone carries with it that necessity which we demand for a law. (GMS, AA 04: 420, orig. emphasis and capitalization)

Kant is here making a point to which I previously called for attention: a practical principle is a practical *law* only when it commands unconditionally. Since hypothetical

imperatives always command on a condition, they do not express practical laws: only mere *recommendations* or mere *precepts*.[68] Importantly, then, categorical imperatives express practical laws that have as one of their properties being *necessarily* binding. The other essential property of a practical law is its being *universally* binding. Kant has said this already at the beginning of the second section when he explains that "the law" pertaining to "the concept of morality […] is so extensive in its significance that it must hold not merely for human beings but for all *rational beings as* such, not merely under contingent conditions and with exceptions, but with *absolute necessity*" (GMS, AA 04: 408, orig. emphasis). We also see this at the very paragraph of the derivation, for Kant says that once we get rid of a law's material content, then "nothing is left but the universality of a law as such"; moreover, we have seen that for Kant a conception of morality as involving necessary and universally valid laws is present in common moral cognition (GMS, AA 04: 389). For this reason, we can conclude that *necessity* and *universality* constitute the form of a law.

At this juncture, there is another important question: after all, what gives a practical principle necessity and universality, thus making it a law? To answer this, it is perhaps more suitable to see why hypothetical imperatives do not express practical laws. As we have seen, hypothetical imperatives enjoin an action has a means to something else – as such, they are conditional commands, because the subject will want the means only as long as she wants the end. Since it is always possible to get rid of the end, the imperative expresses a principle that is not necessarily binding; besides, since not every subject has the same ends, it is also not universally binding. Now, Kant says that all previous moral theories could only account for hypothetical imperatives. He calls them *heteronomous*, since what connects the will of the agent to the principle – what makes the agent follow the principle – is something outside the will itself: it is some object of desire suggested by an inclination or a rational concept (such as the idea of perfection). Heteronomous theories reduce moral commands to mere recommendations: one ought to do something only to the extent that it will bring something else about: happiness, the satisfaction of a moral sense, and so on.

Kant distinguishes two kinds of heteronomous principles. The first kind he calls *empirical* principles, saying that they are all based on the principle of *happiness* and, therefore,

[68] See KpV, AA 05: 20: "But imperatives themselves, when they are conditional – that is, when they do not determine the will simply as will but only with respect to a desired effect, that is, when they are hypothetical imperatives – are indeed *precepts* but not *laws*. The latter must sufficiently determine the will as will even before I ask whether I have the ability required for a desired effect or what I am to do in order to produce it, it must thus be categorical: otherwise they are not laws because they lack the necessity which, if it is to be practical, must be independent of conditions that are pathological and therefore only contingently connected with the will."

"not fit for the foundation of moral laws at all" (GMS, AA 04: 442). In fact, there is a subdivision here: empirical principles can be grounded either on a physical or on moral feeling. Principles grounded on a physical feeling command an action on the condition that it satisfies an inclination; thus, their binding is dependent on "the *particular arrangement of human nature*, or the contingent circumstances in which" one is placed (GMS, AA 04: 442, orig. emphasis). Though Kant acknowledges that, by a necessity of nature, we all have happiness as an end, this end is twofold indeterminate. First, we are not sure about what will make us happy;[69] second, because people have different conceptions about what they think will make them happy.[70] This means that practical principles that have inclinations as their determining ground can only furnish hypothetical imperatives, hence they do not bind absolutely and universally.[71]

Kant is particularly at odds with *ethical egoism*, i.e. the theory according to which agents ought morally to do something only if it advances their self-interest.[72] He calls this the principle of "one's own happiness", and in the *Mrongovious* lectures he explains that it is an "utterly false" principle mainly for three reasons: (1) not every morally virtuous person is happy; (2) virtue gives the person a sense of inner worth, but that is different from happiness (*pace* the Stoics). In fact, since morality commonly requires that we give up the satisfaction of many inclinations, it is often the case that many virtuous persons are indeed *unhappy*; (3) it contradicts common sense morality, for everyone sees a difference between doing something because it is right (from duty) and doing it because it is prudent or advantageous. Kant remarks

---

[69] See GMS, AA 04: 418: "Now, it is impossible that the most insightful and at the same time singularly able, but still finite being should make for himself a determinate concept of what he actually wants here. If he wants riches, how much worry, envy and intrigue might he not by this bring down upon his shoulders! If he wants much cognition and insight, that might perhaps only sharpen his eyes all the more, to show him as all the more terrible the ills that are still concealed from him now and yet cannot be avoided, or to burden his desires, which already give him enough trouble, with more needs still. If he wants a long life, who will guarantee him that it would not be a long misery? If at least he wants health, how often has not bodily discomfort kept someone from excess in which unlimited health would have plunged him, and so on. In short, he is not able to determine with complete certainty, according to any principle, what will make him truly happy, because omniscience would be required for this". Here Kant is also suggesting that the "principle of happiness" is tautological when used as a *principium diudicationis*.

[70] See KpV, AA 05: 25: "That is to say, in what each has to put his happiness comes down to the particular feeling of pleasure and displeasure in each and, even within one and the same subject, to needs that differ as his feeling changes; and a law that is *subjectively necessary* (as a law of nature) is thus *objectively* a very *contingent* practical principle, which can and must be very different in different subjects, and hence can never yield a law because, in the desire for happiness, it is not the form of lawfulness that counts but simply the matter, namely whether I am to expect satisfaction from following the law, and how much."

