Practical Reason and Morality: Kant’s Autonomy and the Debate on Aquinas’s Natural Law

In this paper I examine a thesis originally put forward by Max Weber and Ernst Troeltsch according to which Aquinas, in his doctrine of natural law, anticipated Kant’s, notion of the autonomy of practical reason. This, primarily, involves a close analysis of Kant’s argument in the “Groundwork of the Metaphysics of Morals” in which his conception of autonomy takes centre stage. My claim is that the Weber Troeltsch thesis is false, but not for the reasons usually indicated by its critics. Kant’s notion of autonomy neither contains the notion of discretion nor does Kant fail to provide an objectivist conception of morality. Rather, it is his anti-realist objectivism and its metaphysical foundation, i.e. the idea of transcendental freedom, which is at odds with Aquinas’s creationist metaphysics and realist metaethics. Aquinas’s metaphysics and metaethics allow only a weaker variant of human freedom and they require a reliabilist moral epistemology.
In this paper I examine a thesis originally put forward by Max Weber and Ernst Troeltsch according to which Aquinas, in his doctrine of natural law, anticipated Kant's notion of the autonomy of practical reason. This, primarily, involves a close analysis of Kant's argument in the "Groundwork of the Metaphysics of Morals" in which his conception of autonomy takes centre stage. My claim is that the Weber Troeltsch thesis is false, but not for the reasons usually indicated by its critics. Kant's notion of autonomy neither contains the notion of discretion nor does Kant fail to provide an objectivist conception of morality. Rather, it is his anti-realist objectivism and its metaphysical foundation, i.e. the idea of transcendental freedom, which is at odds with Aquinas's creationist metaphysics and realist metaethics. Aquinas's metaphysics and metaethics allow only a weaker variant of human freedom and they require a reliabilist moral epistemology.

* Erlangen University, Germany/Keio University, Tokyo.

1 I should like to thank the Japan Society for the Promotion of Science for their generous support of my work. My thanks also go to Damian Quinn and Sheila Regan (both Erlangen) for checking my English.
Practical Reason and Morality

In recent years it has become popular to compare the teachings of major thinkers on those issues which are still of interest for discussions in contemporary philosophy. These enquiries often challenge cherished and deeply entrenched positions and sometimes even manage to stir up debates among those who claim not to be concerned with questions of history at all. This phenomenon has been particularly evident in ethics. If we think of the virtue ethics movement, for example, questions are being asked as to whether and, if so, to what extent Aristotle and Kant really disagree after all\(^2\), a suggestion which would have been considered preposterous only a couple of years ago. In the field of philosophers with a Christian, and in particular with a Catholic background people are still struggling to understand how Thomas Aquinas dealt with the challenge of a powerful pagan conception of the good life which the discovery of Aristotle's ethics posed to Christian theology and its teachings about the end of man. Did Aquinas simply integrate the Aristotelian model so that there is a purely philosophical ethics side by side with an account which rests on the interpretation of what he takes to be divine revelation, or does his theological superstructure actually render such a construction impossible? If not, how did he come to grips with the obvious differences in metaphysics? Is metaphysics supposed to play a role in ethics in the first place? Does it play such a role in Aristotle, does Aquinas think it does in Aristotle, and does it in Aquinas? These are just a few of several questions which have dominated recent literature in the field.

In what follows I shall try to deal with another suggestion which many commentators no doubt will consider surprising to say the least, and which has not yet gained an audience in the philosophical world to the extent to which the discussions on Aristotle and Kant, and on Aristotle and Aquinas have done. The claim put forward by

Max Weber and Ernst Troeltsch about a century ago, argues that Aquinas anticipated Kant's idea of the autonomy of practical reason in his doctrine of natural law\(^3\). This thesis has figured prominently in the vigorous debates surrounding the *Humanae Vitae* encyclical\(^4\) and its teachings on artificial contraception. According to one influential commentator (i.e. Honnefelder 1988: 256), the 'Weber-Troeltsch thesis' (WTT henceforth), as we may call it, has been given firmer grounding by recent scholarship on Aquinas, whereas Weber and Troeltsch themselves based their claim on a mistaken interpretation of the Aquinian doctrine of natural law itself which was heavily influenced by the Neo-thomist teachings of their time. Not surprisingly, the WTT has not remained unchallenged. Detailed studies, such as Rhonheimer's (1987), have been produced whose intention is to show by means of minute textual analysis that on the question of autonomy Kant and Aquinas are as far apart as received opinion would expect them to be.

In order to understand the ferocity of the debate we should examine briefly the immediate theological context\(^5\). In the aftermath of *Humanae Vitae* some of those who disagreed with the official teaching of the church tried to invoke the authority of Aquinas — still the Catholic master-thinker in spite of his downgrading in the Second Vatican Council — to give a boost to their case. The strategy behind it was fairly simple: if Aquinas did anticipate Kant's doctrine of autonomy, then there was room for what they called "creative reason" on the part of man. If there was creative reason then reason's function in at least some matters of morality is not just epistemic, but also constitutive. It does not simply read off pre-existent norms which are to be enforced, but "creates" such norms in the first place. The key assumption for the strategy to go

---

\(^3\) Vd. Honnefelder (1988).

\(^4\) *Acta Apostolicae Sedis* 60 (1968), pp. 481–503.

through is the notion that creative reason involves some form of discretion, i.e. reason can choose among alternatives what to enact as norms. The proponents of this view have been called "autonomists" and, among other things, they have been charged, on the one hand, with neglecting the laborious task of textual analysis in favour of political manoeuvring, and, on the other, of misunderstanding, or at any rate misconstruing Kant's conception(s) of autonomy. Partly as a result of these charges, a group which soon obtained the name "moderate autonomists" emerged which tried to adhere to the WTT without exhibiting the methodological shortcomings of their predecessors.

