# Kant on Radical Evil and Reflective Self-Judgment

Kant's “Religion Within the Boundaries of Mere Reason” seems an odd element in Kant's oeuvre. Parts of it seem like scholastic theology or an arbitrary effort to reconcile the Kantian philosophical system with the doctrines of Christianity[[1]](#footnote-1). One of the most troubling notions is that of radical evil. Not only is the motivation for introducing the notion unclear, it is also difficult to grasp the line of argumentation, and furthermore accept its conclusion that there must be an innate propensity for evil in human nature. The vague introduction of the concept of grace exacerbates the puzzlement even further.

Nevertheless, I claim that the notions of radical evil and grace form a significant addition to Kant's conception of moral life lacking in his earlier works in ethics, *Groundwork to the Metaphysic of Morals* and the *Critique of Practical Reason*. I contend that the depiction of human moral life is not exhausted by the consciousness of duty and the endless pursuit toward virtue. Moral life is characterized also by the ongoing introspection of character accompanied by the hope for moral transformation as manifested by the notions of radical evil and grace.

The first section of the paper examines the concept of radical evil and the common problems related to it when taken in a metaphysical or practical sense. The second section presents some motives from *The Critique of Judgment* related to its central concept*,* that of reflective judgment. in the last section I relate reflective judgment to the problematic of the *Religion.* Through a reading of book II of the *Religion* I show that the notion of radical evil is introduced not as a metaphysical claim about human nature, nor does it have a direct practical bearing on moral obligation, but rather as a mode of self-judgment about one's own moral character. Such a judgment can be described using the notion of reflective judgment as a judgment on the availability of moral possibilities. Using these observations it is possible to give a more satisfying account of the notions of radical evil and grace presented in Kant's *Religion*.

# Radical Evil

The *Religion* begins with the ancient question whether human kind is good or evil by nature. Kant does not disclose his motivation in tackling this question and it seems peculiar in view of the standard understanding of Kantian ethics. In the *Groundwork* Kant gives a full account to the meaning of the moral good: The moral good is a good will, a will motivated by duty, i.e. respect for the moral law. The essence of morality is encapsulated in the categorical imperative which gives the general form of morally good maxims for willing actions. But if the categorical imperative guiding specific actions is the essence of morality, what is the meaning of the question about the moral character of a person as a whole, and furthermore the moral nature of human kind in general?

There seems to be a discrepancy between the notion of a moral nature and the basic premises of Kantian ethics. The moral law is inextricably bound to the idea of freedom. Only acts of freedom can be morally accountable. Therefore if there is such a thing as a moral nature, it must be related to an act of freedom, something that can be imputed to a person. But if it is an act of freedom, how can it belong to the essence of the human species? And if it is an inborn state of humanity, in what sense a person can be imputed with it? Kant was well aware of this problem:

But lest anyone be immediately scandalized by the expression *nature*, which would stand in direct contradiction to the predicates *morally good* or *morally evil* if taken to mean (as it usually does) the opposite of the ground of actions from *freedom*, let it be noted that by “the nature of a human being” we only understand here the **subjective ground** – wherever it may lie – of the exercise of the human being's freedom in general[[2]](#footnote-2) (emphasis mine)

The key notion here is that of a “subjective ground”. For Kant every action is an expression of a maxim, a rational ground for action. The maxim of a moral action conforms to the form of the categorical imperative. A subjective ground for the exercise of freedom in general, is therefore a supreme maxim guiding the whole use of the power of choice. Such a supreme maxim can thus be equated with the concept of a moral character, a subjective trait grounding the totality of conduct. If a moral character is to be imputable, the choice of this supreme maxim is itself an act of freedom[[3]](#footnote-3). However, although it must be considered a free act, it cannot be described as governed itself by a maxim and in this sense, the presupposed subjective ground must remain inscrutable. For if the choice of the supreme maxim was governed by a maxim, the maxim would not be supreme but dependent on another maxim:

Since the adoption is free, its ground must not be sought in any incentive of nature, but always again in a maxim; and since any such maxim must have its ground as well, yet apart from a maxim no *determining ground* of the free power of choice ought to, or can, be adduced, we are endlessly referred back in the series of subjective grounds without ever being able to come to the first ground.[[4]](#footnote-4)

The adoption of the supreme maxim is then a free action that is itself not governed by a maxim, and it is also a very peculiar kind of action, since it does not occur at any point in time. To explain this peculiarity, Kant uses his common distinction between the phenomenal and the intelligible. The incorporation of the supreme maxim is “an intelligible deed, cognizable through reason alone apart from any temporal condition”[[5]](#footnote-5)

The choice of the supreme maxim is an inscrutable intelligible act, but we can still give a description of its content as equivalent to the notion of moral character. A good moral character means the incorporation of the moral law as a determining incentive into the exercise of the power of choice. The more difficult task is to explain the nature of an evil character. The obvious thought would be to seek the source of the evil character in the natural inclinations, as though they have the power to seduce and overcome the incentive of the moral law and subordinate the freedom of the will. Yet Kant strongly rejects such a view. In the inclinations themselves there can be nothing evil, as they are a product of nature, and all moral value is the product of freedom. Therefore the source of an evil character lies in the free choice to adopt of an evil supreme maxim, a maxim countering the obedience to the moral law. Kant identifies two possible prime incentives for the will. The first is self love, the aspiration for happiness as the maximum satisfaction of inclinations, and the second is the moral law. Both incentives belong to human nature, the first as a natural being, and the second in virtue of having practical reason. Therefore the distinction between the good character and the evil character presents itself as the question about the priority of one of the incentives over the other[[6]](#footnote-6).