[71] See V-Mo/Mron II, AA 29: 599: "Morality cannot be construed out of empirical principles, for this yields, not absolute, but merely conditional necessity. Morality says, however, you must do it, without any condition or exception."

[72] See also KpV, AA 05: 35-38.

that "[i]f virtue is always followed by happiness, it then has a price, and if vice then rested upon more advantages, we would have to prefer it to virtue" (V-Mo/Mron II, AA 29: 624).

Principles grounded on moral feeling, accordingly, also rely on the contingent constitution of human beings. They depend on the peculiar receptivity of the agent to certain feelings such as sympathy, compassion, etc. This means that the agent is only obligated to, say, help someone, only if she feels a correspondent feeling. It also means that whoever feels attracted to vice has the obligation to engage in it, since in the sentimentalist picture of moral obligation, moral laws are "merely arbitrary, and simply a childish game." (V-Mo/Mron II, AA 29: 625). As we saw in the first chapter with the discussion of moral worth, feelings for Kant are morally blind: an immediate inclination towards beneficence, for example, if taken as the sole determining reason for action, may very well lead to *immoral* action.[73] Therefore, moral feeling also can only yield hypothetical imperatives, for one is obligated to act only if one experiences the necessary feeling (and so the obligation is conditional) – and because not everyone "has the same feeling" (V-Mo/Mron II, AA 29: 625) (and so the obligation is not universal).

Now, opposed to empirical heteronomous principles there are rational heteronomous principles. Yet for Kant the latter kind is also not fit for giving categorical imperatives. In this case, he distinguishes between grounding morality either on the concept of *perfection* (*Vollkommenheit*) (as something that can be brought to existence as an effect of our will), or, on the concept of a "divine all-perfect will" (*göttliche allervollkommenste Wille*). The first kind of principle is that which is advocated by Wolff and Baumgarten,[74] and it is rejected by Kant because of its indeterminacy. Evoking his previous charge against Baumgarten's tautological moral principles, Kant explains that any principle based on the concept of *perfection* will issue prescriptions such as "make yourself perfect" or "seek all the perfections in your person" (V-Mo/Mron II, AA 29: 626). But if one understands perfection to mean that which is morally good in itself, then one has a tautological command, since obviously any moral principle will command what one seeks all that is morally good. If one understands it to mean

[73] In a footnote, Kant justifies his listing sentimentalist theories (mentioning Hutcheson) together with hedonistic ones explaining that both rest on the interest for our own well-being. In the former case, it needs to concern our own self-interest – as the man who helps others not out of self-interest but from immediate inclination to beneficence. In the latter, self-interest and personal advantages are always the determining factor for the agent taking an interest in the action (GMS, AA 04: 442n).
He explains that while in latter what concerns the agent is the feeling of pleasure or satisfaction concerning his own well-being and in the former it is the

[74] See Allison (2011, p. 268). Indeed, as we saw in the introduction, one of Baumgarten's moral principles is "*Quaere perfectionem, quantum potes*" (*seek perfection as much as you can*). See note 1 for Kant's reasons for dismissing it as a supreme principle of morality.

the best mean to an end, however, it will obviously yield only hypothetical imperatives. The second kind of rational principle is rejected because it faces a dilemma: one either obeys God because (i) God commands that which is right and good, or (ii) because what God commands is right and good. If one opts for the first horn, this means that one obeys God because one has a prior concept or standard of moral goodness that is independent of what God wills (on pain of circularity), and one sees that God's willings conform to this standard. If one opts for the second horn, moral laws "would be *statuta*, having no power to bind on their own account, but acquitting it through the will of another" (V-Mo/Mron II, AA 29: 627, orig. emphasis). Again, one would have crude heteronomy: one would obey God's laws not because they are the laws of duty, but only because God has ordained it and one fears God's punishments.[75]

Appropriately, theories that ground morality either in empirical or in rational principles are such that they give the will a law that has "to carry with it some interest as stimulation or constraint, because it did not as a law arise from *his* [i.e. the human being's] will, which instead was necessitated by *something else*, in conformity with a law, to act in a certain way" (GMS, AA 04: 433) and this "makes possible hypothetical imperatives only: I ought to do something *because I want something else*" (GMS, AA 04: 441). Whenever the will is determined by a law given by our sensuous nature – through inclinations or moral sense – or by the idea of some other being's will, we have conditional commands that cannot be moral commands. For there to be categorical imperatives, the will must be determined by a law that does not condition the action to the realization of some interest, since "the dissociation from all interest in willing from duty" is the "specific mark distinguishing categorical from hypothetical imperatives" (GMS, AA 04: 431) and "duty […] can be expressed only in categorical imperatives, but by no means in hypothetical ones" (GMS, AA 04: 425). Since the will must be determined by a law,[76] yet not by any law that is externally imposed, it can only be that in the case of categorical imperatives the will is determined by a law that it gives itself. In other words, the will must be autonomous for there to be categorical imperatives.

