

On the Alleged Vacuity of Kant's Concept of Evil

by Pablo Muchnik, Loudonville/New York

In recent years, there has been a growing interest in Kant's doctrine of radical evil, arising from as diverse quarters as philosophy, psychoanalysis and the social sciences.¹ This interest has contributed to the revival of the notion of evil, which had been displaced from the center of philosophical discussion in the 20th century.² A common trait in the recent literature is that it takes the relevance of the use of the concept of evil for granted. Yet, before understanding what Kant really means by radical evil, it seems appropriate to ask first whether the notion of evil as such is necessary. For, given its religious background, this notion elicits expectations that may be incompatible with the secular concerns of moral philosophy. Such misgivings are aggravated by additional doubts about the concept's explanatory function in the first place. The pertinence of reviving a concept so allegedly flawed cannot be taken at face value. Thus, the task I set for myself here is to show the necessity of the concept of evil as it lies at the core of Kant's moral philosophy already in the *Groundwork*, far before he actually coins the notion of "radical evil" in the *Religion*. Whether this latter account is defensible or not, we will at least be sure that talk about an *evil* which is said to be *radical* is not in vain.

Let me begin by determining some of the reasons why the notion of evil seems dubious. I will then try to justify my proposal for overcoming those doubts.

¹ The English Translations I used are:

- *Critique of Pure Reason* (KrV), translated by Norman Kemp Smith, St. Martin's Press 1965. Quoted, as usual, A and B edition.
- *Groundwork of the Metaphysics of Morals* (GMS), translated by Mary Gregor, Cambridge University Press 1999.
- *Critique of Practical Reason* (KpV), translated by Lewis White Beck, Macmillan Library of Liberal Arts 1993.
- *Religion within the Limits of Reason Alone* (RGV), translated by Theodore Green and Hoyt Hudson, Harper Torchbooks 1960.

A good example of such crisscrossing of interests is *Radical Evil*, ed. by Joan Copjec, Verso, London-New York 1996.

² Cf. Otfried Höffe, Ein Thema wiedergewinnen: Kant über das Böse. In: *Über das Wesen der menschlichen Freiheit*, hrsg. von O. Höffe und Annemarie Pieper, Akademie Verlag, Berlin 1995, 11–34. Höffe argues for the necessity of retrieving the question of evil, which, he believes, was displaced from the center of philosophical debate in the 20th century, despite the many events that made its confrontation most urgent.

I. *Skepticism About the Notion of Evil*

Current writers in moral philosophy may share the distrust with which some of Kant's contemporaries received his doctrine of radical evil. This doctrine can be summarized in two main claims. (1) Evil corrupts the ground of all maxims, for it is rooted in the person's disposition (*Gesinnung*) – and is thus called “radical”. (2) Such corruption is not a matter of the individual but of the species as a whole – it holds universally, even in the best of men, for evil is a propensity (*Hang*) woven into the fabric of human nature. Goethe, focusing on the implications of this second claim, considers Kant's doctrine a “stain” in the mantle of critical philosophy, for it reintroduces religious “prejudices” (the doctrine of original sin), which Kant so carefully tried to supersede with his conception of autonomy. Schiller, focusing on the background of the first claim, considers it “scandalous”, for the corruption Kant sees at the basis of all maxims is nothing but a symptom of his dismissal of sensibility.³ Indeed, the notion of evil has a touch of Christian dogmatism, which the title of Kant's work seems only to confirm, and a smack of harshness, which renders it inaccurate in describing the character of average individuals. Thus, many could object, there is no need to entertain the notion of evil in moral philosophy, far less to assume that evil pertains to the very fabric of humanity. Instead, the notion of wrong is able to do the same philosophical job, but without awakening suspicion. By “wrong” is meant the set of immoral actions an agent decides to perform, i.e., immoral actions based on principles.

A possible reply to the suggestion of replacing “evil” with “wrong” is that it leads to the conflation of agent and action. For, while the notion of wrong denotes immoral actions and their principles, the notion of evil should be reserved for designating the person's disposition (*Gesinnung*).⁴ When we speak of an “evil action”, we do so figuratively and commit a kind of subreption.⁵ In a technical sense, “evil” is meant to designate (a) the (relatively) invariant form of the will in the various circumstances of action, and (b) the will's liability to a certain kind

³ Cf. Goethe's letter to Herder (June 7, 1793) and Schiller's to Koerner (February 22, 1793). These references are from Emil Fackenheim, *Kant and Radical Evil*. In: *University of Toronto Quarterly* 23, 1954, 340.

⁴ This view characterizes Kant's position in the *Religion*. RGV, AA 06: 20.22–34: “Man nennt aber einen Menschen böse, nicht darum weil er Handlungen ausübt, welche böse (gesetzwidrig) sind; sondern weil diese so beschaffen sind, daß sie auf böse Maximen in ihm schließen lassen. Nun kann man zwar gesetzwidrige Handlungen durch Erfahrung bemerken [...]; aber die Maximen kann man nicht beobachten [...]. Also müßte sich aus einigen, ja aus einer einzigen mit Bewußtsein bösen Handlung *a priori* auf eine böse zum Grunde liegende Maxime, und aus dieser auf einen in dem Subject allgemein liegenden Grund aller besondern moralisch-bösen Maximen, der selbst wiederum Maxime ist, schließen lassen, um einen Menschen böse zu nennen.”

⁵ The first in doing so is Kant himself, who often speaks of good and evil *actions*. For reasons evident below, I will consider this as a category mistake that applies to actions a predicate of persons.

of expression, which indicates what the agent takes as her ultimate reason for action. Just as we need a criterion to evaluate *decided* actions (in terms of right or wrong), we also need a criterion to evaluate persons' dispositions (in terms of good or evil), and these criteria, despite their evident connections, are not interchangeable.

Such a reply, however, will not satisfy those convinced of the vacuity of the notion of evil. They will argue that the proposed distinctions are quite artificial in Kantian ethics, whose unity of analysis is the agent's maxim. A maxim is a subjective principle of action, which expresses what the agent is moved to do and for what reasons. If we interpret maxims in a comprehensive way, namely, as presenting a rule of life that expresses the kind of person one wants to be, the set of wrong actions based on immoral maxims would account for the meaning reserved for "evil".⁶ For, the kind of maxim that makes a particular action wrong determines, at the same time, the immorality of the person – her life-policy or fundamental intention. A skeptic about the usefulness of the notion of evil concludes, we do not need a mysterious invariant form of willing, over and above particular wrong maxims, to account for an agent's immoral actions.⁷ A so-called "evil" person is simply one who repeatedly does wrong, and the meaning of such wrongdoing, taken as an expression of her disposition, is captured by the sum of immoral (wrong) maxims upon which those actions were decided. The continuum of intentions, from the more specific to the general life-policy (all implicit in the maxim), organizes conduct in a way that can be tested by the Categorical Imperative. Thus, there is no need for an *additional* criterion to evaluate the moral worth of persons and dispositions; one is enough for all purposes.