This description of the difference between the two moral characters, also explains Kant's ethical rigorism[[7]](#footnote-7) regarding moral character. The two characters are qualitatively distinct, as they entail opposite orders for the subordination of incentives. Therefore each person is either good or evil, there can be no middle ground, no combination of the two, nor a neutral moral character, deprived of moral value whatsoever. For this reason as well, the transition from good to evil cannot be thought of as a gradual process, but only as a complete transformation, a change of heart so to speak. When understood as a statement of a fact, this rigorism seems problematic, as the common intuition is that people can be somewhere in between, not totally evil, but also not quite righteous.

Following this definition of moral evil, Kant ascribes it to a propensity innate in human nature. A propensity is “the subjective ground of the possibility of an inclination”[[8]](#footnote-8), a latent potential assumed to exist in the subject explaining the actual forming of habits and inclinations. A propensity for evil is thus the propensity that explains the tendency to deviate from the moral law. The crucial and controversial step in Kant's argument, is the claim that this propensity to evil applies to the human species as a whole, that evil is inherent in the human nature. The support for this claim is most bewildering:

not that this quality may be inferred from the concept of his species (from the concept of a human being in general, for then the quality would be necessary), but rather that, according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, we may presuppose evil as subjectively necessary in every human being, even the best.[[9]](#footnote-9)

On the one hand, it is not an analytical proposition, meaning that the property of evil does not belong to the concept of a human being, so that is not objectively necessary that all people are evil. On the other hand, it is a necessary presupposition, although only subjectively. But what is the ground for this necessity? Kant notoriously states that "we can spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human *deeds* parades before us"[[10]](#footnote-10). This alludes at least to the possibility of such a proof, which would be required to establish even the merely subjective necessity of evil in human nature. Necessary knowledge cannot have its source in experience, even if it is evident that evil existed in all cultures throughout history.

This alleged laziness to supply a proof is puzzling, and commentators found different ways to fill the gap in Kant's argument. Allen Wood in his early writings dismissed the whole pretension for a formal proof and took the theory as a mere empirical generalization[[11]](#footnote-11). Later he revised his position interpreting radical evil not as a property of the individual human being, but of the human social existence, as part the notion of the unsocial sociability of human kind[[12]](#footnote-12). Henry Allison, on the other hand, presented a reconstruction for Kant's neglected formal proof[[13]](#footnote-13). The proof is based on human finitude and unholy will. The finite and sensible nature of human beings implies that they cannot be motivated solely by the moral law, but also by natural inclination. Therefore the moral law always appears as a duty that can be disobeyed. This possibility of transgression, together with the tendency to be actually tempted by the principle of self-love, is for Allison the ground of evil in human nature. In addition to other problems in this argument[[14]](#footnote-14) Allison's reconstructed proof seems too formal on the one hand and too trivial on the other hand. Kant's clearly states that the propensity to evil cannot be derived from the concept of a human being, and it implies much more than just the unholiness of the will, but rather an active force resisting the moral law. However, Wood's reduction of radical evil to the maladies of society, while based on many allusions in Kant's writings as well as the requirement for a social instrument to overcome evil (the ethical community presented in book III), takes too lightly the individual anxiety over the self moral worth expressed in the notion of radical evil. I suggest that both Allison and Wood give a partial account of radical evil because they take it as an assertion about human nature, whether it is a metaphysical a-priori proposition in the case of Allison or an empirical proposition in the case of Wood. Both of them disregard the relevance of book II with its emphasis on exemplarity and judgment. Instead of interpreting the doctrine of radical evil as a content of a proposition, in section 3 I will try present it as a mode of self-judgment.

In addition to the unclear argument grounding the innateness and universality of evil, the conclusion itself is problematic. For this propensity to have a moral value, it must be an imputable act of freedom, as though brought by the human being upon himself. Here we face not only the problem of being responsible to the choice of a supreme maxim not taken at any point in time, but also that there is a universal necessity in choosing the evil maxim. How can one be responsible for something she has never chosen and which belongs to her by the very fact of being part of the human species?

The problem is deepened by the obligation for moral transformation. The good and the evil moral characters express opposite maxims, and therefore they differ qualitatively, so that no gradual transition from the original evil to the required good is possible. A total transformation is required, an adoption of a new maxim establishing the superiority of the moral law. This adoption, just like incorporation of the original evil maxim, cannot be explained by any maxim, as it is itself the action fixing the constitution of maxims[[15]](#footnote-15). This perplexity involved in the transformation from evil to good is unavoidable, and it is part of the incomprehensibility of the idea of freedom itself. But as in other elements of Kant’s philosophy of religion, the guiding principle is that "ought implies can". Since there is a duty to overcome the initial evil propensity, it must also be possible, because a duty to perform the impossible is a contradiction[[16]](#footnote-16).

According to some Christian conceptions, one cannot be rescued from the grip of evil without some kind of divine assistance in the form of grace. For obvious reasons Kant's attitude toward such views is highly ambivalent and cautious, to say the least[[17]](#footnote-17). The adoption of an evil character is an act of freedom, brought upon oneself. No external force can take away the burden of this initial guilt and justify atonement. Since there is a duty to overcome evil, no external assistance can provide substitution for the autonomous legislation of the will in establishing a good moral disposition. Furthermore, if one has done everything within her powers to act virtuously, there should not be any need for an additional graceful act from God's side, and if one has not done everything she could as duty demands, she is surely not worthy of divine grace[[18]](#footnote-18). Kant is aware of these problems and dismisses the ideas of moral transformation by divine aid or vicarious atonement as superstitious. Not only are these ideas theoretically incomprehensible since there can be no notion of an external influence on the free will, they have no practical use either. On the contrary, such a use would lessen the personal responsibility for the demands of duty. Nevertheless, Kant does not dismiss the notion of grace entirely, and as I will show, it is not just lip service for the common Christian discourse but can play a role in his conception of moral life.