---

[75] Kant says in one of the *Lectures in Ethics*: "If, however, we are to carry out His commands because He has ordained it, and because He is so mighty that He can make us do it by force, then we act under orders, from fear and terror, and simply fail to perceive the justice of the injunction; nor do we know why we ought to do what God has commanded, and why we should be obedient to him; for the *vis obliganti* cannot consist in force, since one who threatens does not obligate, but extorts. So if we are to abide by the moral law out of fear for God's punishment and power, and this because it has no other ground than that God has commanded it, then we do so not from duty and obligation, but from fear and terror, though that does not better the heart" (V-Mo/Mron, AA 27: 1426). Naturally, Kant is engaging here with the Euthyphro dilemma. And he opts for the first horn: "So an action must be done, not because God wills it, but because it is righteous or good in itself; and it is because of this that God wills it and demands it of us" (V-Mo/Collins, AA 27: 262; 278).

[76] GMS, AA 04: 412 and GMS, AA 04: 446.

Appropriately, this much is enough for the reconstruction of the derivation of the FUL. For now we know that if categorical imperatives are possible only if the will gives itself the law, this means that *only autonomous willing can give a practical principle the form of a law*. In other words, only principles that arise (*entspringen*) from the agent's own will have the form of a practical law. And with this we can understand the derivation of the FUL. First, it ought to be emphasized again that the FUL is a *formula*, which means that it specifies a certain procedure the following of which gives the subject the solution to a problem. The problem that agents face is to know whether their proposed maxim conforms to universal law, because it is this conformity alone that a categorical imperative requires.[77] Second, we have seen that a maxim conforms to universal law when it has the form of a law, which means that a maxim conforms to universal law when it can be universally and necessarily binding. Third, we have just concluded with Kant that only if a law is self-legislated can it bind categorically, hence necessarily and universally.[78] Therefore, by asking agents to consider whether their proposed maxim can be *willed at the same time as universal law*, the FUL assesses whether the proposed maxim can have the *form of a law*. If it can, then it conforms to universal law and can be expressed in a categorical imperative, since this conformity to universal law is the only thing that is demanded.

Perhaps it is easier to comprehend how the test proposed by the FUL works by paying attention to what happens when a maxim *cannot* be willed as a universal law. Right after deriving the FUL, Kant says that since the formal concept of nature signifies the existence of things according to universal laws, the FUL can also be expressed as "so act as if the maxim of your action were to become by your will a universal law of nature" (GMS, AA 04: 421). Accordingly, he applies this variation of the formula to four different cases. The first two cases are such that the maxim cannot be willed as universal law because it simply *cannot be conceived to hold* as universal law in the first place. This is the case with the maxim of making a false promise to repay a loan. The agent "knows full well that he will not be able to repay, but also sees that nothing will be lent to him unless he solemnly promises to repay it at a determinate

[77] I take this idea to be present in this remark of Herman (2016, p. 276): "From the fact that the Categorical Imperative rules on maxims we can see that its primary function as a principle of judgment is to forbid or allow *proposed* actions rather than to prescribe one among a number of alternative possible actions. It is not to be employed as one might employ, for example, the principle of utility. The question is not, 'What ought I to do?' but rather, 'Is what I propose to do morally right (or permissible)?' That is, does what I propose to do violate the principle of morality–the Categorical Imperative."

[78] Setting aside for now how one is supposed to interpret the way in which agents can be said to "legislate" a practical law I will briefly address this at the end of this section), Reath (2006, p. 127) remarks that "since a practical law applies unconditionally, its authority cannot be based on an agent's desires, or any contingent or empirically given interests. Instead, it comes from the fact that the agent is its legislator: only if the agent subject to law is its legislator will its authority be independent of any empirically given interest, and so unconditional."

time" (GMS, AA 04: 422). He considers adopting the following maxim: "when I believe myself to be in need of money I shall borrow money, and promise to repay it, even though I know that it will never happen" (GMS, AA 04: 422). Now, to see if this proposed maxim conforms to universal law, the agent tries to will it as such. And he sees that this is impossible, because such a maxim of making false promises cannot even be conceived as holding as a law in the first place. This is because the analogy with laws of nature implies that the universalization of the maxim, and the consequences thereof, would be a matter of common knowledge between agents. Everyone would feel tempted to make false promises when in need of ready cash, and everyone would *know* that everyone else considers making false promises when in financial need. Consequently, this would eliminate all trust between people. Yet trust between the promisor and the promisee is a necessary condition for the practice of promise-making. Hence, if made universal law, this maxim "would make the promise and the end one may pursue with it itself impossible" because "no one would believe he was being promised anything" (GMS, AA 04: 422). Now, if the agent knows that this would be the case if such a maxim were universalized – which is implied by the law of nature analogy[79] – that is, if he knows that it would be impossible to get a loan, and yet *still* and at the same time wills this maxim, then he wills a contradiction: he wills that no one gets a loan (*a fortiori* by virtue of the universalization of the maxim) while willing that he gets one. Kant says that "[s]ome actions are such that their maxim cannot even be *thought* without a contradiction as a universal law of nature; let alone that one could *will* that it *should* become such" (GMS, AA 04: 424, orig. emphasis).