I believe that, although centuries apart, the discomfort of Kant's contemporaries and the line of thought I just described stem from a common root: they are based on a misunderstanding of the notion of a good will as the fundamental source of value in Kantian ethics.⁸ It is not by chance that the opening statement of the *Groundwork* is: "It is impossible to think of anything at all in the world, or indeed

⁶ Rüdiger Bittner interprets maxims as Lebensregeln. Cf., R. Bittner, *Maximen*. In: *Akten des 4. Internationalen Kant-Kongresses*, Mainz 1974, ed. by G. Funke, 485–498. Onora O'Neill has a similar position. She takes maxims to be fundamental intentions that control and guide our more specific intentions. Cf. *Constructions of Reason*, Cambridge University Press 1988, chap. 5, "Consistency in action", 84ff.

⁷ Here I am indebted to conversations with Rainer Forst, whom I want to thank for raising this point.

⁸ There have been many recent attempts to correct this misunderstanding. Worth mentioning are the works of Onora O'Neill (Op. Cit., particularly "Kant after Virtue", chap. 8, 145–162), Christine Korsgaard (*Creating the Kingdom of Ends*, Cambridge University Press 1996, particularly in "Kant's Formula of Humanity" [106–132], "Aristotle and Kant on the source of value" [225–248] and "Two distinctions in Goodness" [249–274]), and Barbara Herman (*The practice of Moral Judgment*, Harvard University Press 1993, particularly in "On the Value of Acting from the Motive of Duty" [1–22] and in "Leaving Deontology Behind" [208–240]).

even beyond it, that can be considered good without limitation except a good will."⁹ Nor can the importance of the good will in organizing the argumentative moves in the *Groundwork* be neglected. For, without giving the good will its due, we get a picture of Kant's morality as subordinating all concerns of value to those of duty. Yet, such a picture distorts the order he himself has envisaged. "We shall set before ourselves the concept of duty, which contains that of a good will though under certain limitations and hindrances".¹⁰ The good will is precisely without these limitations – a feature that allows it to function as the unconditioned condition that sets the structure of obligation and justification into work, giving actions out of duty the value they have.

My claim is twofold. (1) Only by reducing the *Groundwork* to mere deontology can the doctrine of radical evil be perceived as a betrayal of the principles of Kant's critical philosophy – as Kant's contemporaries did. (2) Only by detaching Kant's discussion of duty from the good will can the notion of wrong replace that of evil – as our contemporary skeptic argues. Assuming that the criterion for actions (right and wrong) is all there is to Kantian ethics disregards the difference between duty and the kind of will that gives duty its meaning and worth.¹¹

All these points may sound familiar to the current generation of Kant scholars. What I want to emphasize, however, is an implication of this reading that is frequently overlooked. For, by parity of reasoning, if the good will is the source of the goodness of right actions, we must conclude that an evil will is the source of the evilness (as the opposite value) of wrong actions. The value of the right or wrong action derives from the goodness or evilness of the will, whose ultimate form determines how to value them. I wish to argue, then, that Kant's quest for the source of corruption of all maxims in the *Religion* squares with his foundational project in the *Groundwork*: it fleshes out some of the prior assumptions regarding the source of value of actions and is demanded as a conceptual correlate of the notion of the good will.

⁹ GMS, AA 04: 393.5–7: "Es ist überall nichts in der Welt, ja überhaupt auch außer derselben zu denken möglich, was ohne Einschränkung für gut könnte gehalten werden, als allein ein *guter Wille*." Kant returns to this assumption in several places. Two instances are particularly revealing. One is the concluding remark in the discussion of the different formulations of the Categorical Imperative (GMS, AA 04: 437.5–9). The other is the transition from autonomy as a property of the will to autonomy as a principle of action (Ibid, 446f.). In this case, the good will is not mentioned but goodness is implicitly identified with the *property* of autonomy.

¹⁰ GMS, AA 04: 397.6–8: "[...] wollen wir den Begriff der *Pflicht* vor uns nehmen, der den eines guten Willens, obzwar unter gewissen subjectiven Einschränkungen und Hindernissen, enthält".

¹¹ Kant's deontology can indeed show that an action is right, but it is incapable of indicating *what* is good about it, i.e., *why* acting out of duty is itself good. Only the doctrine of the good will can answer such questions. In a nutshell, the answer is: dutiful actions are unconditionally good *because* they are the expression of a good will, which in turn is the unconditioned condition of all goodness.

Such a basic agreement does not mean, of course, that the positions in the *Groundwork* and the *Religion* are interchangeable. There is more to radical evil than the lateral references to an evil will in the *Groundwork* and it is undeniable that Kant's moral view and terminology undergo important changes.¹² I will not deal with these issues here. Rather, my intention has been to present some basic objections to the pertinence of using the notion of evil and trace their root to a reading of the *Groundwork* that gives a marginal role to the good will. By questioning such a tendency and placing the good will at the center of attention, I propose to sever the root of those objections. This alternative reading yields two consequences:

(1) It defends the notion of evil against those who want to replace it with the notion of the wrong. For just as the good will provides the key to understand what is good about right actions, the notion of wrong, by itself, could not reveal what is evil about acts of wrongdoing.

(2) It forestalls the attacks of those who see in the doctrine of radical evil a relapse into religious dogmatism. For, fleshing out the characteristics of the human will is not a concession to religious prejudices, but a demand posited by what was left undone at the time of establishing an a priori ethics in the *Groundwork*.¹³

In sum, if this view is correct, placing the good will at the basis of Kant's deontology shows that the notion of evil is entailed in the logic of the *Groundwork*. It represents the source of value of wrong actions and is necessary to characterize the immorality of a person (her disposition). The force of this view depends now on finding the appropriate relation between the criteria to evaluate rightness or wrongness of actions and goodness or evilness of persons. The rest of the paper is devoted to exploring this relation. If I here speak mostly of the good, it is because evil in the *Groundwork* operates, as it were, *in absentia*. After reconstructing the role the good will plays in the justification of actions, I will interpret the meaning of an evil will as its necessary correlate.

¹² For instance, while the doctrine of the fact of reason and the belief that the goodness or evilness of the will is itself chosen gain full swing in the *Religion*, they are absent in the *Groundwork*. These views set in motion the problem of justification of the choice of one's own noumenal character and disposition – a choice whose reason must remain inscrutable. Furthermore, in the *Religion* Kant introduces a systematic difference between the concepts of *Wille*, *Willkür* and *Gesinnung*, which can be read only retrospectively in the *Groundwork*. Finally, the doctrine of radical evil introduces a different unit of moral analysis, i.e., the human species as a whole, drastically expanding the prior concerns of Kant's practical philosophy, which typically revolve around the individual's morality.

¹³ I take Kant's moral works of the 1790's, particularly, the *Religion* and the *Metaphysics of Morals*, as examples of his ethics applied to the limitations of the human will (i.e., what Kant calls "anthropology" in the *Groundwork*). Such application entails a crucial enlargement of the scope of Kant's practical philosophy, which led him to the development of the twin doctrines of the highest good and radical evil.