It is fair to say, that all these attempts are based on a picture of Kant's moral philosophy which has not yet taken into account the extraordinary revival of interest in Kant's moral thought among both scholars and philosophers working in the relevant fields. This revival is particularly strong in the English-speaking world, which has not, for a long time, exactly been fond of his approach to philosophy. In virtue of this development many widespread caricatures of Kant's teachings have been revealed to be just that: caricatures with little or no footing in the texts themselves. These admittedly sometimes contentious corrections mainly concern Kant's view of the emotions and their role in moral motivation as well as his theory of action. In addition to this, recent commentators claim that Kant's moral theory rests on a firm metaphysical basis, a suggestion that echoes a more general trend towards reassessing the role of metaphysics in Kant's entire thinking. If in the debate on the WTT, these thinkers' distorted and anachronistic view of Aquinas is in need of correction, then in my opinion, so too is the picture of Kant.

In what follows I shall try to provide some material for such a revised picture of Kant's moral philosophy, and I shall do this by watching him laying its foundations. That is to say, I shall try to give an account of the way in which Kant's argument in the *Groundwork of
the Metaphysics of Morals works. Since in this short but extraordinarily important book his conception of autonomy plays the central role, a proper understanding of this work may help us better grasp what he actually had in mind. This, in turn, is crucial, of course, for a proper assessment of the WTT. Unfortunately, there is not enough time to do the same with the other side of the equation, namely Aquinas's theory of natural law, so that the matter cannot be settled conclusively. All I can do in this respect is give an outline of the main aspects relating to Aquinan thought that need to be considered when examining the WTT.

The paper is divided into four parts. In the first part I shall examine the overall structure of the Groundwork as well as the key ideas and arguments of sections I and II. The second part deals with the sometimes quite obscure, but nevertheless crucial section III. Although the notion of autonomy is introduced by Kant as early as section II, it will be dealt with separately in part three. In the fourth part, the view of Aquinas mentioned above will be presented.

Although restrictions of time and space do not permit me here to defend my arguments on the Aquinan half of the comparison, I shall give a brief outline of my general view. Although both Aquinas and Kant claim that unaided natural reason has access to moral truths, this is not sufficient evidence to maintain that Aquinas anticipated Kant's notion of autonomy. Kant's notion of autonomy is closely connected to his strong conception of selfhood, and it is this conception of selfhood which is incompatible with Aquinas' creationist metaphysics. In addition, whether this is true or not, the Kantian notion of autonomy does not include the idea of discretion. The Kantian moral principle may, when compared to the Aquinan principles, lead to different results in the attempt to answer moral questions, but this is an entirely different matter. Let us now, though, turn to the Groundwork.
I

The crucial passages which require consultation if we are to achieve a clear picture of what Kant had in mind when talking about "autonomy" are found in sections II and III. To understand them properly, we should first take a look at the overall strategy of this work. In the oft-neglected preface, Kant presents the idea of situating ethics parallel to physics in at least one important respect. Just as physics has a pure part consisting presumably (mainly) of Newton's laws of motion which ground the status of law of any special law of physics, ethics is also in need of a similar set of basic principles in terms of which the normative force of all obligations can be accounted for (which is not to say that there is no extra element needed to account for the obligations). This pure part of ethics is called the "Metaphysics of Morals". The strategy just described raises the question of how these basic principles, both of physics and ethics, can themselves be accounted for. Although Kant nowhere in the *Groundwork* makes it explicit how this is to be achieved, we might allow ourselves a fair guess: the laws of motion (metaphysical principles) are specifications of the so-called transcendental principles of the Analytic of the first critique, that is to say the principle of inertia, to take just one example, is a specification of the causal principle, and the factor which renders them a specification is physics treating objects in space not simply as extended objects, but as material objects which, since material objects are equipped with certain basic forces, can in turn be accounted for in terms of mathematics. The transcendental principles themselves are certainly not specifications of yet higher principles, but rather conditions of the possibility of the unity of the world of experience, or of the unity of consciousness amid the floating world of appearances.

With this in mind, it may be possible to understand a little better
the, initially no doubt puzzling, claim of a parallel between physics and ethics. Kant sees a similar relationship of specification between the moral law and the categorical imperative. The moral law is supposed to be valid for all rational beings equipped with a will, whereas the categorical imperative is the mode in which the moral law is addressed to those rational beings who are also equipped with sensuous components. The question is of course, whether there is such a moral law and its specification and, if so, how it can be accounted for. The task of the *Groundwork* is precisely to identify and vindicate this moral law, which can in turn produce the foundation on which the system of ethics can be erected.

Kant goes about this task in three steps, these corresponding to the three sections. The first section analyses the moral practice of everyday life as Kant sees it unfold in front of his eyes. The second section consists of a metaphysical analysis of the faculty of the will or of practical reason. It is in this section that Kant claims that the moral law originates in the will of a rational being itself. The third section, labelled “transition from metaphysics of morals to the critique of pure practical reason” is intended to: (a) confirm the findings of the previous sections as to what the moral law or the categorical imperative consists in, (b) to vindicate the moral law, and (c) to find an account for its specification in the categorical imperative.

We are now in a position to examine the particular sections more closely. The crucial strategy of section I follows an almost Humean procedure. According to Wood (1999: 30–33) Kant is interested in a particular kind of reactive attitude which people show towards certain human actions, namely esteem (“Hochschätzung”), as opposed to mere approval (“Billigung”). He endeavours to discover the assumption which is the basis of esteem being shown towards an action, and it turns out that these actions are thought to originate in a good will. The criterion which makes the will ‘good’ is said to be the quality of volition itself, and to clarify this notion Kant invents
a number of examples meant to demonstrate that the quality of volition, on the assumption of which esteem is shown, is a question of motivation. To clarify exactly what (assumption concerning what) element in motivation leads to esteem, Kant assumes that there are no inclinations, direct or indirect, on the basis of which the action occurs. Thus, the action must (be thought to) have been done from duty, i.e. simply because it is the right thing to do. This does not mean that whoever does something on the basis of inclinations is morally to blame. Kant merely believes that if people think something right has been done out of inclination they show only approval (presumably since the agent was just lucky to have had the relevant inclination)\(^6\). Given all this, the moral law in virtue of which there are duties in the first place must be able to directly motivate agents. Kant somewhat hastily concludes that the moral law can therefore only be the famous universalisability requirement (UR) of the maxim of the action, since a maxim states the reason for which the agent acts. He hastens to show, however, that this very test is actually used in common moral practice in order to determine what is morally forbidden in a certain situation, for example making a false promise in order to cheat one's way out of financial difficulties. At this point of the investigation we do not know whether these assumptions on the part of ordinary people are in any sense universally justified or completely fantastical, or, to take another possibility, due to a rather idiosyncratic attitude taken by 18th century Prussians. In particular, we do not yet know whether there really are any duties and whether the underlying moral law can in fact directly motivate agents.