In the second kind of cases we have a different way to fail the test proposed by the FUL. It happens when the maxim can hold as a universal law of nature, but it is impossible to will it and its universal law counterpart simultaneously. Kant's example is about a man who does well for himself but sees that other people need help. Indeed, he plainly recognizes that he "could just as well help", but he does not. Although Kant does not say what is the maxim in this case, the man's way of thinking (*Denkungsart*) is that of adopting an attitude of general indifference. He thinks: "what's it to me? May everyone be as happy as heaven wills, or as he can make himself, I shall take nothing away from him, not even envy him; I just do not feel like contributing anything to his well-being, or his assistance in need!" (GMS, AA 04: 423). Now, as the passage shows, it is important to get it clear that this is not someone who wishes or does ill to others. It is actually someone who just does not care how others are doing, and who does not want to be bothered with having to do favors or providing help of any kind – "every man

[79] See Rawls (2000, p. 169) and also Velleman (2005b).

for himself", he thinks. Now, Kant does say that this way of thinking could hold as a universal law of nature. This way of behaving would not "destroy itself" if universalized – as in the case of false promising. All there would be is a world where everyone is indifferent to everyone else's needs. It could be a very miserable place, but "the human race could very well subsist" and indeed, Kant says, at least there would be no more cynics who occasionally help others and talk highly of "compassion and benevolence" in public, but who covertly cheat whenever possible. Yet Kant says:

> But even though it is possible that a universal law of nature could very well subsist according to that maxim, it is still impossible to WILL that such a principle hold everywhere as a law of nature. For a will that resolved upon this would conflict with itself, as many cases can yet to come to pass in which one needs the love and compassion of others, and in which, by such a law of nature sprung from his own will, he would rob himself of all hope of the assistance he wishes for himself. (GMS, AA 04: 423, orig. capitalization)

As this shows, the man of the example is an egoist. He does not want to help others, and he does not want their help because*, at the moment*, he is someone "who is prospering". Yet he knows that given the circumstances of life, it is possible (perhaps likely) that he will need the help of others; in fact, if a situation were "to come to pass" in which he needed the aid of others, he wishes that they assist him. The contradiction in his will arises because by willing a maxim of general indifference as a universal law he is willing (1) a world where no one is ever helped – he is willing a world in which everyone may just say "what's it to me?" [*was gehts mich an?*] when seeing someone in distress; nonetheless, at the same time he (2) wills that others help him if he needs it. Consequently, he wills that *no one* (via [1]) and that *someone* (himself, via [2]) receives help.

So there is a contradiction in the will of the agent because while he denies help, he wants to be helped if he needs it. But what if he does not? Surely, we can imagine someone who does not want to help, yet who also does not want to be helped at all. Being committed to a thoroughgoing Stoic ideal of self-sufficiency, this hypothetical man decides to only pursue ends that he is confident he can achieve on his own, or ends that he can easily let go of if they are hard to achieve. How can we get a contradiction in the will out of this? I believe that to push Kant's argument for the duty of beneficence (as grounded on the FUL) to work in this case will give us an important insight into the derivation of the formula.

Now, for it to work in this case,[80] Kant needs another argument for the conclusion that every agent *necessarily wills* to receive the assistance of others. Only then we will always have a contradiction in the will in the case of willing a maxim of non-beneficence. Appropriately, the solution comes from Kant's account of our "true needs" as human beings: he argues that some ends are necessary for the continuation of our existence as rational beings, and that, therefore, it is irrational to forgo of them. A most immediate example is the end of obtaining basic sustenance for oneself.[81] This end constitutes a true need in the sense that without basic sustenance we cease to exist as rational, and at some point, as living beings.

Accordingly, this is an end that even our hardcore Stoic cannot let go of. And since whoever wills the end wills the necessary means, the Stoic could not rationally reject the help of others in fulfilling this end in case he could not do it by himself. Because there are some ends that we cannot give up willing, we must will that others help us in obtaining them. In addition, this means that as rational beings we are all necessarily committed to a law of beneficence. Moreover, this explains why Kant says that the "will would contradict itself" or the will would be in "conflict with itself" (GMS, AA 04: 423, 437) if it tried to will a maxim of non-beneficence as universal law. The egoist cannot will a maxim of universal indifference because, as a being with *pure* practical reason, he is already (and necessarily) bound by a law of beneficence – it is a law that springs from his pure rational nature, from his "authentic self" (*eigentliches Selbst*) (GMS, AA 04: 461). It is for this reason that there happens a contradiction in his will when he tries to will a maxim of egoism at the same time as a universal law. Indeed, Kant hints to the fact that there are some laws to which we are necessarily bound because they originate from our own pure will – something he will only explain in GMS III – when he says that "in every transgression of a duty, we find that we actually do not will that our maxim should become a universal law, *since that is impossible for us* [*denn das ist uns unmöglich*] [...]"

---

[80] Kant needs to show that the FUL can ground a duty of beneficence because in that way he shows that his derivation of a moral principle from the abstract concept of a rational being in general is consistent with common moral knowledge, which obviously takes beneficence (and truth-telling, not committing suicide, etc.) as paradigmatic cases of moral duties. Right after deriving the FUL, he says: "Now, if from this one imperative all imperatives of duty can be derived [*abgeleited*] as from their principle, then, even though we leave it unsettled whether what is called duty is not as such an empty concept, we shall at least be able to indicate what we think by it and what the concept means" (GMS, AA 04: 421). And right after the fourth example, he says: "These, then, are some of the many actual duties, *or at least of what we take to be such*, whose division [*Abtheilung*] can clearly be seen from the one principle stated above" (GMS, AA 04: 423-4, emphasis added).

[81] This example comes from Allison (2011, p. 190). On a similar line, Timmermann (2007, p. 85) says: "Kant assumes, not implausibly, that human needs present claims that cannot simply be rejected. Hunger, for instance, cannot just be reasoned away. A hungry person desires to eat, and it is entirely rational for him or her to hope for the assistance of those who could easily help. This assumption goes some way towards explaining the impossibility of a principle not to care about the hardships of others. If we cannot but desire the help of those more fortunate than ourselves, it is inconsistent for us to deny our assistance to those who are now in that kind of predicament."