II. The Criterion for Evaluating Actions

The most direct way to grasp why the Categorical Imperative is the criterion for evaluating the morality of actions is to analyze Kant's definition of the will (*Wille*): "the capacity to act in accordance with the representation of laws, that is, in accordance with principles".¹⁴ This definition brings together two different features: (a) a capacity that characterizes the human will and distinguishes it from other types of volition, and (b) the various types of relation to law such a will establishes by means of its representations. The first feature draws the limit for the imputation of actions, i.e., it describes the capacities an agent must have in order to be considered responsible; the second indicates the structure of justification, i.e., the conditions for the morality of the principles of action.

A. Imputation of Actions

Kant contrasts the peculiar capacity that characterizes the human will with two other kinds of volition that obliterate all traces of humanity: the animal will and the holy will. For brevity's sake, let me simply mention that, although based respectively on the incommensurable concepts of nature and freedom, the volitions characteristic of animality and holiness are identical with what the law commands. I call this form of willing "homogeneous", because the object of volition necessarily coincides with the demands of the law.

The human will, instead, is characterized by a *structural heterogeneity*. Kant argues:

If reason solely by itself does not adequately determine the will; if the will is *exposed* [*unterworfen*, i.e., submitted] also to subjective conditions (certain incentives) that are not always in accord with the objective ones; in a word, if the will is not in itself completely in conformity with reason (as it is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is *necessitation* (*Nöthigung*). That is to say, the relation of objective laws to a will that is not thoroughly good is represented as the determination of the will of a rational being through grounds of reason, indeed, but grounds to which this will is not by its nature (*Natur*) necessarily obedient.¹⁵

¹⁴ GMS, AA 04: 412.27–28: "das Vermögen, *nach der Vorstellung der Gesetze*, d.i. nach Principien, zu handeln".

¹⁵ GMS, AA 04: 412.35–413.8, my italics: "Bestimmt aber die Vernunft für sich allein den Willen nicht hinlänglich, ist dieser [der Wille] noch subjectiven Bedingungen (gewissen Triebfedern) unterworfen, die nicht immer mit den objectiven übereinstimmen; mit einem Worte, ist der Wille nicht *an sich* völlig der Vernunft gemäß (wie es bei Menschen wirklich ist): so sind die Handlungen, die objectiv als nothwendig erkannt werden, subjectiv zufällig, und die Bestimmung eines solchen Willens, objectiven Gesetzen gemäß, ist *Nöthigung*; d.i. das Verhältnis der objectiven Gesetze zu einem nicht durchaus guten Willen wird vorgestellt als die Bestimmung des Willens eines vernünftigen Wesens zwar durch Gründe der Vernunft, denen aber dieser Wille seiner *Natur* nach nicht nothwendig folgsam ist." It is

The human will is *heterogeneous* because the cognition of something as good is not always sufficient reason to act on it. Such insufficiency must be understood as a result of the will's being affected by impulses (*Antriebe*) of sensibility, which are precisely the forces determining the animal will and absent in the holy will. The human will is sensuous, for it is *pathologically* affected by impulses; yet, it is at the same rational, for it is moved to action only by the representation of laws. This internal heterogeneity is *structural*, for, were sensuous impulses capable of determining the will directly, it would become animal, and were rationality always sufficient to move it, it would become holy.

The structural heterogeneity of the human will, marked by the difference between being exposed and being determined by sensuous impulse, is the condition for the possibility of imputation and necessitation. Of imputation, because the objective conditions for the goodness of an action are subjectively contingent; of necessitation, because rational grounds do not forfeit their claim on the will, even if, prompted by sensibility, it does not actually respond to them. Let me spell out some implications of this view.

A.1 Necessitation

Kant believes that the validity of rational grounds does not depend on the agent's actual recognition of them. "Necessitation" reveals a twofold aspect in the relation of a heterogeneous will to the law: (a) the independence of the law's *validity* from the agent's recognition, and (b) the necessity that such validity acquire *authority*, i.e., that the demand of the law be fully acknowledged and carried out in action.

Although both the holy and animal will stand under laws, they are not necessitated (*genöthigt*) to action. To be necessitated implies that the command of the law appears as an '*ought*' (*Sollen*). The imperative indicates the goodness of an action whose performance is not guaranteed by the cognition of such goodness.¹⁶ This uncertainty reveals the unique character of the good *presented* by reason. While its validity is independent from the agent's recognition, it can become authoritative, i.e., an effective determining ground of action, if, and only if, the agent *represents* it as being good *for her*. Kant believes that in its *presentation* the good posits a demand for recognition in the agent, which can be fulfilled only if its *representation* is

worth noticing Kant's peculiar use of the word "nature" (*Natur*) in this sentence, which presents an ambiguity comparable to the one in the statement "*Der Mensch ist von Natur böse*" in the *Religion* (RGV, AA 06: 32.11). If "nature" meant here "the sum of objects of experience", as it does in the *Critique of Pure Reason*, Kant's statements would make no sense. For, neither evil nor the contingent obedience to the grounds of reason would be then imputable.

¹⁶ Imperatives do not apply to the holy will, for whom the '*ought*' (*Sollen*) expressed by the rational law is always already a volition (*Wollen*) (GMS, AA 04: 414), nor to the animal will, for whom the command of reason remains completely alien (KrV, A 548/B 576).

such that the agent accepts the good *because* of its validity. The good in question may be the moral good or the useful, for the “ought” simply indicates that objective demands are subjectively contingent.¹⁷

Since Kant severs authority from validity, in order to become effective the good must produce an *interest* in the agent. Kant defines “interest” (*Interesse*) as “the dependence of a contingently determinable will on the principles of reason”.¹⁸ Being rational, these principles cannot be imposed upon the agent; rather, they must be the expression of the agent's own *spontaneous* activity. And, since the human will is not automatically in conformity with reason, the good is only contingently contained in the agent's principles. Such contingency is explained by the fact that human will is exposed (*unterworfen*) to two different sources of interest, sensible impulses and the demands of reason, each one expressing an essential feature of its heterogeneous structure. Neither of these sources is sufficient to determine the will, but each presents a candidate ground to determine action. Kant calls such grounds “incentives” (*Triebfedern*), which can actually settle a course of action if, and only if, the will actually takes them to serve as determining ground (*Bestimmungsgrund*). Once the incentive is incorporated into the ground of action, it is called a motive (*Bewegungsgrund*). A motive is an incentive that has been taken as determining ground, thereby impeding the other candidate from occupying the position of determination. Since in a heterogeneous will no incentive is by itself determinative, the agent is always accountable for what she takes to be motivating. Hence, her motives are always indicative of where she places value.

It is important to note that, although the agent sets both the motive and the end of an action, these rarely coincide.¹⁹ The motive expresses the type of interest (sensible or rational) which an agent takes to be her determining ground, whereas the end is the state of affairs the action intends to bring about. A given end, say, helping a person in distress, may have different motives: compassion, prudence, duty, reputation, and so on. Just as these motives affect the conception of the action and the way in which the intention is carried out, the same end can be based on multiple motives: the relation here is not a one-to-one correspondence. Such non-coincidence is an essential feature of Kant's theory of action, and indicates the need for a systematic distinction between different functions of the concept of “duty” – a point I will develop below.