In section II Kant shifts the realm of the enquiry from the possible contingencies of common moral practice to the sphere of meta-

\(^6\) The questions as to whether duty can be a second order motive (cf. Walker 1998: 19) and whether there can be motivational overdetermination (cf. Baron 1995) of action address related, but different issues.
physics, and analyses the faculty of the will or practical reason itself. Here Kant tries to establish that this very faculty is the source of normativity, but the way in which Kant’s argument is intended to work is anything but clear. One reason for this difficulty lies in the structure of the text. The line of the argument is often interrupted by renewed treatments of the examples. There is a lengthy section introducing a vast array of technical terms that are, subsequently, not used at all in elucidating the core ideas, and the crucial sections themselves are disconcertingly short and therefore both open to and in need of interpretation. Finally, we are told that the results of this section are, in an important sense, still conditional (e.g. GMS AA IV, 429 Fn 1) but Kant does not elaborate upon the precise condition which needs to be met. It may be the fact that conceptual analysis cannot guarantee the instantiation of its object, i.e. the will; or it may be that there is a certain feature of this object which conceptual analysis alone cannot uncover.

I shall therefore concentrate on the core passages of section II. In Kant’s treatment of the will the following two aspects are of particular importance:

(i) The will is defined as the capacity to act on the representation of laws (GMS AA IV, 412 and 427), and it is taken to be a faculty different not only from sensual appetite, but from appetite in general.

(ii) In apparent conflict with the conclusions of section I, which dismissed the idea that the criterion of the goodness of will derives from an end, Kant now tells us that the objective basis of the will’s self determination is, in fact, an end (GMS AA IV, 428 sq.). He further argues that this end, if given by reason alone, must be valid for all rational beings, just as the moral law which we are trying to find is supposed to be valid for all rational beings. In what follows and in accordance with Kant’s own terminology, I shall use “objective”
to indicate this feature of the end we are looking for. In order to estab-
lish what this end could be, Kant proceeds in two steps. First, by
way of elimination, he tries to single out this end from a list of possi-
bile candidates. Secondly, he tries to establish why this end is in
fact the end we have been searching for. In a curiously twisted, but
nevertheless typically Kantian line of thought, the no doubt baffled
reader learns in fewer than two pages that this end is in fact the ra-
tional being itself. In virtue of its special worth, known as dignity,
it requires a special form of treatment and thus the existence of
duties can be established, even if their precise nature is not yet deter-
mined. Recall that this was still an unwarranted assumption in sec-
tion I. A number of further comments on both steps of the argu-
ment are certainly necessary.

Let us first look at the list of possible candidates. An important
point to heed is that Kant limits the list of possible candidates to pre-
existing ends, as opposed to those to be brought about or brought
into existence. This, arguably, removes the air of contradiction
about what has been said in section I, but it surely requires an expla-
nation which Kant, unfortunately, does not provide. Avoiding too
much detail, my interpretation would be as follows: Since the bring-
ing into existence of an end is never a matter entirely within our
power as a free agent, and since this end is essentially connected to
the moral law which in turn applies only to actions as far as they
are in our power, those ends are of no concern for our inquiry.
Within the list of pre-existing ends we find, in this order: (i) objects
of the inclinations; (ii) the inclinations themselves; (iii) non-rational
creatures, and (iv) rational creatures. Of particular importance for
the overall question we are addressing here — as will transpire a
little further on — are the reasons why Kant dismisses (i) and (ii).
The objects of the inclination, he says, have worth relative only to
the inclinations in question and would lose their worth together
with them. This is presumably meant to be in tune with the objectivi-
ty requirement, but is not entirely satisfactory, because it may be that a rational creature needs to respect precisely those inclinations which only particular rational beings happen to have. A more promising way of reading Kant has been suggested by Christine M. Korsgaard (1996: 118–124). She distinguishes two types of objective ends: conditional ends, the conditions of which are fulfilled, and unconditional ends. The very passages in question here are, in her interpretation, to be read as a regress argument meant to find the unconditioned condition for possible conditioned objective ends. Consequently, the question arises as to whether the condition of the worth of the objects of inclinations, namely the inclinations themselves, can provide this condition. Kant’s position is simultaneously unambiguous and surprising:

“But the inclinations themselves, as sources of needs, are so far from having an absolute worth, so as to make one wish to have them, that it must instead be the universal wish of every rational being to be altogether free from them” (GMS AA IV, 428, trans. Gregor 37).

Now even a sympathetic a reader such as Korsgaard parts company with Kant on the subject of the outright dismissal of the inclinations tout court. She also limits Kant’s claims to those inclinations, which are disruptive to our happiness (Korsgaard 1996: 120). It seems clear, however, that Kant indeed wants to be understood in a much stronger sense. It is, unfortunately, equally clear that Kant fails to provide an argument for his claim. All we can assume is that the reason for this dismissal must have something to do with Kant’s conviction that all rational beings must, in an important sense, be on the same level in their relationship with the moral law. What this “important sense” might be will hopefully become clear as we move on.
In any event, the same line of thought is clearly at work when Kant is presenting his own objective end, or to be more precise, unconditioned objective end, namely rational beings as such. Before we turn to his own justification of this claim, it should be pointed out that, especially in the context of our general aim to compare Aquinas’s and Kant’s account, it is striking that Kant does not consider the option that only one special kind of rational beings, namely God may be the objective and indeed unconditioned end of the self-determination of the will. According to this conception, knowledge of God is the natural end for man and the beatific vision is his supernatural end\(^7\), the achievement of which is possible only through divine grace and, in turn, a perfect fulfillment of the natural end. As far as the question of happiness is concerned it is clear that for Kant, who turns on its head the argument for misology (or hatred of reason) to which common moral cognition is prone to fall victim, reason’s function — assuming natural teleology — is not the achievement of happiness. Thus, following Rousseau, for man equipped with reason there is not just one privileged way of materially filling the concept of happiness according to Kant. Rather, reason opens up a potentially infinite set of new needs, the fulfillment of which is at least a precondition for happiness. Accordingly, in his discussion of heteronomy as the source of all spurious principles of morality (GMS AA IV, 441), happiness is classified as an empirical principle\(^8\). In the case of God, and quite independently of treating him as the natural or supernatural end of our actions, Kant thinks that treating him as the source of moral law amounts either to circularity or to crude voluntarism\(^9\).