As a supernatural action divine grace is completely incomprehensible so the most we make of it is the notion of receptivity for grace as being worthy of it:

receptivity is all that we, on our part, can attribute to ourselves, whereas a superior's decision to grant a good for which the subordinate has not more than (moral) receptivity is called grace.[[19]](#footnote-19)

But if being receptive for grace is no more than being virtuous, why is there a need to introduce such a concept at all? Kant strangely remarks that although we can neither understand nor assign a practical use for the idea of grace, we should "keep ourselves at a respectful distance from it, as from a sacred thing"[[20]](#footnote-20). The belief in this idea is therefore only a “*reflective* faith”[[21]](#footnote-21) without any claims for knowledge or a practical use. What is the import of a ‘reflective faith’ and what is involved in it? For this it is necessary to expound Kant's notion of reflective judgment.

# Reflective judgment

The problems described above concerning the doctrine of radical evil and moral character can be seen as analogous to the difficulty of accounting for the power of judgment in general. The power of judgment is the faculty of "subsuming under rules, i.e., of determining whether something stands under given rule or not"[[22]](#footnote-22). The difficulty with the power of judgment is that the judgment itself, i.e. the application of rules, cannot be explained itself by a rule, on pain of inﬁnite regress[[23]](#footnote-23). A similar difficulty applies to Kant's account of moral character. Establishing a moral character requires a choice which precedes every rule, because this is the choice of the supreme rule for all subsequent choices. In both cases we need some way to articulate the possibility of establishing a rule when no prior rule is available. This possibility is expressed in the notion of reflective judgment developed in the Critique of Judgment explored in this section. In the last section I will show that this structural similarity between the two problems is more than an analogy, so that my reading of the *Religion* will restate the problem of moral character as a species of the problem of judgment: as a part of moral life, the problem of moral character is not a problem of choosing, but one of judgment, appearing as a constant necessity for reflective self-judgment.

The Critique of Judgment, as its name suggests, deals with the faculty of judgment in its own right, and not as dependent on concepts and rules supplied by the other faculties of cognition, understanding and reason. Thus the problem of judgment can be formulated as the question about the conditions of possibility of judgment as such, which is to be answered by finding its own a-priori principle. The key to finding the a-priori principle of the faculty of judgment is the distinction between determinant and reflective judgments. Determinant judgments are those in which the universal is given and the particular is subsumed under it. With reflective judgment, on the other hand, the particular is given and the universal subsuming it is to be sought[[24]](#footnote-24). For this reason, if the power of judgment has a principle for its use, it would be exhibited in reflective judgments.

 In the introductions to the third critique, Kant seeks the principle of judgment by analyzing the role of reflective judgments in the scientific investigation of nature. Such a principle is necessary, because the categories of understanding described in the first critique which make objective experience possible, are not sufficient for explaining nature as a unified system as demanded by science. For this purpose more assumptions are needed, such as that the manifold of laws could be unified under more general laws, that similarity can be found between different domains of nature and that in general the phenomena of nature can be categorized systematically[[25]](#footnote-25). These assumptions are equivalent to the idea that nature was designed to befit our cognitive faculties, to be understandable by us. Therefore this idea can be designated as the principle of the purposiveness of nature. Since it cannot be inferred from experience, it is an a-priori principle of the power of judgment. This principle is required to establish the possibility of the human thought in the broad sense, not just by providing the conditions for representing an objective environment, but by expressing the giveness of a space of possible progress toward systematic knowledge, thus realizing the human interest in the intellectual project.

The same function of guiding the investigation of nature is assigned in the Critique of Pure Reason to the regulative principles of reason[[26]](#footnote-26). Reason has no direct relation to experience, it does not create concepts of objects, and hence it is not constitutive of objects. But reason does have a relation to concepts already provided by the understanding[[27]](#footnote-27). It does so by providing rules or guidance for the ordering of concepts and for seeking new concepts in relation to others. The general structure prescribed by reason for the ordering of concepts is that of a unified system, so that they can be thought of as interrelated under one principle. The systematicity of nature is given as a task for the understanding by reason, under the conception of the maximum. This is a theoretical maxim to always seek for the more homogeneity, diversity and affinity in the empirical concepts of nature.

However in the introduction of the Critique of Judgment, seeking systematicity is assigned to reflective judgment. What is the meaning of the shift of this function to the power of judgment? I suggest that the notion of reflective judgment adds an affective aspect to the task of systematizing knowledge, as a concrete experience of harmony in addition to the general regulative maxim. There is a kind of experience which exemplifies the possibility of systematic knowledge. Already in the introduction Kant relates the cognitive interest in systematicity with the feeling of pleasure. Systematicity is not guaranteed by the categories of the understanding, hence it is not objectively necessary, but it is still an interest of reason. Since the attainment of goals is accompanied by a feeling of pleasure, finding systematicity in nature causes pleasure[[28]](#footnote-28). With reflective judgments purposefulness is manifested by a feeling and not only prescribed by a regulative command for endless progress towards a postulated idea of a maximum. In this sense, reflective judgment is the condition of judging the adoption of a maxim for maximizing systematicity to be possible. Thus it articulates the manner in which ideas of reason are taken to be useful in concrete way through a certain kind of experience.

This relation between reflective judgment, feeling and possibility is exhibited most clearly in aesthetic judgments. In addition to the regulative use of the principle of purposiveness for discovering the systematic laws of nature, or in judging the purposive structure of organisms, there are mere reflective judgments that do not contribute anything to objective knowledge. These are aesthetic judgments, judgments that relate solely to the subjective feeling of pleasure or displeasure without yielding any objective knowledge. The capacity for reflection is a capacity of comparison between given representations or between representation and the faculties of cognition in general[[29]](#footnote-29). In aesthetic judgments the comparison between a representation and the faculties of cognition is found to be purposive, by the way that the representation elicits a harmony between the faculties, a harmony which is required for cognition in general[[30]](#footnote-30). This harmony is not determined by a concept prescribed by the understanding, but comes forth freely in the process of aesthetic contemplation. In the second moment of the analytic of the beautiful, this harmony is described as the free play between the faculty of imagination and the faculty of understanding. Both faculties are required for every cognition, however in determinant judgments the imagination is subordinated to a concept supplied by the understanding. Aesthetic judgments demonstrate that they can cooperate freely without coercion. In this sense, in an aesthetic judgment the representation is judged as representing the possibility of judgment in general, without being subsumed under a specific concept[[31]](#footnote-31). The representation is deemed merely as understandable, as providing material for further thought. Thus what is revealed in aesthetic judgment is the general capacity to perform judgments, to cognize something at all.