¹⁷ This view of normativity is usually referred to as “internalist”. It holds that the reason why the action is presented as good is a reason for doing it. My point is that Kant's internalism comprises not only categorical imperatives, as is normally understood, but also hypothetical imperatives, for it is embedded in the structural heterogeneity he assigns to the human will. Cf. C. Korsgaard, “Kant's Analysis of Obligation: The argument of Groundwork I”, *Op. Cit.*, 43.

¹⁸ *GMS*, AA 04: note 413.27–28: “[d]ie Abhängigkeit eines zufällig bestimmbaren Willens [...] von Principien der Vernunft”.

¹⁹ They do so only in the special case in which the end of an action is also a duty. The clearest example in the *Groundwork* appears in the discussion of beneficence (Section I), as I will argue below.

Now, the human will is affected by the two kinds of incentives that put it at a crossroad when choosing its motive. According to Kant, such under-determination requires the agent to reflect upon the validity of her reasons for action. Thus, Kant's view of necessitation draws upon his theory of justification, the second feature in his definition of the will.

B. *The structure of Justification*

Kant's distinction between maxims and laws provides the general framework of this structure and introduces a set of interlocking conceptions of the good. In order to appreciate how these conceptions originate and relate to our previous discussion, let me begin by drawing the main lines of the justificatory framework.

B.1 Maxims and laws

The structural heterogeneity that makes imputation and necessitation possible also requires the distinction between *maxims* and *laws*.

A *maxim* is the subjective principle of action, and must be distinguished from the *objective* principle, namely the practical law. The former contains the practical rule determining reason conformably with the *conditions of the subject* (often his ignorance or also his inclinations), and is therefore the principle in accordance with which the subject *acts*; but the law is the objective principle *valid for every rational being*, and the principle in accordance with which he *ought to act*, i.e., an imperative.²⁰

Since the representation of laws from which the action follows is the distinctive trait of the will (according to Kant's definition), there arises a systematic distinction between the object that is *presented* (the law or objective good) and the way in which it is *represented* in the maxim.²¹ Kant assumes that the objective component in the representation provides an *immanent* standard for evaluating the accuracy of the subjective one. Thus, in order to be objective, the subjective principle must rest on *grounds valid for every rational being*. Let me call this the *demand for universality*. Furthermore, in order to be objective, the subjective principle must *be free from the restrictive conditions of subjectivity*, which Kant identifies with ignorance

²⁰ GMS, AA 04: 420.36–421.30 note; “*Maxime* ist das subjective Princip zu handeln und muß vom *objectiven Princip*, nämlich dem praktischen Gesetze, unterschieden werden. Jene enthält die praktische Regel, die die Vernunft *den Bedingungen des Subjects gemäß* [my emphasis] (öfters der Unwissenheit oder auch den Neigungen desselben) bestimmt, und ist also der Grundsatz, nach welchem das Subjekt *handelt*; das Gesetz aber ist das objective Prinzip, *gültig für jedes vernünftige Wesen*, [my emphasis] und der Grundsatz, nach dem es *handeln soll*, d. i. ein Imperativ.”

²¹ Such a difference is negligible in the holy will, whose maxims are always lawful, and is absent in the animal will, which does not represent the law and has no maxims.

(i.e., cognitive limitations) and inclinations (i.e., pathological limitations). Let me call this the *demand for unconditionality*.²²

To understand how these demands work, we need to keep in mind that for Kant every action necessarily has an end (GMS, AA 04: 427–428), and that “rational nature is distinguished from the rest of nature by this, that it sets itself an end”.²³ The *demand for universality* concentrates on the *validity of the grounds* (motives) upon which the action is willed, leaving the content of ends undetermined, whereas the *demand for unconditionality* focuses on the subjective relation to one's own rationality, thereby guaranteeing the *validity of ends themselves*. This indicates a basic difference in the orientation of the two demands. Whereas *universality* is oriented toward the rationality of other beings as a check upon one's grounds for action, *unconditionality* is oriented to one's own rationality as a check upon subjective limitations. This latter is capable of advancing conditions to determine the end and making it objective, i.e., acceptable for all agents. In both cases rationality imposes a restriction: the *demand for universality* requires the agent to enlarge her perspective by including the view of all other agents, while *unconditionality* demands her to identify herself with her rational nature.

The difference in scope and orientation of the two demands is due to Kant's theory of action, in which motives and ends do not generally coincide.

B.2 Universality

The *demand for universality* presupposes the agent's capacity to recognize the validity of grounds and act on them, i.e., the capacity to subordinate the manifold of her desires to the constraints of reason. Since the content of the end is not specified, the grounds for adopting it may stem from empirical or from pure practical reason. Hence, this demand is the source of both hypothetical and categorical imperatives.

Setting aside the details of the discussion of hypothetical imperatives, we can concentrate on how the agent represents herself here as the cause of actions whose ends are not evaluated in the process of deliberation. Whether she wants to lose weight, construct a bomb, or solve a mathematical problem, deliberation focuses only on finding the most appropriate means to attain these ends. At this level, the demand for universality determines the acceptability of means, whose “goodness” (i.e., usefulness) can be recognized by all agents who happen to share the same end. The value of means is conditional, for it presupposes the desirability of ends reason does not itself determine.

²² The reading of this passage is inspired by B. Herman's “On the Value of Acting from the Motive of Duty”, *Op. Cit.*, 6–17.

²³ GMS, AA 04: 437.21–22: “Die vernünftige Natur nimmt sich dadurch vor den übrigen aus, daß sie ihr selbst einen Zweck setzt.”

If, on the other hand, the action as such is represented as “good” (i.e., right), not as a means to an end, but as being itself necessary and containing value within itself, then the ground for adopting the end must itself be unconditioned. *Universality* determines here the value of ends whose goodness could not be assessed by hypothetical imperatives. What makes the action right is the universalizability of grounds, which all rational agents can recognize as inherently valid – whether they share the end or not (GMS, AA 04: 400). For, *universality* incorporates the perspective of all other agents within one’s own reasons for action, limiting the arbitrariness of one’s will and the urgency to act in self-serving ways.

Kant insists that only the motive of duty can produce such a limitation. Take, for instance, the example of the shopkeeper in *Groundwork I* (GMS, AA 04: 397.21–32). Although the action of not overcharging inexperienced customers preserves the appearance of morality, it rests merely on prudential considerations (a variant of self-love). In the merchant’s business community, it happens that preserving one’s reputation for honesty is more effective business policy than pursuing profit indiscriminately. The shopkeeper is reasonable enough to check the temptation of taking advantage of his current customer, and, in doing so, implicitly consults the view of those who want to succeed in business. Yet, this arbitrary restriction in the scope of rationality presents a twofold problem. (I) It makes his action contingent. Were the merchant capable of overcharging customers with impunity, were he part of a community in which the reputation of honesty had ancillary value, or thought his customer belonged to a class of people who did not deserve moral consideration, the shopkeeper would drop at once any pretence of decency. The infirmity of his conduct lies in the fact that he has restricted the demands of rationality to the preferences of like-minded agents, turning universality into mere generality. Furthermore, (II) by neglecting the overarching perspective of other agents, the shopkeeper forfeits the capacity to evaluate the morality of his own actions. Having embraced a given set of values, his only concern is to find effective means to attain them, disregarding the question of the acceptability of any such pursuit for other agents. In his mode of deliberation, questions of goodness become mere matters of efficiency.²⁴