\(^7\) Cf. Denis J. M. Bradley’s (1997: 395–404) critical discussion of this point.
\(^8\) We may say that for Aquinas there is a threat of too much natural reason, whereas for Kant there is the threat of reason as a total failure in the sphere of the practical.
\(^9\) As to the charge of circularity this needs to be discussed in connection with the question whether Aquinas anticipated Grotius’s famous etiamsi
We are now, finally, in a position to turn to Kant's claim that we find the (unconditioned) objective end in nothing else but the rational being itself. Kant justifies it in an extremely condensed passage. He says:

"The ground of this principle [sc. the universal practical law — W.E.] is: rational nature exists as an end in itself. The human being necessarily represents his own existence in this way; so far it is thus a subjective principle of human actions. But every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me; thus it is at the same time an objective principle from which, as a supreme practical ground, it must be possible to derive all laws of the will" (GMS AA IV, 429, trans. Gregor 37 sq.).

As mentioned above, this argument is in great need of interpretation. Apart from the question of what rational ground human beings along with all other rational beings in fact have to represent their existence as being an end in itself, it seems to me that we are in need of even the slightest clue as to what exactly Kant is trying to get at. This clue, in my opinion, has again been provided by Korsgaard and developed further by Allen W. Wood. According to Korsgaard what Kant is trying to get at can be described as follows.

Every fully rational action must aim at the good and the good is something at which the fully rational action of all rational beings aim. Since according to our list there is no other pre-given end, it is the end-setting capacity itself which determines what the good is and it does so in virtue of what she calls its "value-conferring status" (Korsgaard 1996: 122). This corresponds to the claim that whatever end I set myself, as long as it is morally permissible I

daremus, cf. Honnefelder (1988) and many others. Hopefully, as to the crucial question it doesn't matter whether this is correct.
expect it to be respected. In other words, it must be up to me to what I attach my heart, as it were\textsuperscript{10}.

According to Wood (1999: 124–132) this means neither making an empirical claim, nor attempting a reduction of value. He complements Korsgaard’s ideas in suggesting that Kant is trying to avoid the pitfalls of an objective conception of morals which is construed in realist terms. Wood, here, invokes the terminology of 20th century meta-ethics, in particular the distinction between externalism and internalism as well as John Mackie’s so-called “queerness argument”\textsuperscript{11}.

According to Wood, Kant’s idea of the value-conferring status of the end-setting capacity itself amounts to an anti-realist, but nevertheless objectivist conception of morality. It frees Kant from the need to assume a quality, such as ‘to-be-doneness’ residing in the object, and from the need to assume a no less queer human faculty by means of which we can pick up this quality and instantly become motivated.

Accordingly, in virtue of its value-conferring status and thus its absolute worth, Kant takes the end-setting capacity (a feature of the will) to be the source of normativity\textsuperscript{12}.

At precisely this point, however, we may again ask whether Kant has failed to consider an alternative, namely a naturalistic alternative. Why cannot both a feature of an object and ordinary human in-

\textsuperscript{10} This, I take it, is also the deeper aspect of Hegel’s criticism in the famous §135 of the “Grundlinien”.

\textsuperscript{11} Nevertheless, Kant’s predecessors, the moral sense theoreticians and the rationalists faced similar problems as indeed again Korsgaard (1998) has shown in another important paper of hers.

\textsuperscript{12} The question is whether it is already in virtue of the good sub cuius ratione appetimus (cf. Critique of Practical Reason AA V, 59) that we get the universalisability requirement, or rather: under the condition that there is a fully rational action it must — as an action — have an end which qualifies as good. As to Kant’s moral anti-realism cf. Rauscher (2002).
clination work in harmony to produce a workable externalist and realist alternative to Kant’s antirealist and internalist conception.\footnote{Whether this is ruled out by Hume’s law as well as by problems caused by supervenience is yet another problem.}

II

We can hopefully answer this question by turning now to section III.\footnote{Cf. Schönecker (1999) for an overall and in many respects different account.} As I am trying to show, this question is closely connected to the question identified above, a question which still needs to be answered, namely in what sense the results of sections I and II are still provisional. Two possible alternatives have been considered: (i) conceptual analysis cannot guarantee the reality of its object, or (ii) there is a feature of the object which cannot be detected by conceptual analysis.

Section III is entitled “transition from metaphysics of morals to the critique of pure practical reason” and it is in precisely this section that Kant hopes to revise the hitherto provisional nature of the enquiries. Its structure is quite straightforward, but nevertheless, involves a number of problems.

Kant starts off by presenting what has been called the “reciprocity thesis” (RT) by Allison (1998: 273), according to which a free will and a will under moral laws is one and the same (GMS AA IV, 447). For the sake of convenience we may interpret the RT as a biconditional which consists of the ‘freedom (of will) thesis’ (FT) and the ‘moral law thesis’ (MT) and so we may delete the plural in the phrase “moral laws”, since this refers either to the fact that the moral law will be shown to require a certain feature of all the maxims we have, or else indicates that there are various formulations of this one moral law (as a consequence of the respective formu-
lation of the categorical imperative). In the subsequent paragraphs Kant tries to establish the FT which, in virtue of the RT, would also produce the MT. So far, so good. Looking more closely, however, things become considerably murkier. The most important, but by no means only, points in need of clarification are the following four:

(i) Kant’s own argument for establishing the RT is highly problematic since it seems to imply that immoral actions are not free and hence not imputable. This is one horn of what has been called “Reinhold’s dilemma” (Allison 1998: 295), as the German philosopher Carl Leonhard Reinhold is considered the first to have spotted this possible defect in Kant’s thought. Although he refers to the Wille-Willkür distinction in the *Metaphysics of Morals* the problem is essentially the same.