This is made clearer in the third moment of the analytic where judgment of taste is described as purposiveness without purpose. The judgment of taste is disinterested, so it is not related to a specific purpose. However it shows a form of purposiveness without purpose in the harmony the faculties of cognition. In this respect the beautiful object appears as designed to fit our subjectivity in general without forcing on it a specific purpose. In other words, since in aesthetic judgments the capability of cognition is manifested, they can be considered as an appearance of possibilities without determination of their mode of actualization. There is therefore a basic human capability which locates in the particular concrete object the possibility of things to be cognizable generally, i.e. the possibility of sense and significance.

This capability is not limited to the theoretical domain, it has deep implications also for the practical use of reason. More specifically it is supposed to be related to the unity of both uses of reason. It is in fact the declared overall goal of the Critique of Judgment to describe the possible transition "from the domain of the concept of nature to that of the concept of freedom"[[32]](#footnote-32), thus constructing a bridge across the "incalculable gulf"[[33]](#footnote-33) between these two domains. The need for a bridge arises because moral action has to influence the natural world although its conditions of possibility are completely unrelated to the demands of morality:

Understanding and reason thus have two different legislations on one and the same territory of experience[[34]](#footnote-34). . . The concept of freedom **should** make the end that is imposed by its laws real in the sensible world; and nature must be consequently also be able to be conceived in such a way that the lawfulness of its form is at least in agreement with the possibility of the ends that are to be realized in it in accordance with the laws of freedom[[35]](#footnote-35).

An account of the exact way in which this transition is beyond the scope of this paper, but for my purposes here it suffices to say that by introducing the notion of reflective judgments, Kant conceives of a way to view nature's aptness for human moral aspirations. The appearance of such a space for transition between the natural and the moral facilitated by the power of reflective judgment also underlies my reading of the *Religion*[[36]](#footnote-36), but this time the question revolves around the suitability of human inner nature, rather than nature in general.

# Moral character and reflective judgment

Let us recall the common conundrums associated with the notion of radical evil:

1. How can we be accountable for our moral character if it is not an action taken at any point in time and if it is a propensity innate in human nature?

2. How can we choose a moral character if this is the act that establishes the available reasons for free choice in the first place?

3. What sense can be given to the idea of grace without impairing the demand for autonomy and moral responsibility?

In this final section I will show how the notion of reflective judgment allows the resolution of these problems. The first part in the argument will show that the notions of radical evil and grace are to be thought of as modes of judgment about our own moral character. The second part shows that this kind of judgment fits the description of reflective judgments articulated in the third critique.

I have already noted the structural analogy between the problem of moral character and the general problem of judgment, the first in the realm of cognition and the second the realm of conduct. The problem of judgment is that no rule can be supplied for it, because judgment itself consists in the application of rules. Likewise, moral character is the choice of a supreme maxim which is itself not governed by any maxim. In what follows I will show that this is not just a structural similarity, as the problems of 'choosing' a moral character and the possibility of judging it are tightly intertwined in the first two books of the *Religion*. This means that the crux of the question of moral character lies in the possibility of judging it, moreover in the obligation of judging it, something that is expressed in the nagging anxiety over our own moral worth.

But how can we evaluate our own moral worth? Such an evaluation assumes that we can find a unifying principle giving coherence to the multitude of actions and patterns of behavior. The supreme maxim underlying all other particular maxims is purely intelligible, inscrutable in experience which reveals only particular deeds. Even the maxim motivating each separate action is not transparent for the agent – one can only speculate but can never be assured whether an action is enacted out of duty or merely in accordance with it but from another motive[[37]](#footnote-37). It seems that passing judgment on the moral character requires something like a reflective judgment, some way of representing the rule and potential implicit in particular deeds.

Another discussion about moral character can be found in the Critique of Practical Reason concerning the concept of the highest good. Practical Reason can conceive the idea of the complete correlation between happiness and virtue as the highest good. The idea is not only conceivable but is the necessary object of the moral will. But complete virtue, a will in complete conformity with the moral law, is not attainable for natural beings. Since there is a duty to promote the highest good, it must also be possible. Therefore the immortality of the soul is postulated to enable the endless progress toward virtue[[38]](#footnote-38).

In the *Religion* however, the question regarding moral character is quite different and thus the answer is of a different kind. While in the second critique, what is sought is the condition for making endless progress toward virtue possible, the *Religion* deals with the problem of the possibility of even beginning the quest toward virtue, in light of the innateness of evil and the requirement for a total change of heart. More accurately, the question of the second critique is absorbed into the one of the Religion: The change of heart to the good cannot be thought of to be genuine if it does not encapsulate the constant progress towards virtue. For this reason the solution of the problem is different. While in the second critique a metaphysical solution is put forward, by postulating the eternal soul as the substratum for moral progress, in the Religion the problem is presented as a problem of judgment: in what sense can one think himself worthy before a divine judge. This is at large the purpose of book II of the *Religion*.