This example shows that only the motive of duty can determine the morality of actions, despite their legality and contingent beneficial consequences. Unlike self-love, which is in principle restricted to one’s own perspective (or at most comprises

²⁴ Changing circumstances, it is not difficult to imagine that our shopkeeper would act like Adolf Eichmann, who, according to Arendt, “thoughtlessly” confounded the general law of the land with the law of morality. It was “his inability to think, namely, to think from the standpoint of someone else” that allowed him to thrive in the Nazi Party and encouraged his uncanny sense of self-righteousness. Cf. H. Arendt, *Eichmann in Jerusalem. A Report on the Banality of Evil*, Penguin Books, N.Y. 1994, 49 and 136. Allison makes a similar point. Cf. H. Allison, “Reflections on the Banality of (Radical) Evil. A Kantian Analysis”. In: *Idealism and Freedom. Essays on Kant’s Theoretical and Practical Philosophy*, Cambridge University Press 1996, 170ff.

that of like-minded agents), the motive of duty requires us to evaluate subjective preferences in terms of their acceptability to other agents. Note, however, that duty does not determine the *content* of our ends – it only requires us to limit the grounds upon which contingent ends are pursued, forcing us to overcome the idiosyncrasies of self-love. Such a function is expressed in the first formulation of the Categorical Imperative: “Act only in accordance with that maxim through which you can at the same time (*zugleich*) will that it become a universal law.”²⁵

The adverb “*zugleich*” is the key here. It indicates that what must be universalizable is neither the action nor what is aimed at by it (the end), but the ground (motive) upon which the compound is decided to be good. The action and its end are like placeholders whose value gets determined by their relation to the ground. The value depends on whether the agent's rational nature, expressed by the perspective of universality, determines the ground upon which the action is performed. Such determination guarantees the acceptability of a contingent action by all agents.²⁶

This model illustrates one of the senses of what Kant calls *acting out of duty*: pure practical reason, which determines the acceptability of the ground, relies here upon ends whose content is given by empirical practical reason. The motive of duty works as a *limiting condition* on the agent's adoption of ends in general, for it establishes the terms for which such an adoption is valid. Duty is not the end of an action, but the reason for performing it. The Categorical Imperative requires the agent to discard actions whose grounds cannot be universalized, but allows her to act on all other cases. The content of particular duties is thus determined *indirectly*. The failure of the universalized maxim yields a duty by requiring the agent to perform the contrary action she originally intended to. All other compounds are considered permitted. The pursuit of permitted ends is legitimate *because* the claims of empirical practical reason, whose systematic arrangement falls under the notion of happiness, are previously submitted (*unterworfen*) to the control of morality. Their goodness is thus conditioned by the goodness of acting out of duty, which functions as the limiting condition for adopting ends in general.

This *indirect* derivation of duties suggests the presence of two different levels of commitment on the side of the agent. (1) A commitment to perform actions whose compound (relation of ground and end) is universalizable or to perform the contrary action when the proposed compound fails the test. (2) A commitment to make consistency in the *ground-end relation* the general criterion to select among maxims. The first one determines the agent's performance of certain types of action; the second determines the adherence to a certain criterion in the selection of maxims. Thus, in the logic of Kant's argument, morality concerns

²⁵ GMS, AA 04: 421.7–8: “handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde.”

²⁶ This interpretation blocks the charge of empty formalism, common in the literature since the time of Hegel. For, since every action has necessarily an end, content is always already present in the formal procedure evaluating the maxim of action.

both the process of choosing actions and the process of choosing principles of action. It is the second level of commitment that justifies Kant in his belief that he can derive duties from the Categorical Imperative, i.e., from the fundamental commitment to the motive of duty.

My purpose in introducing these distinctions is to make explicit that, according to the process of justification in the *Groundwork*, the goodness (or evilness) of an agent must be expressed at the second level of commitment – a point that becomes fully evident in the *Religion*. For, at this level, the will discloses the character of its fundamental valuations and exhibits what the agent takes as a valid reason for choosing maxims of action. Hence, in order to understand the relation between Kant's criteria to evaluate persons and actions we must make explicit his reasons to associate duty and self-love with the goodness or evilness of an agent's will. I will return to this point below.

B.3 Unconditionality

Unlike the *demand for universality* that focuses on *grounds* and leaves ends undetermined, the *demand for unconditionality* has a more restrictive scope. It applies to actions whose ends are themselves duties, i.e., ends whose goodness depends on a direct *relation* to the agent's own rationality. In actions of this kind, the content of ends and motives is required to coincide. Such a requirement advances conditions on the character of ends themselves, whose content pure practical reason left undetermined in the *demand for universality*.

This interpretation rests on a certain understanding of Kant's phrasing, "conformably with the conditions of the subject", in his definition of the principle on which the agent *acts* (i.e., her maxim).²⁷ This has frequently been interpreted as positing the negative version of the requirement to adopt a principle "valid for every rational being" (*ibid.*), which I identified as the *demand for universality*. Although such an interpretation is plausible, it overlooks the consequences, in terms of justification, of the different ways in which ends and motives can relate in an action. To account for this, I propose to take *unconditionality* as a different demand, holding for *ends* and their relation to the agent's *own* rationality. Such a demand complements the one we found in the first formulation of the Categorical Imperative. While the latter determines the value of actions without reference to the coincidence between the content of ends and their motives, the former is needed to posit the conditions for the validity when they coincide. By this means, Kant's structure of justification can exhaust the possible combinations of ends and motives in an action – an advantage that would be abandoned if we interpret both demands as equivalent.

²⁷ Quoted above: GMS, AA 04: 421.26.

I take the new criterion to complement the conditions presented by the first formulation of the Categorical Imperative, which provides the standard for the value of actions whose ends are not determined by pure practical reason. Such a relation is possible because the *demand for universality* is not concerned with the subjective genesis of ends, but with their *consistency* with the conditions holding for the *ground-end relation*. This condition assures the validity of contingent actions for all rational agents, but leaves open the possibility for the *demand for unconditionality* to set conditions on the subjective origin of ends, so that their content becomes itself necessary. Since *unconditionality* requires ends to be adopted independently of the agent's limitations, it imposes constraints on what counts as the agent's valid self-relation in setting ends as such. The genetic conditions are by definition *unconditioned*: an end *ought* to be adopted *because* of its intrinsic rational character.

Empirical practical reason has no place in this new configuration. Yet, among the set of actions that are done out of duty and show that pure reason is practical, I interpret the *demand for unconditionality* as applying to actions whose ends are also dutiful. This unique kind of end is not simply *consistent* with the ground of its adoption when universalized – it is *identical* with it.

For Kant, a maxim expressing this kind of end has *moral content* (GMS, AA 04: 397.36–398.1). Maxims with moral content arise *out of duty* (the ground) and have *a duty as their end*. Let me call the content of the end “objective” in order to distinguish it from all other contingent or “subjective” ends, which are prompted by empirical practical reason.²⁸ Whereas the *demand for universality* determines the validity of subjective ends by checking on their grounds, i.e., *indirectly* and drawing upon empirical practical reason as a source of ends, *unconditionality* sets the condition of their validity *directly*. Namely, it stipulates that ends are objectively valid if their adoption does not depend on the agent's limitations (cognitive and pathological), a condition that only pure practical reason having itself as an object can fulfill.