(GMS, AA 04: 424, emphasis added). What we do will when violating a duty is, at most, that we be exempt, only "to the advantage of inclination" (GMS, AA 04: 424), from abiding to a law that we necessarily will and recognize as valid. One finds evidence for this in Kant saying, at the beginning of the section, that moral laws "originate completely a priori from pure but practical reason" (GMS, AA 04: 408).[82]

2.4 – In what sense are moral laws self-legislated?

Throughout this dissertation, I have been arguing that for Kant moral laws are self-legislated, since only self-legislation can account for their unconditional and universal bindingness. But what exactly does this mean? On the one hand, at first glance Kant's notion of autonomy can suggest a picture according to which "anything goes". For if all it takes to a moral law to be binding is that the agent legislates it, it looks like *whatever* agents legislate could become a moral requirement. Relatedly, if the self-imposition of a law is what makes it binding, it appears that agents can rescind laws they had previously legislated, thus getting rid of obligations at will. On the other hand, if there are indeed some kinds of norms governing the process of legislation, then it looks like these norms, and not the legislated laws, have unconditional authority. As Wood (1999, p. 156) has said, "it is easy to regard Kant's conception of autonomy as either incoherent or fraudulent".

In this section I wish to briefly discuss these issues and try to make Kant's notion of autonomy clearer. Now, as Sensen has suggested, it is instructive to first distinguish the metaphysical side of Kant's notion of autonomy from its normative side. On the metaphysical level, which is mostly expounded in *Groundwork* III, Kant claims that when one acts morally one cannot be bound by a law which gets its authority from some external (i.e. external from the agent's own will) source because this yields only hypothetical imperatives. Rather, since the will is a kind of causality, which implies that it necessarily works according to some law (also see GMS 04: 412), this means that the law one follows when acting morally must have

[82] See also Herman (2016, p. 153-4, orig. emphasis): "If a hypothetical imperative is valid for a rational being only so long as he sustains his commitment to the end contained in the imperative (the end for whose realization the imperative is a directive), a categorical imperative is to be valid for an agent without regard to his commitment or lack of commitment to ends, his desires, or the courses of action he, in the ordinary ways, has an interest in pursuing. Thus to the degree that different rational agents are distinguished from each other by their desires and interests, these are differences that determine the applicability of hypothetical imperatives but make no difference where categorical imperatives are concerned. It is this feature of categorical imperatives that suggest that they are properly to be looked at (if they exist) as *pure* principles of practical reason (or principles of pure practical reason)–they do not depend for their validity on contingent empirical features of agents, but merely on the fact of the agent's rationality–that the agent has a will."

the own will as its source. This is why Kant says that autonomy is a "property *of the will* [*Eigenschaft des Wilens*] of being a law to itself" (GMS 447, emphasis added). In other words, autonomy is a property of a will: a will is autonomous when it follows its own law.

This metaphysical side of autonomy also has a normative implication, however. Given that Kant's analysis in *Groundwork* II has revealed that categorical imperatives have "the dissociation from all interest in wiling from duty" (GMS, AA 04: 431) as a distinctive mark from hypothetical imperatives, then all categorical imperatives must be enjoined by a self-imposed law, for all other laws must have a non-moral interest if the agent is to adopt it. Kant makes this clear when he says:

> Thus the *principle* of every human will as *a will universally legislating through all its maxims*, if it is otherwise correct, would be very *well fitted* to be the categorical imperative, in that, precisely for the sake of the idea of universal legislation, it *is founded on no interest* and can thus alone, among all possible imperatives, be *unconditional*; or better still, by converting the proposition, if there is a categorical imperative (i.e. a law for every will of a rational being), then it can only command to do everything from the maxim of one's will as one that could at the same time have as its object itself as universally legislating; for only then is the practical principle, and the imperative the will obeys, unconditional, because it can have no interest whatsoever at its foundation. (GMS, AA 4: 432)

Consequently, a command to act on categorical imperatives is the same as a command to act autonomously, which lets Kant cash out the so-called Formula of Autonomy: "to do no action on a maxim other than in such a way, that it would be consistent with it that it be a universal law, and thus only in such a way *that the will could through its maxim consider itself as at the same time universally legislating*" [*daß der Wille durch seine Maxime sich selbst zugleich als allgemein gesetzgebend betrachten könne*] (GMS, AA 04: 434, orig. emphasis). Now, it is in this formula that we can find an answer to the objection that Kant's notion of autonomy is deeply flawed. Note that Kant says that the will must be able to *consider* itself as universally legislating at the same time through its maxim, not that the will must *legislate* universally through its maxim at the same time. What is more, one finds Kant using this counterfactual language about the notion of autonomy on many other passages, such as "so act *as if* your maxim *were to serve* at the same time as a universal law (of all rational beings)" [*handle so, als ob deine maxime zugleich zum allgemeinen Gesetze (aller vernünftigen Wesen) dienen sollte*] (GMS, AA 04: 438, emphasis added), and more evidently so when he says that

> According to this principle [i.e. the principle of autonomy], all maxims are rejected that are not consistent with the will's own universal legislation. Thus the will is not just subject to the law, but subject in such a way that it must also be viewed *as self-legislating* [*als selbstgesetzgebend ... angesehen werden muss*], and just on account of this as subject to the law (of which it can consider itself the author) in the first place. (GMS, AA 04: 431, orig. emphasis)