Most of the commentators of the *Religion* ignore book II which seems redundant in the total line of argumentation. While book I states the problem of evil in human nature and book III describes the required cure for evil in the formation of a universal church, the purpose of book II is unclear. It is titled “concerning the battle of the good against evil principle for the dominion over the human nature”, yet it does not provide any description of a 'battle', or a mechanism that explains how one principle presides over the other. Both the adoption of the initial evil principle and a transformation towards the good, as we have seen above, are totally inscrutable. Thus book II cannot indeed give us a rule for adopting a supreme maxim and the victory of the good remains a mystery. Yet because the moral transformation is obligatory, it must also be possible. The mode of this possibility is presented in the ideal of the human being well pleasing to God, the *son of God[[39]](#footnote-39)*. As an ideal, its origin is not in experience but in reason alone, yet it can still be presented concretely for serving as an example for emulation:

an experience **must** be possible in which the **example** of such a human being is given. For, according to the law, each and every human being should furnish in his own self an example of this idea. And the required prototype always resides only in reason, since outer experience yields no example adequate to the idea.[[40]](#footnote-40) (emphasis mine)

Some things are important to note in this passage which relate it to the issue of judgment. Why an example is necessary for progress towards virtue? Isn't the moral law sufficient for guiding the formation of character? How can an idea which cannot be exhibited in experience serve as an example? Kant adds the qualification that as an idea of reason this example is something that each has to seek within himself. Thus embracing the example involves recognizing oneself as capable of judgment according to the ideal. This mode of exemplarity is inherent also in aesthetic judgments. As aesthetic judgments pertain to the purposiveness of the cognitive faculties and thus to the possibility of judgment in general, they make a claim for universality and necessity. But since aesthetics judgments are not bound by a concept, this necessity is only exemplary - providing an example for a rule that cannot be formulated[[41]](#footnote-41). Thus the norm involved in aesthetics is only an ideal norm, and if there is an ideal of beauty as an archetype of taste it must be:

a mere idea, which everyone must produce in himself, and in accordance with which he must judge everything that is an object of taste, or that is an example of judging through taste[[42]](#footnote-42).

Although there is an obvious difference between the ideal of beauty and the moral ideal of the son of God, as the latter is based on the definite concept of virtue, the similarity between this passage about the ideal of beauty and the former is not accidental. Since the faculty of judgment as revealed in aesthetic judgments depends on an ideal norm and since examples are needed for the cultivation of judgment, the use of the ideas of reason as examples possesses the same source of authority from “one's own”, which is inherent in the nature of judgment as such[[43]](#footnote-43).

In order to show the possible reality of the idea and thus its ability to serve as an example, Kant states that three obstacles have to be settled[[44]](#footnote-44). By solving these difficulties, not only the theoretical possibility of the ideal is shown but also the possibility of considering ourselves in conformity with this ideal. This is so because the solution involves engaging in a certain mode of judgment about the ideal, which I will claim is akin to reflective judgments. The first difficulty is similar to that of the antinomy of practical reason in the second critique, the impossibility of attaining a holy will. The solution to this difficulty is similar: from an intellectual point of view what matters is the disposition driving the progress and not the measure of virtue at a specific point in time. However here, there is no need to postulate an immortal soul, just the possibility of a divine judgment by an intuitive intellect:

we can think of the infinite progression of the good toward conformity to the law as being judged by him who scrutinizes the heart (through his pure intellectual intuition) to be a perfected whole even with respect to the deed (the life conduct)[[45]](#footnote-45)

The second difficulty concerns the constancy of the good disposition. Even if we acquired a good character what can guarantee its immutability? Here again a judgment of the totality from the given is required, conjecturing the constancy of the disposition according to past behavior. The last and greatest difficulty is that of justifying the judgment of a converted person as morally good in view of her necessary past as evil. For whatever good she has done, it was her duty and there can be no surplus of good to compensate for past mischiefs. The solution is that judging from an intelligible point of view, the identity of the person has changed after the change of the moral disposition. Moral transformation is to be judged as the rebirth of the person. All the three difficulties pertain to the judgment of the moral character: In what sense can the disposition be judged as good, unchangeable, and redeemed despite its impure past. Since the moral ideal to be used as an example has to be found in one's own reason, these are judgments of one's own possibility to embrace the ideal as a norm. Therefore the judgment that is expressed from the perspective of the possibility of the ideal as a divine judgment, as a mode of self-judgment it can be thought of as a case of reflective judgment.

I'll say a little bit more about reflective judgments in general to clarify their role with regard to the notion of moral character in the *Religion*. The casting of the problem of moral character as a problem of judgment instead of as a postulate of practical reason is analogous to a shift of emphasis between the Critique of Pure Reason and the Critique of Judgment. As noted above, the introduction of the third critique presents the a-priori principle of the power of judgment, the purposiveness of nature, by locating its use in the scientific inquiry of nature which presupposes the systematic unity among the empirical laws of nature[[46]](#footnote-46). The same concern is presented in the first critique, in the appendix to the transcendental dialectic[[47]](#footnote-47), regarding the regulative use of the ideas of pure reason regulating the discovery of systematic unity in nature. Yet there is a difference between the regulative maxims of reason commanding endless progress toward the maximum and the concreteness of the reflective judgment which is related to particular objects. This concreteness is shown by the relation that Kant establishes between the power of judgment and the feeling of pleasure and displeasure, the most subjective, non conceptual and immediate of all mental representations. Therefore what the Critique of Judgment examines is concrete examples of the various uses of the reflective power of judgment, in the scientific inquiry of nature, in biology and in the field of aesthetics. This connection between the feeling of pleasure and the reflective judgment is most salient in aesthetic judgments of taste which are purely subjective, yet disclose the possibility of cognition itself. My claim is that something similar to this, a reflective judgment accompanied by a certain kind of feelings, those of hope and guilt, is presupposed in the discussion of radical evil and moral character in the *Religion*.