The implication of acting *out of duty* is different in each case. When duty is understood as a limiting condition operating on the *grounds* for the adoption of ends, pure practical reason checks the ends of empirical practical reason and requires the agent to act only on ends that are *consistent* with the maxim's universalizability. This results in the determination of what Kant calls perfect duties and permitted actions. When, on the other hand, duty provides the ground and also the content of the end of action, pure reason has itself as an object, checking the agent's subjective limitations in order to raise them to the standard of pure rationality. This specifies the duties Kant calls “imperfect”, which, in the *Metaphysics of Morals* will turn out to be duties connected with seeking one's perfection and the happiness of others. Imperfect duties are those whose ends are a duty to have. The agent's commitment here is to unconditionally adopt unconditional ends. In contrast with per-

²⁸ Cf. GMS, AA 04: 428.25–27. Kant's distinction in the *Groundwork* differs from mine. I am adopting here the terminology of his later works. Cf., for example, RGV, AA 06: 6f., note.

fect duties, which command a specific action, imperfect duties command a form of willing, leaving its application to the agent's own judgment.²⁹

The complementary relation of the demands of *universality* and *unconditionality* points to a connection between two of the main formulations of the Categorical Imperative. As the so-called formula of universality expressed the condition for the valid *ground-end* relation, the formula of the end-in-itself expresses the conditions for the possibility of the *end-subjective condition relation*. This is so because the formula of end-in-itself is tailored to make more intuitible the *matter* of the moral law, and the material of volition is precisely what the *demand for unconditionality* regulates. It requires that the setting of subjective ends be measured by the capacity for setting ends as such, which Kant identifies with our rational nature (GMS, AA 04: 437.21–22). To treat humanity also as an end-in-itself means to comply with the conditions that allow an agent to reflect upon her capacity for setting ends us such. In this reflection, rational nature posits itself as its own object. This expresses the *positive role* of the end-in-itself formula and accounts for the genesis of imperfect duties.³⁰

B.3.1 The Friend-of-Man

To bring the discussion of *unconditionality* home, let me analyze Kant's example of the sympathetic individual (*Menschenfreund*, the friend-of-man), whose actions present all the appearance of being morally good – for “to be beneficent where one can is a duty”.³¹ Furthermore, “without any other motive of vanity or self-interest [the friend-of-man] find[s] an inner satisfaction in spreading joy around [him]”.³²

²⁹ Just as the goodness of the will varies in perfect and imperfect duties according to the different meanings of *acting out of duty*, the immorality of actions must reveal the evilness of the will according to the different types of transgression. That is, the meaning and implications of the evilness of an agent's will must depend on whether her actions disclose a failure in her commitment to duty as a *limiting condition* or to duty as *giving moral content* to maxims. This is a conclusion Kant himself does not seem to have acknowledged. I develop this point in my dissertation, *The Consistency of Kant's Doctrine of Radical Evil* (unpublished).

³⁰ The end-in-itself formula also has a negative role, for it works as a *limiting condition*: “[...] so wird der Zweck hier nicht als ein zu bewirkender, sondern selbständiger Zweck, mithin nur negativ gedacht werden müssen, d. i. dem niemals zuwider gehandelt [...] werden muß.” (GMS, AA 04: 437.26–30). This implies that an agent must restrict her ends in actions involving other agents so that they “jederzeit [...] von eben derselben Handlung auch in sich den Zweck müssen enthalten können” (GMS, AA 04: 430.7–8). The restrictive character of the end-in-itself formula is what it shares with that of universality and justifies Kant in considering them as interchangeable. An action whose end contains its possible acceptance by all other agents must be based on “einer Maxime, die ihre eigene allgemeine Gültigkeit für jedes vernünftige Wesen zugleich in sich enthält” (GMS, 437.36–438.1).

³¹ GMS, AA 04: 398.8: “Wohlthätig sein, wo man kann, ist Pflicht”.

³² Cf. *Ibid.* 9–11: “ohne einen andern Bewegungsgrund der Eitelkeit, oder des Eigennutzes” the friend-of-man find[s] “ein inneres Vergnügen, [...] Freude um sich zu verbreiten”. I have altered the quotation from plural to singular. Agreeing with Herman, I find that the clue to

If we disregard the agent's subjective relation to the end, the lack of vanity and self-interest in his motivation seem to guarantee the universalizability of his maxim and exhaust what Kant means by acting out of duty. Yet, Kant concludes that despite its dutiful and lovable appearance, the action has no moral worth and stands at the level of other inclinations – particularly the inclination to honor (ibid.15–16). This conclusion seems to disqualify the morality of the action altogether: inclinations yield only subjective grounds, which cannot possibly be acceptable to other rational beings.

The problem disappears if we interpret Kant's example in terms of the *demand for unconditionality*. The action is questionable not because of a straightforward deficiency in its grounds, which are universalizable (disinterested and free from vanity), nor is it because of the intrinsic nature of the end, which deserves praise and encouragement. Rather, the question lies in the *subjective relation* the agent establishes with the goodness of such an end. The end is adopted “because of the inner satisfaction in spreading joy around” (my translation), which makes it depend upon the subject's state (*Zustand*) in performing the action. Hence, the adoption of the end is conditional. That is why Kant compares beneficence with honor (*Ehre*). Just as the latter is praiseworthy “if it fortunately lights upon what is in fact the common interest and in conformity with duty”³³, it may well bring about frightful consequences if it falls upon dubious ends. Kant's point is that the worth of beneficence in the example can only derive from having been adopted not from inclination but *from duty*.³⁴

The friend-of-man relates to the objective end of beneficence depending on the contingency of his subjective limitations. His helpfulness towards others is conditioned by his moral luck (his sympathetic *temperament*), a contingency that, in turn, diminishes his capacity to evaluate the moral context of the action. For, indiscriminate beneficence may promote iniquity in the world.³⁵ Imagine the case of somebody in a serious predicament, but intending to commit an awful crime. The sympathetic individual would lack an internal criterion to distinguish between the request for help of the criminal, who is about to be captured by the police, and those of her victims, who are also very much in trouble. Since the friend-of-man's reasons for helping depend on his pleasure in doing so, any person caught in a serious predicament would awake his sympathy and match equally well his motivational structure.

understand the example is to see that “the souls so sympathetically attuned” (in the first part of the example) and the “person of cold temperament” (in the second part) refer to the same individual. I use the masculine to preserve the German “*Freund*”. Cf. B. Herman, “On the value of acting on the motive of duty”, Op. Cit., 18.

³³ GMS, AA 04: 389.16–17: “wenn sie glücklicherweise auf das trifft, was in der That gemeinnützig und pflichtmäßig, mithin ehrenwerth ist”.

³⁴ This is Korsgaard's reading. Cf. “Kant's Analysis of Obligation: The argument of Groundwork I”, Op. Cit., 59–61.