These passages make it clear that Kant's notion of autonomy does not have the incoherent or the voluntaristic aspects often associated with it. The Principle of Autonomy, through its formula, commands agents to act *as if* their maxims were to hold at the same time as universal law, or, equally, that agents act *as if* their will could be viewed or considered as *self-legislative*. It is with this reading in mind that I have been saying that moral laws are self-legislated, and how the FUL works as a test to decide whether a proposed maxim could have the form of a law. Kant's later treatment of autonomy in *Groundwork* II shows that when he says in the first section, concerning the feeling of respect, that moral laws are self-imposed laws, what he means is that moral laws are such that agents must be able to regard them *as* self-imposed. I take that this is enough to show that his conception of autonomy does not result in flagrant voluntarism.[83]

2.5 - Objection: autonomy in *Groundwork* II before the FA?

One could object that this account of the derivation in the second section also involves an unwarranted appeal to autonomy. Wood seems to have this in mind when he says that the derivation of the FUL relies on a "premature and surreptitious appeal to FA [i.e., Formula of Autonomy]" (1999, p. 81). More specifically, the complaint is that we would have reason to accept the derivation of the FUL if we knew that "the will is the *author* of objective practical laws, hence that the moral law is a principle of *autonomy*" (1999, p. 81, orig. emphasis). But since Kant only argues for this claim later when deriving the Formula of Autonomy (GMS, AA 04: 432), the derivation of the FUL rests on a weak presupposition. For Wood, this implies that neither the FUL nor the FLN "can be regarded as an adequate and self-sufficient formulation of the supreme principle of morality Kant is searching for in the *Groundwork*" (1999, p. 81).

The reply to this objection again relies on the fact that Kant is employing a regressive mode of argumentation. More accurately, he argues from some accepted moral judgments that form a *conception* of morality – one that is about duties, moral laws, the unconditional worth of a good will, and so on – to the ideas that are *presupposed* by that

[83] Arguing that the notion of self-legislation in the *Groundwork* parallels Kant's account of just political legislation in the *Naturrecht Feyerabend* lectures, Kleingeld (2018, pp. 173-4) says: "The Principle of Autonomy does not demand that one actually give moral laws. It involves the use of a legislation *analogy* by which Kant articulates a procedure for testing the moral permissibility of one's maxims. Understood in this way, his use of the idea of autonomy does not have voluntaristic implications: in no way does it imply that the content or obligatory force of moral laws is dependent on an act of the agent or on anyone's actual consent. By extension, it does not imply that the agent can rescind moral laws".

conception. In other words, the regress searches for the conditions of possibility of that conception, of what would make possible duties that are enjoined by universally and necessarily binding moral laws. Since the idea of *autonomy* is Kant's solution to this problem, it is understandable why a complete elucidation of this idea must come last. Yet on this methodological approach, he is justified in presupposing that concept in his arguments *before* he introduces it.[84] We can see that he is already arguing with the assumption that only autonomy can make a principle a law for only with self-legislation a principle can bind with absolute and universal necessity, *which is part of the concept of a very categorical imperative*. He says right before the derivation:

> For the time being, however, this much can be seen: that the categorical imperative alone expresses a practical LAW, and that the others can indeed one and all be called *principles* of the will, but not laws; since what is necessary to do merely for attaining a discretionary purpose can be regarded as in itself contingent, and we can always be rid of the prescription if we give up the purpose, whereas the unconditional command leaves the will no free discretion with regard to the opposite, and hence alone carries with it that necessity which we demand for a law. (GMS, AA 04: 420)

As I understand it, it is clear that here Kant is arguing with the assumption that moral laws are self-legislated. The unconditional command "leaves the will no free discretion" because it is the will that imposes this command on itself, in an unconditional way. In fact, Kant comes close to admitting that he had been arguing for the whole time with the *assumption* that moral laws are self-legislated, when he says – by the end of *Groundwork* II – that he had "to assume" (*annehmen*) categorical imperatives as excluding "from their commanding repute any admixture of interest, as an incentive […] because such an assumption was necessary if one wanted to explicate [*erklären*] the concept of duty" (GMS, AA 04: 431).

[84] For more on this issue, see Allison (2011, p. 238n4; p. 277n6), Timmermann (2007, pp. 42-3), and Paton (1947, p. 26).

Conclusion

The Formula of Universal Law is surely the most known version of Kant's Categorical Imperative. Not only that, but and I think it is not exaggerate to say that it might be the most famous principle of ethics in general. Its popularity, however, is in inverse proportion to the clarity of Kant's reasoning that leads to it, and this dissertation was motivated by the interest to understand the arguments he provides for this formula. Commentators, most famously Aune and Wood, argue that Kant's analyses of duty and of the concept of a categorical imperative allow him to provide a principle that is in fact tautological, and that Kant illegitimately moves from this principle to the FUL. As I argued in the introduction, however, it is only possible to interpret the derivation in this way if one attributes to Kant a principle he simply does not hold. Kant says that a good will acts out of conformity to universal law alone, but this is different from saying that to have a good will one must orient one's actions according to a principle like "Act on maxims that conform to universal law".