(ii) It is unclear in what sense we are to understand “freedom” in the FT? Although as many as five different conceptions of freedom have been detected in Kant, the most likely candidates are — following Kant’s preceding discussion in the *Critique of Pure Reason* — practical freedom and transcendental freedom. The crucial difference between practical and transcendental freedom is the degree, or more precisely, the kind of independence from the inclinations it asserts. With practical freedom each choice and action is taken in virtue of some inclination, whereas in the case of transcendental freedom this link is disconnected. That is to say, that a transcendentally free being can act counter to the dictates of one natural drive without having to rely on another natural drive (cf. Allison 1998: 287 and Allison 1996: 111).

(iii) Why, after establishing the FT relatively early on in the text, does Kant apparently still carry on arguing for the MT? Given the RT, establishing the FT should surely be enough to achieve the goal

---

15 Exactly how many of these formulations we need to distinguish is a matter of debate.
of the argument of section III. In particular, is there a need for an extra argument in order to establish the validity of the moral law as a categorical imperative for human beings?

(iv) And, finally, how and in what sense is Kant actually establishing the FT?

We need to address all these problems in order to find an answer to our leading questions, and to give a preview of what is to come later the following outline may suffice:

In (i) Kant is not addressing the problem of causality, at least not primarily, but rather the problem of justification. The claim, following Allison, will be that universalisability of the maxim is the only available means by which the transcendentally free being — answering (ii) — is not to be condemned always to act arbitrarily and hence irrationally. That Kant has in mind the transcendental variant of freedom will become clear shortly. Kant’s aim, furthermore, is to identify morality with practical rationality. Hence, regarding (iii), there is no need for an additional argument in order to establish the validity of the moral law for human beings, but a need only for an explanation as to how this identification is supposed to work and how the concern for practical rationality as such can outweigh other concerns human beings as sensual beings have.

Finally, as to (iv), according to Kant only in virtue of transcendental freedom can a rational being engaged in the field of the practical be a ‘self’. The other models of practical rationality (and consequent-

---

16 in line with contemporary attempts to do so by means of a different, objective end conception of practical reason by, for example, Foot (2001) and probably also Murphy (2001) and, of course, in the works of John Finnis, Germain Grisez and Joseph Boyle (e.g. 1987).

17 This of course leads us back to the question as to what extent the quest for happiness (and indeed how we are to understand “happiness”) is an ineradicable factor in all human actions and the difficulties this causes for Kant’s moral theory. Cf. Johnson (2002) for a thoroughgoing discussion of these topics.
ly weaker notions of freedom, such as practical freedom) deprive the
rational being of the possibility of self-determination. Without
doubt, this rather strong conception of the self is controversial, in par-
ticular since it comes at the expense of an ability to account for the ac-
tivities of this self in terms of the categories and principles suitable
for "natural" events. Admittedly, Kant hastens to point out that "trans-
cendental freedom" cannot be established in the sense of theoretical
knowledge, but only as an assumption related to the self-
understanding of a rational being engaged in action.

We can now return to our question regarding the hypothetical or
provisional nature of the findings up to section II, which section III,
discussed here, will hopefully address. Two possibilities are mention-
ed above: (i) the need for a reassurance that the concept of the ca-
pacity, which was analysed, is in fact instantiated, i.e. that there is
will or practical reason. Or (ii) if the instantiation of the will is not
really a problem then section III might be meant to reveal an extra fea-
ture of the will which validates the argument of II. In light of what
has been said a little earlier about the argument of section III, both
(i) and (ii) still make perfect sense. Together with the establishment
of the FT in the sense of transcendental freedom, Kant has a fortiori
shown that the will is instantiated. On the other hand, it is the
strong connection between selfhood and transcendental freedom
which does the trick for section II. According to this reading, the
notion of selfhood in virtue of transcendental freedom might be
meant to establish the value conferring status of humanity as the ca-
pacity to set ends\(^{18}\).

In any case and whichever of the two options Kant actually did

\(^{18}\) There may indeed be a third possibility that by establishing that there is
perfectly rational action in virtue of transcendental freedom only insofar
as the UR is fulfilled, this somehow reinforces the claim that the concept
"good" must be applicable, but can only be applicable in virtue of the
value-conferring status of humanity.
pursue, two things become now clear: (i) the motivational assumptions of common moral cognition hit a deep truth and are not just a freak of contingency; (ii) it now turns out that Kant has not overlooked a possibility in section II, namely the externalist and realist conception. It is precisely this idea which he obviously takes to be incompatible with his conception of selfhood. This is confirmed by taking a look at some of the key passages of section II, so far omitted in our account, which also indicate that Kant is indeed working with the transcendental variant of freedom:

"Now I assert that to every rational being having a will we must necessarily lend the idea of freedom also, under which alone he acts. For in such a being we think of a reason that is practical, that is, has causality with respect to its objects. Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgment not to his reason but to an impulse. Reason must regard itself as the author of its principles independently of alien influences; consequently, as practical reason or as the will of a rational being it must be regarded of itself as free, that is, the will of such a being cannot be a will of his own except under the idea of freedom, and such a will must in a practical respect thus be attributed to every rational being" (GMS AA IV, 448, trans. Gregor 53 sq.).