I noted above the difference betweenthe Critique of Practical Reason and the *Religion* with regard to the problem of virtue. The first solves it by postulating the immorality of the soul, the second by introducing a certain mode of judgment. This difference emphasizes that the problem of moral character is not only an epistemological or a metaphysical problem regarding its conditions of determinability or possibility, but also an existential problem: what is the moral meaning of one's life? This mode of reflective self-judgment is the articulation of what it is like to live a life accompanied by the consciousness of the moral law considering inner human nature and outer nature as a whole.

There are several other places in which Kant describes the affective aspect of moral life. One kind of feeling related to the consciousness of the moral law, that of respect, is articulated in the *Groundwork* and more extensively in the second critique[[48]](#footnote-48). This feeling is necessarily produced by the acknowledgment of some product of reason that can subordinate and humiliate the natural inclinations, and in this sense the moral law is sublime. The awareness of a moral law that commands absolutely and overrides all other incentives produces a feeling of veneration. Similarly, the aesthetic judgment of the sublime, described in the Critique of Judgment, is about the felt elevation produced by the failure of cognition to grasp the sensible magnitude or might of nature, which leads to the consciousness of a faculty that is independent of and surpasses every natural measure or purpose, i.e. the faculty of reason[[49]](#footnote-49).

But besides the sublimity of the moral law, the *Religion* adds more affective facets to the description of the moral life. In addition to the anxiety over our own moral worth in comparison with the holiness of morality, there is an element of hope contained in the demand for a conversion of heart. There is a need to exemplify the possibility of moral progress, a feeling that we are heading in the right direction although no objective judgment of moral character can be made. Such an appearance of the possibility of progress can be related to the judgments of beauty as presented in section 2. Indeed, there are several places in which Kant associates the aesthetic and the moral in such a way. Although the judgment of taste is disinterested, meaning that there is no desire in the existence of the beautiful object itself, in §42 "On the intellectual interest in the beautiful", a special kind of interest is related to the existence of beauty in nature in general. Reason has an interest in finding in nature traces or signs for the objective reality of its ideas. The free and disinterested harmony found in judgments about the beauty of nature exhibits such a sign for the suitability of nature for our purposes[[50]](#footnote-50). Furthermore, this interest is not only related to outer nature, it also hints something about the inner nature, about a predisposition toward the good:

He who takes such an interest in the beautiful in nature can do so only insofar as he has already firmly established his interest in the morally good. We thus have cause at least to suspect a predisposition to a good moral disposition in one who is immediately interested in the beauty of nature[[51]](#footnote-51)

Taking interest in the existence of beauty in nature as such, that is in its eliciting a feeling of purposiveness, points to a similar purposiveness and harmony in character, a kind of harmony which is associated with the resolute adoption of the good maxim. In other words, the possibility of judging the beauty in nature is related to the possibility of having a 'beautiful soul'. Taken in this sense of felt harmony, a good moral character has a kind of aesthetic quality. This is indeed the essence of Kant's response in the *Religion* to Schiller's criticism about the harsh attitude accompanying the moral law[[52]](#footnote-52). The “aesthetic constitution” of virtue is “courageous and hence joyous” and not fearful or dejected, accompanied by love for the good, and not by a secret resentment. A good moral disposition requires a resolve to improve morally combined with the hope that improvement is possible, and is signified by a certain feeling:

This resolve, encouraged by good progress, must effect a joyous frame of mind, without which one is never certainof having gained a love for the good, i.e. of having incorporated the good into one's maxim[[53]](#footnote-53).

We see here that only through this kind of joyous feeling, which I argued is to be identified as a reflective judgment, a self-judgment of moral character is possible. Having this relation between beauty and character in mind, the possibility of a good moral character is itself judged as a marvel of nature, exhibiting a harmonious and simple existence in sensual creatures like human beings. The reflective power of judgment in its application to inner and outer nature, thus points to something beyond the rigorous opposition of freedom and nature, some common ground for both that is manifested in their harmony. This is expressed in one of the more adventurous passages of the third critique:

it [the power of judgment] sees itself, both on account of this inner possibility in the subject as well as on account of the outer possibility of a nature that corresponds to it, as related to something in the subject itself and outside of it, which is neither nature nor freedom, but which is connected with the ground of the latter, namely the supersensible, in which the theoretical faculty is combined with the practical, in a mutual and unknown way, to form a unity.[[54]](#footnote-54)

Here it seems that Kant almost transgressed the boundaries of reason by making reflective judgment lead to the unconditioned ground of everything, the holy grail of traditional metaphysics whose possibility Kant rejected in the first critique[[55]](#footnote-55). Yet having its source in reflective judgment, this unity is completely unknowable and therefore also of no immediate practical use because we cannot draw moral guidance from nature. Precisely because of this, finding signs or clues for this unity in nature can only be regarded as an act of grace, a favor of nature[[56]](#footnote-56), and hence related to religion[[57]](#footnote-57) and not to metaphysics.

With this connection between beauty, nature's grace, and morality, we are in a better position to understand Kant's perplexing notion of grace as a reflective belief[[58]](#footnote-58). Moral consciousness requires not only strict obedience to the law, but also a dimension of receptivity and gratitude that such a realm of rational order exists at all, accompanied by the hope of reaching this inner harmony and having a part in it. This position is implied in the *Conflict of the Faculties*:

grace is none other than the nature of man in so far as he is determined to actions by a principle which is intrinsic to his own being, but supersensible (the thought of his duty). Since we want to explain this principle, although we know no further ground for it, we represent it as a stimulus to good produced in us by God, the predisposition to which we did not establish in ourselves, and so, as grace.... It has to be made clear from them that we ourselves must work at developing that moral predisposition, although this predisposition does point to a divine source that reason can never reach (in its theoretical search for causes), so that our possession of it is not meritorious, but rather the work of grace [[59]](#footnote-59)