Kant's reasoning, in fact, is that one conforms one's action to universal law when one acts on the FUL, so my aim in the dissertation has been to show how this is so. In the first chapter, we saw that Kant first argues that when we act from duty we act out of respect for a moral law, and we feel respect for moral laws because they are products of our own will: they are self-imposed. Second, Kant argues that an agent with a good will adopts a practical principle not by virtue of the effects that could be obtained by acting on it, but rather because the practical principle is lawful; that is, has the form of a law. A good will, then, can be described as acting out of the "mere conformity of actions as such to law". Third, Kant gives a formula, the application of which lets agents check whether their proposed practical principle (their maxim) is indeed lawful, thus being the kind of principle a good will acts on. Since the laws to which agents should conform their maxims are self-imposed laws, the FUL works as a device with which agents can assess whether their subjective practical principles are in conformity to these laws: they do so by checking if they can will their maxims at the same time as universal law. To put it shortly, Kant can cash out the idea that a good will acts out of mere conformity to universal law into what agent can will as universal law just because the laws to which their action should conform are self-imposed, or self-legislated laws.

In the second chapter, we have seen that Kant's analysis of the concept of categorical imperative shows that the laws to which maxims should conform are universally and necessarily binding. Universality and necessity constitute thus the form of a practical law. Accordingly, Kant then provides a formula: a procedure which agents can use to assess if their

proposed maxims conform to those laws by checking whether they have the form of a law. Now, since a practical principle has the form of a law only if it is self-legislated, the FUL commands agents to consider whether their proposed maxim can be willed at the same time as universal law: if it can, this means that it could be regarded as a self-legislated law, and consequently, that it has the form of a law and that it conforms to universal law. To summarize with Reath's (2006, p. 110) words, the FUL "is a deliberative procedure that confers the status of law on those maxims (or their generalized versions) that it passes". As such, Kant can justifiably infer it from the analysis of the concept of a categorical imperative.

In conclusion, if the interpretation put forward in this dissertation is sound, then Kant does not commit a remarkable fallacy when deriving the principle he considers to be the canon of moral judgment (GMS, AA 04: 424). In addition, giving prominence to the fact that the notion of autonomy is already present at the derivation in *Groundwork* I gives us reason to believe that the widely held view according to which *Groundwork* II makes a "fresh start" (Allison, 2011, p. 35; Timmermann, 2007, p. 13) or consists of an "entirely new search" (Wood, 1999, p. 49) for the principle of morality needs some qualification. Indeed, it makes a fresh start by beginning with the analysis not of the concepts of good will or duty, but with the concept of a rational being as such. As I have tried to show, however, there is a certain continuity between *Groundwork* I and *Groundwork* II owed to the fact that Kant continues to argue with the assumption that moral laws are self-legislated, because only self-legislation can account for "that necessity which we demand for a law" (GMS, AA 04: 420).

Bibliography

Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Allison, H. (1991) 'On a Presumed Gap on the Derivation of the Categorical Imperative', *Philosophical Topics,* 19 (1), 1-15.

Allison, H. E. (2011) *Kant's Groundwork for the metaphysics of morals: a commentary*. Oxford ; New York: Oxford University Press.

Ameriks, K. (2003) *Interpreting Kant's Critiques*. Oxford : Oxford University Press.

Aune, B. (1979) *Kant's Theory of Morals.* Princeton: Princeton University Press.

Bacin, S. (2015) 'Kant's lectures on ethics and Baumgarten's moral philosophy', in Sensen, O; Denis, L. (eds.) *Kant's Lectures on Ethics: A Critical Guide*. Cambridge: Cambridge University Press, pp. 15–33.

Baumgarten, A. G. (2020) *Baumgarten's Elements of First Practical Philosophy: A Critical Translation with Kant's Reflections on Moral Philosophy*. Edited by C. D. Fugate and J. Hymers. London: Bloomsbury Academic. [1760]

Engstrom, S. (2009) *The Form of Practical Knowledge: A Study of the Categorical Imperative.* Cambridge (Mass): Harvard University Press.

Gaut, B. and Kerstein, S. (1999) 'The Derivation without the Gap: Rethinking Groundwork I', *Kantian Review*, 3, pp. 18-40.

Gressis, R. (2010) 'Recent Work on Kantian Maxims I: Established Approaches', Philosophy Compass, 5(3), pp. 216–227.

Guyer, P. (2000) *Kant on freedom, law, and happiness*. New York: Cambridge University Press.

Guyer, P. (2017) 'The Struggle for Freedom: Freedom of Will in Kant and Reinhold' in Watkins, E. (ed.) *Kant on Persons and Agency*. Cambridge: Cambridge University Press, pp. 120-137.

Herman, B. (1993) *The Practice of Moral Judgment.* Cambridge (Mass): Harvard University Press.

Herman, B. (2016) *Morality as Rationality: a study of Kant's ethics*. New York: ROUTLEDGE. (Ph.D. thesis defended at Harvard University in 1976)

Hill, T. E. (2000) *Respect, pluralism, and justice: Kantian perspectives*. Oxford; New York: Oxford University Press.

Horn, C., Mieth, C. and Scarano, N. (2007) *Grundlegung zur Metaphysik der Sitten*. Frankfurt am Main: Suhrkamp (Suhrkamp Studienbibliothek, 2).

Kant, I. (1902-) *Kants Gesammelte Schriften: Herausgegeben von der Deutschen Akademie der Wissenschaften*. 29 vols. Berlin: Walter de Gruyter.

Kant, I. (1996) 'Critique of Practical Reason' in *Practical philosophy*. Edited and translated by M. J. Gregor. Cambridge ; New York: Cambridge University Press (The Cambridge edition of the works of Immanuel Kant).