III

Before addressing the WTT we must clarify what autonomy is for Kant. We can do this using a two-step approach: first, by investigating why the UR is the moral law and then by explaining how the UR is connected to the notion of autonomy.
The UR and the Moral Law

In our investigation, we have to distinguish nicely between Kant's own intentions and an assessment of the quality of his arguments. There can be no doubt that Kant thinks that the UR is the only possible candidate for the moral law, but nevertheless, it is concerning this precise point that the charge of a "gap" in the derivation of the categorical imperative has been leveled against Kant. Recall, that in the text of the *Groundwork*, the UR is identified as the content of the moral law on two occasions: first it is said, in section I, to be at work in common moral practice as the principle which is able to directly motivate an agent, and it is also said actually to be used in determining that making a false promise to get oneself out of financial difficulties is not a moral option (GMS AA IV, 402 sq.). Secondly, in section II, a conceptual analysis — carried out along with the conceptual analysis of practical reason or will — of the concept of a categorical imperative as opposed to a "mere" hypothetical imperative is used to show that the former contains the UR. Thirdly, in section III, we learn that the MT is implied by the FT, and I take this to suggest that what was, also provisionally, identified as the content of the moral law is actually confirmed by the RT, and the argument for the RT is meant to establish this very point.

But let us take a closer look at what the gap charge amounts to in order to understand Kant's point better. The charge, in essence

---

19 by Bruce Aune (1979: 29 sq.) and others; Wood (1990: 161–164) takes this to be Hegel's point as well.

20 In virtue of these considerations, it is again surprising to learn that another sympathetic commentator, namely Allison (1996: 144) himself claims that the "derivation" of the categorical imperative has actually failed in the *Groundwork* in contrast to the so-called "metaphysical deduction" (Allison 1996: 150 (following Lewis White Beck)) of the *Critique of Practical Reason*, although Allison also says that Kant
(leaving aside terminological problems concerning "maxim" and "practical law"), is that of an illicit slide from two very different senses of "formal" or "universal". Whilst it can mean that a maxim is applicable to all rational agents in relevantly similar circumstances, it can also mean something far stronger, namely that it is rational for each agent (including the agent whose maxim is being tested), if all rational agents are acting simultaneously on the maxim in question. Following Wood (1990: 135 sq.), we can call the primary sense of "universality" the "universality of applicability", and the second sense the "universality of concern", or "collective rationality" (cf. Allison 1998: 146). The very different nature of these considerations can be illustrated by taking a look at the example of the rational egoist. He can claim that it would be reasonable for everyone in relevantly similar circumstances to act on this particular maxim, but nevertheless hope that indeed not everybody does act on this maxim, since it would make the attainment of his ends far more difficult, if not impossible. The maxim of the lying promise can similarly be said to be rational when employed by everyone in relevantly similar circumstances (given the wishes and preferences the agent has), but nevertheless the potential false promiser cannot wish, let alone will, that his maxim be followed by everyone (including himself now) since this would render making a promise (false or not) useless as a means of pursuing one's ends. This is, undoubtedly, exactly how the UR is supposed to work, but the essential question is how Kant can get from the harmless sense of universality to this sense of universalisability on which the UR rests.

In Allison's account it is the notion of transcendental freedom somehow manages to justify the UR in virtue of the RT. This can only mean, that the identification of the moral law with the UR in section I and II is to be taken to be merely defective. Indeed, Allison further below in his paper downgrades his charges to not having made the presupposition of transcendental freedom fully explicit.
which makes this transition possible. Because of the nature of transcendental freedom the universalisability of the maxim is the only feature which can render the choice of the agent rational, since no desire — in virtue of transcendental freedom — can do so, since the desire itself is capable and in need of further justification as to whether it can function as the basis of rational choice. Accordingly, we need to distinguish fully rational actions and partially rational actions or agents. An action is partially rational if it is justified, for example, as a means to something else which is itself not justified. A fully rational action, on the other hand, is an action the justification of which is itself justified or cannot itself be justified because it is simply given. This means that, in contrast to the action of a transcendently free agent, the action of a practically free agent can be fully justified in virtue of some desire.

**Autonomy and Validation**

We are now in a position to understand better what the moral law amounts to and, accordingly, what “autonomy” refers to in the Kantian sense. We have seen that the UR states the necessary and sufficient condition for a transcendently free being to act fully rationally. Yet it is far from clear how this requirement is linked to autonomy. Kant makes the following claim: “Autonomy of the will is the property of the will by which it is a law to itself” (GMS AA IV, 440, trans. Gregor 47). Directly after this explanation, he states that the principle of autonomy is the UR. There is, however, no further elucidation as to why this is the case and how precisely we are to understand this connection. Again, we are in need of an interpretation, and I shall try to provide one by dealing with an objection which has been raised against Kant by Rüdiger Bittner and, in a similar fashion, by Gerold Prauss (1983). In what follows, I shall restrict my focus to Bittner’s suggestions.
According to Bittner (1983: 124 sq.), Kant is guilty of confusion in identifying the UR with the principle of autonomy. The principle of autonomy deals with conditions for the validity of moral laws whereas the UR is itself a moral law. This shows, in Bittner's view, that Kant failed to distinguish two different levels of reflection, namely the level which is concerned with moral laws, and the level which is concerned with how we should act. All Kant can show, in Bittner's opinion, is that (i) we can only act on self-chosen laws (Bittner 1983: 151 sq.), and (ii) that there are two conditions which together are sufficient and necessary to secure the validity of moral laws (Bittner 1983: 165–168): On the one hand they need to be self-chosen and on the other they need to be approvable by all rational beings. This, according to Bittner, has fatal consequences. If inherent in the concept of a moral requirement is the fact that it is also addressed to those who are reluctant and unwilling, the condition that only self-chosen laws can be valid makes a moral requirement impossible. In addition to this, the fact that a moral law can be approved of by all rational beings is not a sufficient reason to adopt it.

In a later essay, Bittner addresses Allison's attempt to rescue Kant's argument for the RT by interpreting it in terms of justification. He says:

"In his paper 'Morality and freedom: Kant's reciprocity thesis', which in a revised version went into chapter 11 of his book "Kant's theory of freedom", Allison has presented a Kantian style argument for the lawfulness of all free actions. It is grounded on the idea that a being which in its actions is independent from the determination through inclination and nature and which thus — according to Allison's use of the term — possesses transcendental freedom can only justify the choice of its maxims in virtue of the unconditioned practical law, which in the categorical imperative is expressed as a requirement. This
is a Kantian style argument, but not Kant's argument in the passage quoted above [sc. GMS AA IV, 446 sq., W.E.] (which is not Allison's claim) and therefore I do not wish to address it in the present context. One remark, however, should be allowed: Allison requires an agent, as a rational being, to be able to justify his maxim, but it is not clear why he can do so. Justification is a human practice: why does this practice have the privilege of participating in it being indispensable for rationality? Rational beings appear to be conceivable who act on maxims and thus on reasons but who provide justifications of these reasons maybe only occasionally, maybe not at all" (Bittner 2000: 216, Fn 19, my translation).