³⁵ I follow B. Herman on this point. Cf. B. Herman, Op. Cit., 5ff.

In other words, the actions of the friend-of-man express a twofold limitation: the objective end is adopted conditionally, and there is no criterion to discriminate among candidates for beneficence. These shortcomings exemplify the pathological and cognitive limitations that *unconditionality* requires a maxim to overcome. Only when the friend-of-man is “overclouded by [her] own grief, which extinguished all sympathy with the fate of others”³⁶ – only when we imagine his temperament as “cold and indifferent” can he relate to the end of his action independently of the contingencies of his own state. This independence allows him to judge the moral import of claims for help, and to act only on cases that are consistent with the Categorical Imperative. Thus, a plea for help becomes a reason for action if, and only if, the agent’s response does not further immorality in the world and is given *unconditionally*.

Note that the circumstances of beneficence allow considerable leeway for judging the particular situation. Duty does not require a concrete action (as in the example of the shopkeeper or any other case of direct duty), but a form of willing which triggers reflection upon the adoption of objective ends and the pertinence of others’ requests for help. As Kant puts it above, *unconditionality* demands us to overcome the limitations connected with “the conditions of the subject” (my translation) (GMS, AA 04: 421.note), i.e., pathological and cognitive limitations. The nature of an agent’s will shows itself through her commitment (or lack thereof) to certain deliberative guidelines about how to act in the particular circumstances. Just as in the case of the first formulation of the Categorical Imperative, such a commitment rests at a more fundamental level and provides the standard to gauge the value of particular actions. It is now time to analyze it.

III. *The Criterion for Evaluating Persons*

The structure of justification shows that the goodness of particular actions (their rightness) is due to their complying with the conditions of *universality* and *unconditionality*, displayed by the agent’s mode of deliberation. These are *immanent* conditions implied in the definition of the will (GMS, AA 04: 412.26–30). Yet, what is Kant’s argument for us to accept this set of conditions? How does he guarantee that, while grounding the goodness of particular actions, the conditions themselves are not ungrounded?

These questions are not incidental to the logic of Kant’s argument. They are inscribed in what he assumes to be the nature of human reason, which finds satisfaction only in reaching the unconditioned. This assumption indicates the direction of Kant’s answer. It consists in the notion of the good will, the rock bottom upon which the process of justification rests. The good will complies with the uncondi-

³⁶ GMS, AA 04: 398.21–22: “[...] vom eigenen Gram umwölkt, der alle Theilnehmung an anderer Schicksal auslöscht”.

tional demand of the moral law in all its volitions, for it is good “unconditionally” and “without limitation”.³⁷ Since the human will is good insofar as it shows the good will in its actions, i.e., it is good *on the condition* that it has a good will, then the structure of justification to which every rational agent appeals must be grounded precisely on such a good will.

The assumption upon which the good will acts is that rational nature confers value and that rationality is itself the source of all value. Rationality is the source of value, for it is the only property that can withstand the demand for unconditionality. This does not mean that the conditioned values of sensibility are disqualified, but that they are incapable of bringing closure to justification. Instead, they give rise to an infinite series of further conditions, without ever reaching the unconditioned, which Kant identifies with rationality itself. As he puts it in the second Critique: “The concept of good and evil is not defined prior to the moral law, to which, it would seem, the former would have to serve as foundation. [R]ather, the concept of good and evil must be defined after and by means of the law.”³⁸ The logical pre-existence of the law with respect to its object generates a “paradox of method” (Ibid, 62.36) that introduces in morality a turn comparable to the Copernican revolution in metaphysics.³⁹ Kant names such a turn “autonomy”, which indicates “the property of the will by which it is a law to itself (independently of any property of the objects of volition)”.⁴⁰

An autonomous will renounces all interest arising from given objects and commits itself to its rational nature as the source of all value. Thus, such a will is transcendently free. As Kant tries to demonstrate in *Groundwork III* and assumes as a *fact of reason* in the Analytic of the second Critique, a will with the property of autonomy would express the moral law in all its maxims. Since a being “that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect”,⁴¹ it follows that human agents are *obligated* to comply with the conditions of autonomy. Thus, autonomy, besides designating the property characteristic of a good will, without which the structure of justification would be ungrounded, must also

³⁷ Cf. GMS, AA 04: 437.6 and 393.6.

³⁸ KpV, AA 05: 62.37–63.4: “[...] daß nämlich der Begriff des Guten und Bösen nicht vor dem moralischen Gesetze (dem er dem Anschein nach sogar zum Grunde gelegt werden müßte), sondern nur (wie hier auch geschieht) nach demselben und durch dasselbe bestimmt werden müsse.”

³⁹ The turn could be phrased thus: the conditions for the possibility of moral objects (the good and the evil) is given “after and by means of” the rational law that makes their experience possible. Or, alternatively, good and evil do not determine the content of the moral law; rather, it is the structure of the law that makes those objects possible and determines our conceptions of them.

⁴⁰ GMS, AA 04: 440.16–18: “Autonomie des Willens ist die Beschaffenheit des Willens, dadurch derselbe ihm selbst (unabhängig von aller Beschaffenheit der Gegenstände des Willens) ein Gesetz ist.”

⁴¹ GMS, AA 04: 448.4–6: “Ein jedes Wesen, das nicht anders als *unter der Idee der Freiheit* handeln kann, ist eben darum, in praktischer Rücksicht, wirklich frei”.

provide a principle for the volitions of a structurally heterogeneous will. “*The principle of autonomy is [...] to choose only in such a way that the maxims of your choice are also included (zugleich) as universal law in the same volition.*” [my emphasis]⁴²

Thus, in Kant’s argument, “autonomy” is called to fulfill a twofold role: it makes known the *property* of the good will and it provides the *fundamental principle* guiding all the volitions of a structurally heterogeneous will. As a property, autonomy expresses the *unconditioned condition* grounding the structure of justification. As a principle, autonomy provides the criterion to evaluate the agent’s will as a whole. This is because it expresses the rule for assessing the rationality of an agent’s mode of deliberation, i.e., for evaluating the way in which she adopts maxims in general. Thus, Kant calls autonomy “the supreme (*oberstes*) principle of morality”. The fact that he places it along with the other formulations of the Categorical Imperative should not obscure its distinctive status. Whereas the phrasing of the formulae of universality and end-in-itself has *actions* in mind, for it requires the agent to “*act such that [...]*” autonomy bids us to “*choose such that [...]*”.⁴³ Its universe of application is that of *maxims*, rather than of particular actions, requiring them to be at-the-same-time (*zugleich*) included as universal law.

The principle of autonomy establishes the condition for the validity of maxims, namely, it requires maxims to express reason’s capacity to confer value and to be the source of value. This condition defines what counts as the goodness of a person’s second level of commitment, i.e., the goodness of her mode of deliberation. If the conditions of consistency in the *ground-end relation* and in the *end-subjective condition relation* are made to be the guidelines of selection among morally possible actions, it is *because* the maxims of such a will are autonomous and give at the same time universal law. Thus, autonomy, as the principle for evaluating an agent’s will, makes explicit the *rationale* at the basis of Kant’s requirements for evaluating actions in general. Autonomy is the condition, itself *unconditioned*, on which those other restrictions rest. As such, it rules over the agent’s second level of commitment, i.e., her commitment to duty as a guideline in the selection of maxims of action. This second level defines the moral quality of a person as a whole (her good will), rather than the atomic value of particular volitions (their rightness).