Like a reflective judgment, reflective belief in divine grace pertains to the appearance of a possibility, the possibility of moral transformation which is only the beginning of an infinite progress toward virtue. In the scriptures, this possibility is depicted symbolically as the appearance of the son of God. The validity and knowledge of the moral law is of course not dependent on his appearance on earth, yet it symbolizes the favorable opening of a space for the fulfillment of morality and thus represents the moral hope to overcome evil:

another moral dominion [than that of evil] has been revealed to them [human beings] as freedom, and it it they can find protection for their morality[[60]](#footnote-60)

How are the conundrums of radical evil and moral character related to this notion of grace? The representation of the revelation of morality and the possibility of conversion as an act of grace, presupposes a background of radical evil as the unfavorable reality of human nature. We can only feel grateful for something which is uncertain or undeserved. Hope for moral transformation and guilt for not having yet achieved it are therefore intertwined on this account of the reflective self-judgment of moral possibilities. This is evident by the way Kant summarizes the significance of the theory of radical evil:

The thesis of innate evil is of no use in moral *dogmatics,* for the precepts of the latter would include the very same duties and retain the same force, whether there is in us an innate propensity to transgression or not. In moral *discipline* [moralischen Ascetik], however, the thesis means more, yet not more than this: We cannot start out in the ethical training of our connatural moral predisposition to the good with an innocence which is natural to us but must rather begin from the presupposition of a depravity or our power of choice … and, since the propensity to this is inextricable, with unremitting counteraction against it[[61]](#footnote-61)

The notions of radical evil and grace have no significance regarding the content of moral duty which is exhausted by the notion of a moral law. They do, however, contribute to an adequate account of the presence of morality in actual human life, a presence which is articulated as a constant self-judgment and disciplining. I will demonstrate this by examining Kant's brief example of moral education in the Critique of Practical Reason[[62]](#footnote-62). In this moral scene of instruction a ten-year-old boy is told a story about a man who is being pressured to support the indictment of an innocent person. First he is offered gifts but rejects them. Then his friends abandon him and he loses his social status. Finally threats to his freedom and even his life are made. In all this, he holds up to his firm resolution not to lie. The story is supposed to have the following effect on the sensitive boy:

then my young listener will be raised step by step from mere approval to admiration, from that to amazement, and finally to the greatest veneration and a lively wish that he himself could be such a man (though certainly not in these circumstances)[[63]](#footnote-63)

In this story a new mode of value is revealed to the listener, one beyond any incentive of happiness. Yet if my account of the *Religion* is plausible, this kind of response in front of a moral ideal is not limited to the initiation into morality, but prevails throughout life as a mode of self-judgment. For a person, already aware of the moral law, the response of facing such an ideal example is not only that of veneration in front of the possibility of virtue, but also a feeling of shame. Being aware of the strong temptations, one is never sure how she would have acted in such a situation, thus calling her own moral constitution into question yet hoping to attain the resolve to do the right thing.

Just as the fall to evil is not an event in time but still imputable to us, moral conversion is likewise not an action taken at some point in time, but the appearance of a possibility for transformation in our horizon, always still ahead of us, and always on the background of the nagging question "Am I still evil?". The appearance of such a possibility, as a chance for a rebirth, can therefore be thought of as an act of grace.

**References**

**Works by Immanuel Kant and abbreviations**

All references are given by the *Akademie* edition as following: volume:page

References to the Critique of Pure Reason are indicated as A or B.

A/B Critique of Pure Reason [1781/1787] (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Allen W. Wood and Paul Guyer, Cambridge: Cambridge University Press, 1998.

Gr Groundwork of the Metaphysics of Morals [1785] in Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Mary J. Gregor, Cambridge: Cambridge University Press, 1999.

KpV Critique of Practical Reason [1788] in Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Mary J. Gregor, Cambridge: Cambridge University Press, 1999.

KU Critique of the Power of Judgment [1790] (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Paul Guyer. Cambridge: Cambridge University Press, 2001.

RGV Religion within the boundaries of mere reason alone [1793] in Religion and Rational Theology (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Allen W. Wood and George Di Giovanni, Cambridge: Cambridge University Press, 2001.

SF The Conflict of the faculties [1798] in Religion and Rational Theology (The Cambridge Edition of the Works of Immanuel Kant in Translation). Translated and edited by Allen W. Wood and George Di Giovanni, Cambridge: Cambridge University Press, 2001.

**Works by other authors**

Allison, Henry E.. Kant's Theory of Freedom. New York: Cambridge University Press, 1990.

Allison, Henry E.. Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment . New York: Cambridge University Press, 2001.

Allison, Henry E.. *On The Very Idea of Propensity to Evil.* The Journal of Value Inquiry. 36: 337–348, 2002

Davidovich, Adina, *How to Read Religion within the Limits of Reason Alone*. Kant-Studien, vol. 85, no. 1, pp. 1-14, 1994

Davidovich, Adina. Religion As a Province of Meaning: The Kantian Foundations of Modern Theology (Harvard Theological Studies). Philadelphia: Augsburg Fortress Publishers, 1993.

Silber, John R. *The Ethical Significance of Kant's* Religion. in Religion within the Limits of Reason Alone. Theodore M. Green and Hoyt H. Hudson (ed.). with a new essay by John R. Silber. New York: Harper and Row, lxxix-cxxxiv. 1960

Wolterstoff, Nicholas P,  *Conundrums in Kant's Rational Religion.* In Rossi, Philip J. (Ed). Kant's Philosophy of Religion Reconsidered. Bloomington: Indiana University Press, 1991.

Wood, Allen W.. Kant's Moral Religion. New York: Cornell University Press, 1970.

Wood, Allen W.. Kant's Ethical Thought. New York: Cambridge University Press, 1999.