In short, his main argument in replying to Allison is that: a) justification is a human practice, and b) that Allison failed to provide an argument as to why it should be crucial for rational beings to participate in this human practice.

Bittner's criticism obviously draws on a notion of autonomy which grants an agent full discretion as to the principles he or she wishes to choose to act upon and then asks whether there can be good reasons to choose a principle which qualifies as a moral law. Accordingly, "autonomy" is interpreted in a way which seems to comply perfectly well with the literal meaning of the term, i.e. self-legislation, and with the meaning this term has acquired in the realm of politics (although current trends in international law would require considerable qualifications of this claim). It also seems to comply with the usage of the term "autonomy" in contemporary ethical debates. On closer inspection, however, even the current usage of the term "autonomy" reveals a fundamental problem.

Nowadays, we often use the word "autonomy" to indicate that an important decision, for example whether to have an abortion, should be taken by the pregnant woman herself, but we need to clar-
ify what we actually mean by this statement: We can, for example, mean that this is something which should not be dealt with within the range of positive, or more precisely criminal law. We can also mean that the woman is, morally speaking, free to decide for or against an abortion, but here we need to be cautious. In saying so, we are in fact making a moral claim, namely that abortion is allowed and therefore that the woman in question is morally free to decide. We do not, of course, mean that a woman is free to decide whether it is morally permissible to perform abortions. It would be a grave mistake to conclude, that since we are action-theoretically or metaphysically free to perform a certain course of actions, the principle of autonomy would automatically give the action moral clearance. As we shall see, interpreting Kant in this way amounts to a serious misunderstanding. Certainly for Kant, it is important that the agent as a free agent takes his or her own decisions, but this includes the possibility that the agent in his action might perform some action which is morally wrong and for which he or she is, for that matter, to be blamed.

Quite obviously, the crucial feature of Bittner's understanding of autonomy is that there must be an act by the agent (or the legislating body of a state) which turns something into a law for this agent. Consequently, he can ask whether there is a good reason to perform this crucial act concerning a possible candidate for such a law.

In doing so, however, Bittner fails to appreciate the core strategy of Allison and, in my view, misreads the core passage quoted above in which Kant explains his notion of "autonomy".

Allison's core strategy is the identification of moral correctness and full practical rationality against the background of the notion of transcendental freedom. Of course, we can imagine rational beings acting on maxims and thus acting on reasons and refusing to participate in the justification game. But although acting on maxims does in fact mean — in Kantian terms — to act on reasons,
it (acting on maxims simpliciter) is of course not sufficient for full practical rationality since the reason on which I act may fail the justification requirement. Furthermore, the question is not whether a (transcendently free) rational being is actually participating, but whether they can participate and can provide a justification. Allison’s claim that this is only possible if the maxims meet the UR remains unchallenged. Of course, since full rationality exhausts the reasons available there can be no further reason given why an individual should be moral. The question as to whether it is rational to be moral can be answered quite easily: yes of course, because it is one and the same thing. Internally, of course, the question arises as to how concern for full rationality can outweigh short-term considerations.

As far as the key passage (GMS AA IV, 440) in Kant is concerned, we need to take another look. Recall that “autonomy” for Kant is that property of the will by which it is a law to itself. Kant does not say, by which that property gives or can give a law to itself, but by which it is a law to itself. This indicates that for Kant there is in fact no requirement for an act which renders the moral law a valid moral law (for the agent). On the one hand, this no doubt complies with our conviction that it cannot be up to us to decide what the moral law is, nor that it is valid for us. On the other hand, however, it makes it quite difficult to understand why Kant can use the term “autonomy” after all, because “autonomy” is taken to mean “self-legislation”. To understand this, we need to understand in what sense the UR can be said to be due to self-legislation without there being an enactment to give it validity.

**Autonomy, the UR and the Self**

So far, it could be shown only that the UR is the means for a transcendently free being to retain full rationality. This, however, is ob-
viously not the same as self-legislation. If a law is a means to something else, the means is obviously not automatically due to any form of legislation of that something for which it is a means. Take for instance the law on driving on the left-hand side of the road in Japan. This law is evidently a means to solve a coordination problem, but it can hardly be claimed that the coordination problem in any sense legislated the law to drive on the left side. Nothing in the coordination problem indicates that this law is in any sense its law. However, these considerations give some indication of how we can understand the term "legislation" without there being any act on the part of the alleged legislator in question. If the law in question is somehow the expression of what the alleged legislator really or essentially is, we might be willing to accept the term "self-legislation" as appropriate, even if it may not fully comply with the contemporary usage of the term. This means, far more importantly, that Kant's autonomy presumably needs to be understood in metaphysical terms. According to Karl Ameriks (2000), this indeed provides the clue to Kant's intentions. In Ameriks's opinion, the fact that Kant did have this metaphysical conception of autonomy has been overlooked or ignored by those who think that the Kantian idea of autonomy has been enhanced only by the post-Kantian philosophers, or else by those who see the need to free Kant's philosophy from its un-Kantian elements, i.e. precisely from the metaphysical remnants of the ("dogmatic") past. In contrast to this widespread reading of Kant, Ameriks holds:

"This means that the fundamental 'self' that autonomously 'generates' the basic laws of morality is not defined as a human self, even though we actually become aware of the laws only as exemplified in concrete spatiotemporal, that is, human contexts. Neither the grounding of the laws nor their ultimate scope can be a matter of any kind of temporal, let alone spatiotemporal
action (such as consensus formation limited to the specific conditions of finite beings like us). Why then even bother to say that these laws are 'given' by us at all rather than that they are laws to which we must conform? The best answer, I propose, has to do with the fact that such laws are still not external to our essential nature, which for Kant is our sheer rationality" (Ameriks 2000: 13 sq.).