Kant contrasts autonomy with heteronomy, i.e., the “source of all spurious principles of morality”.⁴⁴ A heteronomous will goes beyond itself and seeks the law that it is to determine it in a property of an already given object. “The will in that case does not give itself the law; instead, the object, by means of its relation to the will, gives the law to it.”⁴⁵ As a criterion for the agent’s adoption of maxims, heteronomy

⁴² GMS, AA 04: 440.18–20: “Das Princip der Autonomie ist also: nicht anders zu wählen als so, daß die Maximen seiner Wahl in demselben Willen zugleich als allgemeines Gesetz mit begriffen seien.”

⁴³ My emphasis.

⁴⁴ GMS, AA 04: 441.2: “[...] Quell aller unächtten Principien der Sittlichkeit”.

⁴⁵ *Ibid.* 7–8: “Der Wille giebt alsdann sich nicht selbst, sondern das Object durch sein Verhältniß zum Willen giebt diesem das Gesetz.”

could give only imperatives that are hypothetical. They require the agent to will something because something else is willed. Hence, principles of morality based on heteronomy remain unjustified. For the will never determines itself immediately, "but only by means of an incentive that the anticipated effect of the action has upon [it] [...]. Here yet another law must be put as the basis in me [...] that [...] in turn needs an imperative that would limit [my] maxim."⁴⁶

The conditional character of such commands forces reason to look for further conditions to ground and justify them. Kant believes that the only way to avoid an infinite regress is to assume the unconditionally good or autonomous will. The property of such a will (being to itself a law) is expressed in the principle guiding the agent's selection of maxims, which guarantees the acceptability of her conceptions of the good for all rational beings. This is why Kant conceives of autonomy as the source of morally acceptable maxims, i.e., as the source of the goodness of the maxims she displays in actions. Actions based on these maxims are *right*. On the other hand, heteronomy, as a criterion of maxim-selection, reveals the agent's liability to take as ultimate reasons for action ends that cannot be fully justified. These are objects of volition assumed on the condition of willing something else. The acceptability of such ends is based on the contingency of other agents having already adopted them, and since this adoption depends on their empirical conditions, there can be no necessity here. This generates a Hobbesian-like situation in which the value of an object would ultimately depend on the agent's capacity for imposing her contingent valuation on other agents. This is why Kant conceives of heteronomy as the source of unacceptable moral principles, i.e., as the source of the evilness of the maxims the agent displays in actions. Actions based on this type of maxims are *wrong*, and the heteronomous will is evil.

The evil that heteronomy makes explicit resides in the fact that the agent has taken as her guideline for maxim-selection a principle whose reason cannot possibly be justified. No *kingdom of ends* can be established on this basis. The reasons offered are mere invitations to continue the war of all against all, i.e., to base on force what lacks in rationality. Just as autonomy is the source of right maxims, since the agent's good will grounds a common world of moral experience (i.e., a world in which good and evil are necessary objects for the faculty of desire), heteronomy is the ultimate source of wrong maxims. It makes known the agent's evil will, her liability to act on unjustifiable reasons, disregarding the question of their acceptability to other agents. Evilness resides in the agent's commitment to a mode of deliberation that annihilates its own rational basis. Such a form of willing is evil because it impedes the constitution of a common world of moral experience and decides questions of right by appealing to force.

⁴⁶ GMS, AA 04: 444.10–15: "[...] sondern nur durch die Triebfeder, welche die vorausgesehene Wirkung der Handlung auf den Willen hat; [...] hier muß noch ein anderes Gesetz in meinem Subject zum Grunde gelegt werden, nach welchem ich dieses Andere nothwendig will, welches Gesetz wiederum eines Imperativs bedarf, der diese Maxime einschränke."

Under this interpretation, autonomy and heteronomy designate competing forms of deliberation and justification, the choice of which decides what a person takes as an ultimate reason for action. This basic commitment further specifies the actions and principles upon which an agent is liable to act, which is what I called the first level of commitment. Right or wrong actions express the kind of moral world the agent makes possible through her maxims. Kant's *kingdom of ends* is the moral world resulting from the systematic integration of the objects of volition of autonomous agents. By contrast, let me call the moral world resulting from the volitions of heteronomous agents a "*jungle of means*", i.e., the dysfunctional integration of objects of volition of agents whose mode of deliberation determines their acceptability by sheer force, rather than by rational considerations.⁴⁷

In other words, the autonomous mode of deliberation gives to all my choices the form: "*I will that you will*". Hence, it commits the agent to act only on those motives and ends that other agents could identify as their own. Heteronomy, on the other hand, gives to all of an agent's choices the form: "*I will what I please*". Here, subjective reasons for action systematically disregard the view of other agents. The mode of deliberation represents diverse interests as mutually *exclusive*. I take the other's will not to be an independent source of value, which I must respect, but a means to my ends – a *thing* for use, not a *person* to respect. The agent builds through her choices a competitive *society*, in which anticipation and distrust of another's intentions lead to (overt or hidden) hostility. Thus, Kant believes that the adoption of self-love as the ultimate reason for action is the basis of all wrongdoing, for it invests the arbitrariness of an agent's will with tyrannical power in the constitution of the moral world. Heteronomy is not simply the source of immorality, but of evil in strict sense, since it indicates the person's willingness to systematically neglect the interests and views of others as reasons for action.

This point brings us back to the beginning of this paper. My goal here was to show the necessity of the concept of evil as expressed in the logic of the *Groundwork*. Skepticism about this notion is based on a misunderstanding about the role the good will has in the structure of justification of maxims: without a rational source of value there would be no justified values. By distinguishing between the criterion to evaluate actions and the criterion to evaluate persons, I drew upon the notions of autonomy and heteronomy to illuminate what Kant means by a "good" or "evil" will. The agent's underlying form of willing and deliberation translates in the kind of moral world she makes possible through her actions, a *kingdom of ends* or a *jungle of means*.

The alleged vacuity of Kant's concept of evil is due to the confusion between the criterion to evaluate persons (maxim-selection) and the criterion to evaluate actions (first and second formulations of the Categorical Imperative). It is fair to say that such a confusion is partly due to Kant's own language, in which

⁴⁷ Karl Jaspers makes a similar point in his "Das radikal Böse bei Kant", in: Jaspers, Karl: *Rechenschaft und Ausblick, Reden und Aufsätze*. München: R. Piper 1951, 109.

autonomy appears both as the supreme principle of morality and as another formulation of the Categorical Imperative. It disappears once we ground Kant's deontology on his theory of value and justification, and distinguish between different levels of commitments within an agent's will. This strategy assures a central place to the notion of a good will, whose necessary correlate is an evil will, not a wrong action.⁴⁸

⁴⁸ I would like to express my gratitude to Lauren Barthold, Tim Boehme, Dimitri Nikulin, Richard Bernstein, and Rüdiger Bittner, who gave me insightful comments and suggestions at various stages of writing this paper.