1. For example Göthe: “Kant required a long lifetime to purify his philosophical mantel of many impurities and prejudices. And now he has wantonly tainted it with the shameful stain of radical evil, in order that Christians, too, might be attracted to kiss its hem” (quoted by Wood 1999, p. 403). [↑](#footnote-ref-1)
2. RGV 6:21 [↑](#footnote-ref-2)
3. Ibid: “but this subjective ground must, in turn, itself be a deed of freedom” [↑](#footnote-ref-3)
4. RGV 6:22n [↑](#footnote-ref-4)
5. RGV 6:31 [↑](#footnote-ref-5)
6. RGV 6:36 “The difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim but rather in their subordination (in the form of the maxim): which of the two he makes the condition for the other. It follows that the human being is evil only because he reverses the moral order of his incentives in incorporating them into his maxim.” [↑](#footnote-ref-6)
7. RGV 6:24 [↑](#footnote-ref-7)
8. RGV 6:29 [↑](#footnote-ref-8)
9. RGV 6:32 [↑](#footnote-ref-9)
10. Ibid [↑](#footnote-ref-10)
11. Wood 1970: pp. 219-226 [↑](#footnote-ref-11)
12. Wood 1999: pp. 283-290 [↑](#footnote-ref-12)
13. Allison 1990: pp. 154-157 and Allison 2002 [↑](#footnote-ref-13)
14. Wood criticized this argument, first that sensuality does not imply unholiness, and secondly that unholiness of the will does not imply the superiority of the inclinations over the moral law [Wood 1999: p. 407 n7]. Allison replied in [Allison 2002]. My position will be closer to Allison's existential interpretation that Wood's social one, with the aid of some insights from the Critique of Judgment. [↑](#footnote-ref-14)
15. RGV 6:45: How it is possible that a naturally evil human being should make himself into a good human being surpasses every concept of ours. For how can an evil tree bear good fruit? But, since by our previous admission a tree which was originally good but did bring forth bad fruits, and since the fall from good to evil is no more comprehensible than the ascent from evil back to good, then the possibility cannot be disputed [↑](#footnote-ref-15)
16. RGV 6:50 [↑](#footnote-ref-16)
17. RGV 6:51-53 [↑](#footnote-ref-17)
18. For further explication for the problematic of the notion of grace see Wolterstorff 1991 and Silber 1960 [↑](#footnote-ref-18)
19. RGV 6:75n [↑](#footnote-ref-19)
20. RGV 6:191 [↑](#footnote-ref-20)
21. RGV 6:52 [↑](#footnote-ref-21)
22. A132 / B171 [↑](#footnote-ref-22)
23. A133 / B172, KU 5:169 [↑](#footnote-ref-23)
24. KU 5:179 [↑](#footnote-ref-24)
25. KU 5:180 [↑](#footnote-ref-25)
26. The appendix to the transcendental dialectic: A642 / B670 – A668 / B696 [↑](#footnote-ref-26)
27. A643/B671 [↑](#footnote-ref-27)
28. KU 5:187 [↑](#footnote-ref-28)
29. KU First introduction 20:211 [↑](#footnote-ref-29)
30. KU 5:192 [↑](#footnote-ref-30)
31. See also Allison 2001: p. 51 [↑](#footnote-ref-31)
32. KU 5:196 [↑](#footnote-ref-32)
33. KU 5:175 [↑](#footnote-ref-33)
34. KU 5:175 [↑](#footnote-ref-34)
35. KU 5:176 [↑](#footnote-ref-35)
36. Davidovich (1994) also argues for the importance of the third critique to the reading of the Religion, however, without a proper account of the uniqueness of reflective judgments. See also Davidovich (1993) for the religious significance of the Critique of Judgment. [↑](#footnote-ref-36)
37. G 4:407 [↑](#footnote-ref-37)
38. KpV 5:122-124 [↑](#footnote-ref-38)
39. RGV 6:62 [↑](#footnote-ref-39)
40. RGV 6:63 [↑](#footnote-ref-40)
41. KU 5:237 [↑](#footnote-ref-41)
42. KU 5:232 [↑](#footnote-ref-42)
43. To make this idea clearer I would need to say more about the relation between ideas of reason and reflective judgments than the scope of this paper allows. My main point here is that with the notion of reflective judgment Kant conceives of a way to articulate the exemplary and concrete use of ideas of reason, which is applied in the *Religion* to the ideal of the son of God [↑](#footnote-ref-43)
44. RGV 6:66 [↑](#footnote-ref-44)
45. RGV 6:67 [↑](#footnote-ref-45)
46. KU 5:183 [↑](#footnote-ref-46)
47. A642/B670 - A668/B696 [↑](#footnote-ref-47)
48. KpV 5:75ff [↑](#footnote-ref-48)
49. KU §27 5:257 ff [↑](#footnote-ref-49)
50. KU 5:300 [↑](#footnote-ref-50)
51. KU 5:301 [↑](#footnote-ref-51)
52. RGV 6:24n [↑](#footnote-ref-52)
53. Ibid [↑](#footnote-ref-53)
54. KU 5:353 [↑](#footnote-ref-54)
55. As Goethe remarked: “at times he [Kant] seemed determined to put the narrowest limits on our ability to know things, and at times, with a casual gesture, he pointed beyond the limits he himself had set”. Goethe, *Scientific Studies*, edited and translated by Douglass Miller, Suhkramp Publishers, New York, 1988. p. 31 [↑](#footnote-ref-55)
56. KU 5:380: "We may consider it [beauty in nature] as a favor that nature has done for us" [↑](#footnote-ref-56)
57. Religion in the broadest sense of ascribing meaning by taking part in an all-encompassing whole [↑](#footnote-ref-57)
58. The account of grace that I present here does not cohere with traditional Christian notions of grace which involve some arbitrariness, but this is the most I can make of it [↑](#footnote-ref-58)
59. SF 7:43 [↑](#footnote-ref-59)
60. RGV 6:83 [↑](#footnote-ref-60)
61. RGV 6:50-51 [↑](#footnote-ref-61)
62. KPV 5:155-6 [↑](#footnote-ref-62)
63. Ibid 5:156 [↑](#footnote-ref-63)