Indeed, interpreting the UR as an expression of our rationality and simultaneously the means by which rationality is made possible, preserves the notion of self-legislation, if we assume that we are essentially rational beings or rather rational selves. And yet, there is still a problem to be solved, since the following objection can be raised: it may be conceded, a critic may say, that we can interpret universality as an expression or manifestation of rationality. This idea seems to be confirmed, at least according to one recent, influential commentator, by the fact that Kant uses this very feature of rationality in his deduction of the right to property in his philosophy of law. Still, there is a considerable distance between the notion of universality and the notion of universalisability. Accordingly, we seem to find ourselves back at the charge of an illicit slide from the harmless notion of rationality in the sense of universal validity to the much stronger notion of collective rationality with which we dealt earlier. The question which now concerns us is how collective rationality can be an expression of what we essentially are as rational selves.

There is, I believe, a way out of the problem. The strategy must, it seems to me, be the following. We need to find a feature in the nature of rational selves which can warrant this very transition. And this feature, to my mind, is the essentially communitarian character of the self for Kant. Despite the widespread assumption which takes Kant's philosophy to be primarily individualistic, a rational
self is, essentially, a member of a community of rational selves. However, there is again no direct evidence for this in the text of the *Groundwork*. And yet, if we look at Kant's conception of a kingdom of ends, we realize that the very idea of the communitarian rational self lies at its foundation. Kant says:

"The concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will, so as to appraise himself and his actions from this point of view, leads to a very fruitful concept dependent upon it. ... For, all rational beings stand under the law that each of them is to treat himself and all others *never merely as means* but always *at the same time as ends in themselves*. But from this there arises a systematic union of rational beings through common objective laws, that is, a kingdom, which can be called a kingdom of ends (admittedly only an ideal) because what these laws have as their purpose is just the relation of these beings to one another as ends and means" (GMS AA IV, 433, trans. Gregor 41).

The German text actually says that the concept of a kingdom of ends is an "anhängende(r) ... Begriff" to that of a "rational being as a universal law giver", which can be taken to indicate that there is something in the intension of the latter which gives rise to the former.

This concludes the interpretation of Kant's argument in the *Groundwork* and so we can now turn our attention again briefly to the WTT.

**IV**

The limitations here of time and space, and simple courtesy to the busy reader, prompt me to offer only a rough outline of my main con-
conclusions regarding our original question: Did Aquinas really anticipate Kant’s notion of autonomy? In order to answer this question the following points need to be considered:

1) As we have seen, according to the interpretation given in the above, morality is understood as full practical rationality. This complies with recent trends in Aquinas scholarship, but it remains to be investigated whether the underlying conceptions of practical reason are the same in Aquinas and Kant. It seems quite obvious to me that these conceptions are different, and that Aquinas relies on a version of freedom which is very close to Kant’s practical freedom.

2) Kant and Aquinas both try to construe an objectivist conception of morality. In addition to this, both claim that by means of his natural faculties man has access to moral truths. The Kantian notion of autonomy is part of his strategy to establish this objectivist conception of morality, but at the same time it is designed to allow a very strong conception of the self as far as all rational beings with a will are concerned.

3) Aquinas’ metaphysics, characterised as creationist and theist, does not allow such a strong conception of the self as far as human beings are concerned. In line with his reliance on the conception of practical freedom, he construes his objectivist conception of morality in externalist and realist terms. The notion of creation is vital to the achievement of the necessary harmony of natural inclinations and morality. It is clear that Kant’s philosophy allows the possibility that human selves have indeed been created, but whether they have been created or not has no bearing on the functioning of Kant’s moral theory. For Aquinas, in contrast, the status of human beings as creatures is vital to getting his moral theory off the ground, at least as far as his moral epistemology is concerned.

4) Aquinas’s conception of the self is very close to the one dismissed by Kant in the passages in which he speaks of a “vormundschaftliche Natur”, or a tutelary nature (GMS AA IV, 425). In fact, howev-
er, this conception is far more sophisticated than the one Kant dismisses. Aquinas construes his doctrine of principles of practical reason in close parallel to the doctrine of theoretical reason. In both cases, created nature provides only the starting points for the principles. A full grasp of the principles is reached only when the agent acquires the virtues of the practical and theoretical intellect. Still, Aquinas cannot be played out against Kant on the ground that a virtue ethical account is superior to Kant. For Kant's ethics, the concept of virtue as a part of morality also plays an important role, but Kant uses a different conception of virtue to that of Aquinas. This also holds for his theory of action and for his anthropology.

5) Kant's strong conception of the self comes at a very high price. Its activities cannot be accounted for in natural terms; in Aquinas's thought, however, this is possible. It is far from clear, therefore, which conception of the self is to be preferred from a systematic point of view. Aquinas's conception may just strike the right balance between full dependence on and total independence of the naturally given.

6) The interpretation of Kant presented above is essentially a metaphysical one, and this not just in the technical sense of a "metaphysics of morals". This indicates that Kant's moral philosophy does indeed rest on metaphysical foundations. It is therefore not justified to consider Kant's ethics itself as autonomous in the sense of being independent of metaphysics. Ironically, the critics of the WTT sometimes think that this is one of the defining features of Kantian ethics and that, at least in this respect, Aquinas and Kant are indeed similar. The claim made by proponents and opponents of the WTT of an autonomous Aquinian ethics has been challenged in recent years. It has been proposed that Aquinas's moral theory is not independent of metaphysics, and it has also been said that it is not independent of revealed theology. In the case of metaphysics, I think this claim is true. In their dependency on metaphysics Aquinas's
and Kant’s ethics are indeed similar, but since their metaphysics is so different, this undercuts the very truth of the WTT. A detailed demonstration of this must, however, be held over for a future occasion.

Bibliography:

A) SOURCES:

B) SECONDARY LITERATURE:


Practical Reason and Morality