Kant, I. (1997*) Lectures on ethics*. Edited by J. B. Schneewind and P. Heath. Translated by P. Heath. New York: Cambridge University Press (The Cambridge edition of the works of Immanuel Kant).

Kant, I. and Kant, I. (1998) *Critique of pure reason*. Translated by P. Guyer and A. W. Wood. Cambridge ; New York: Cambridge University Press (The Cambridge edition of the works of Immanuel Kant).

Kant, I. (2012). *Groundwork of the Metaphysics of Morals.* Translated by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press.

Kerstein, S. J. (2002) *Kant's search for the supreme principle of morality*. Cambridge: Cambridge University Press.

Kitcher, P. (2003) 'What Is a Maxim?', *Philosophical Topics*, 31(1/2,), pp. 215–243.

Kitcher, P. (2004) 'Kant's Argument for the Categorical Imperative', *Noûs*, 38 (4), pp. 555-584.

Kleingeld, P. (2017) 'Contradiction and Kant's Formula of Universal Law', *Kant-Studien*, 108(1).

Kleingeld, P. (2018) 'Moral Autonomy as Political Analogy: Self-Legislation in Kant's Groundwork and the Feyerabend Lectures on Natural Law' in Sensen, O; Bacin, S. (eds.) *The Emergence of Autonomy in Kant's Moral Philosophy.* Cambridge: Cambridge University Press.

Kleingeld, P; Willaschek, M. (2019) 'Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law', *Philosophers` Imprint*, 19 (6), pp. 1-18.

Korsgaard, C. (1996) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Nyholm, S. (2015) 'Kant's Universal Law Formula Revisited', *Metaphilosophy*, 46(2), pp. 280-299.

O'Neill, O. (1989) *Constructions of Reason.* Cambridge: Cambridge University Press.

O'Neill, O. (2015) *Constructing Authorities: Reason, Politics and Interpretation in Kant's Philosophy*. Cambridge: Cambridge University Press.

Paton, H. J. (1947) *The Categorical Imperative*: *A Study in Kant's Moral Philosophy*. London: Hutchinson & Company.

Potter, N. (1975) 'The Argument of Kant's Grundlegung, Chapter 1', *Canadian Journal of Philosophy Supplementary Volume*, 5(1), pp. 73–91.

Rauscher, F. (2018) 'Transcendental and Empirical Levels of Moral Realism and Antirealism' in dos Santos, R. and Schmidt, E. (eds) *Realism and antirealism in Kant's moral philosophy: new essays*. [Berlin] ; Boston: De Gruyter (Kantstudien. Ergänzungshefte, Band 199).

Rawls, J. and Herman, B. (2000) *Lectures on the history of moral philosophy*. Cambridge, Mass: Harvard University Press.

Reath, A. (2006) *Agency and autonomy in Kant's moral theory*. New York: Clarendon Press ; Oxford University Press.

Reath, A. (2010) 'Formal principles and the form of a law' in Reath, A; Timmermann, J. (eds.) *Kant's* Critique of Practical Reason: *A Critical Guide*. New York: Cambridge University Press.

Rickless, S. (2004) 'From the Good Will to the Formula of Universal Law', *Philosophy and Phenomenological Research*, 68 (3), pp. 554-577.

Timmermann, J. (2003) ‘Sollen und Können: „Du kannst, denn du sollst“ und „Sollen impliziert Können“ im Vergleich’, *History of Philosophy and Logical Analysis*, 6(1), pp. 113–122.

Timmermann, J. (2007) *Kant’s Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge: Cambridge University Press.

Timmermann, J. (2018) ‘A Tale of Two Conflicts’, *Kant-Studien*, 109(4), pp. 581–596.

Timmons, M. (2006) ‘The Categorical Imperative and Universalizability (GMS, 421–424)’, in Horn, C. and Schönecker, D. (eds) *Groundwork for the Metaphysics of Morals*. Berlin: De Gruyter, pp. 158–199.

Schönecker, D; Wood, A. (2002) *Kant’s Grundlegund der Metaphysik der Sitten: Ein einführender Kommentar.* Paderborn: Ferdinand Schöningh.

Sensen, O. (2013) ‘The moral importance of autonomy’ in Sensen, O. (ed.) *Kant on Moral Autonomy.* Cambridge: Cambridge University Press, pp. 262-281.

Sensen, O. (2019) ‘How free is prudent behavior?’, *Studia Kantiana*, 17(1), pp. 11–27.

Stern, R. (2004) ‘Does “Ought” Imply “Can”? And Did Kant Think It Does?’, *Utilitas*, 16(1), pp. 42–61.

Velleman, D. (2005a) ‘Love as a Moral Emotion’ in *Self to Self: Selected Essays*. Oxford: Oxford University Press, pp. 70-109

Velleman, D. (2005b) ‘The Voice of Conscience’ in *Self to Self*, pp. 110-128.

Willaschek, M. (1992) *Praktische Vernunft: Handlungstheorie und Moralbegründung bei Kant*. Weimar: Metzler.

Willaschek, M. (2006) ‘Practical Reason. A commentary on Kant’s *Groundwork of the Metaphysics of Morals* (GMS II, 412-417)’ In Horn, C. and Schönecker, D. (eds) *Groundwork for the Metaphysics of Morals*. Berlin: De Gruyter, pp. 121–138.

Wood, A. (1999) *Kant’s ethical thought*. Cambridge; New York: Cambridge University Press.