Amit Kravitz*

Kant's Diabolical Evil Reconsidered

https://doi.org/10.1515/jtph-2025-0002
Received February 5, 2025; accepted July 12, 2025; published online July 25, 2025

Abstract: In this paper I focus on the unique challenge to provide a systematic account of the difference between human and diabolical evil in Kant's philosophy – a theme that, as I show, lies at the core of Kant's moral theory. I explain why I think Kant's structural account, which cannot be refuted on grounds of future historical (contingent) events by definition, has been misunderstood in the literature hitherto. I then propose a novel account of Kant's essential exclusion of diabolical evil from the realm of human freedom, showing how this exclusion is related to relevant epistemic limitations concerning the origin of the moral law.

Keywords: evil; finitude; Kant; moral theory; religion

1 Introduction

In RGV[1] Kant uses the terms "diabolical" [*teuflisch*] or "devil" [*Teufel*] six times; two formulations, however, seem to capture the essence of Kant's intent in this regard:

(A) In order to find a ground of moral evil in the human being, sensibility entails too little; since it renders the human being, by removing the incentives which are rooted in freedom, to be a mere animal. However, a reason which is absolved from the moral law, a quasi [*gleichsam*] vicious [*boshaft*] reason (an evil will per se) entails too much, because thereby the collision [*Widerstreit*] with the law would be elevated to be the incentive (for without any incentive the power of choice [*Willkür*] cannot be determined) and this would make the subject a

1 I will use the following abbreviations: *RGV = Die Religion innerhalb der Grenzen der bloßen Vernunft*; *KpV = Kritik der praktischen Vernunft*; *KU = Kritik der Urteilskraft*; *GMS = Grundlegung der Metaphysik der Sitten*; *MS = Metaphysik der Sitten*; *V-MS/Vigil = Vorlesungen Wintersemester 1793/1794 Die Metaphysik der Sitten Vigilantius*; *Prol = Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können*; *KpV = Kritik der reinen Vernunft*; *SF = Der Streit der Fakultäten*. All translations in this article are mine.

Thomas Buchheim gewidmet.

***Corresponding author: Dr. Amit Kravitz**, Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft, Ludwig-Maximilians-Universität, Lehrstuhl für Philosophie I, Geschwister-Scholl-Platz 1 LMU, 8059 München, Germany, E-mail: amit.kravitz@campus.lmu.de

diabolical being. Neither of the two is applicable to the human being (RGV, 6: 35).[2]

(B) The viciousness [*Bösartigkeit*] of human nature is not so much malice [*Bosheit*], if one understands this word in its strict meaning, i.e. as a fundamental attitude [*Gesinnung*] (subjective principle of maxims), incorporating evil as evil into the maxim as the incentive (for this fundamental attitude is diabolical), but rather as a perversity of the heart, which is therefore to be termed an evil heart (RGV, 6: 37).

In RGV, 6: 36 Kant addresses this same issue, however without explicitly mentioning the concept of "devil":

(C) The human being (even the basest [*selbst der ärgste*]) […] never renounces, as it were, the moral law in a rebellious manner (by the withdrawal of obedience).

Now it is clear that Kant is not so much interested in diabolical evil as such, but rather in human evil; the first designates the metaphysical horizon of the second and thus defines its intelligibility. Therefore, it would be mistaken to argue that diabolical evil is not a constitutive component of Kant's moral theory and does not obtain a genuine philosophical value. Kant, indeed, did not have a philosophical interest in classic theological questions concerning diabolical beings (e.g. in the question of whether a being like the devil actually exists and tempts human agents to do evil, or in questions concerning the precise relation between the devil and God, or why God allows the existence of such a being to begin with and so on). Moreover: Kant's usage of theological language and concepts in RGV deserves a separate clarification. Kant thinks that morality should provide the criterion for theology (e.g. the criterion by which the Bible is interpreted). Thus, when Kant raises the question of "whether morality must be interpreted according to the Bible or vice versa" (RGV, 110), his answer seems unequivocal: it is the Bible that should be interpreted according to morality (see also e.g. ST, 41: "reason is the highest interpreter of things related to religion [*Religionssachen*]"; this echoes Kant's formulation in the first Critique, according to which he suggests "moral theology" and not "theological morality"; KrV B, 660). Concerning diabolical evil, this means that Kant is *not* committed to a specific Christian interpretation of this concept but rather conceptualizes this issue in light of his own independent moral framework. Thus, when Kant argues that it is *impossible* for humans to act on a specific ('diabolical') evil incentive, one ought first and foremost

2 Thus, Kant thought that the concept of 'devil' cannot be applied to human beings (it is *only* for the "human being impossible, and nevertheless […] it cannot be overlooked in a system of morality"; MS, 6: 321–322), and not that this concept is impossible. Diabolical evil is "an unreachable ideal" (V-MS/Vigil, 27: 691, i.e. the "idea of extreme evil"; MS, 6: 321–322); and 'ideas' in Kant's philosophy are not incoherent, but only unreachable for human beings.

to elicit and clarify the metaphysical limitation on *human freedom* implied in this determination.

Before delving into details, let me first say something general about metaphysical differences in the context of Kant's moral theory. Kant's morality is based on the determination that human beings are "finite rational beings" (KpV, 5: 26), meaning, "rational beings in nature" (KU, 5: 435). These definitions imply that the human being is neither divine – it is "finite" and "in nature", i.e. contrary to God it is not a "purely intelligible being" (Prol, 4: 344) and cannot be attributed "the utmost perfection" (KpV, 5: 41) – nor is the human being animalistic, for it is "rational", i.e. it is defined as having access to the moral law.

In both cases the difference is metaphysical, i.e. it concerns the respective structure of the will, which cannot be altered by the agent itself according to Kant. For (i) the animalistic lack of moral incentive (of access to the moral law) is a matter of definition and not an outcome of a free act (an animal cannot decide to gain access to the law). In the same spirit (ii) a divine being cannot be ascribed sensual incentives by definition (see my discussion in Section 2). The same holds true (iii) for human beings, for neither their moral incentive nor their sensual incentive is an outcome of free choice; the first "forces itself [*sich aufdringt*]" (KpV, 5: 31) upon them, and the second concerns the fact that they find themselves interwoven in (an already existing) "nature", so that it is *impossible* for humans to "eliminate" (RGV, 6: 28) one of these two constitutive incentives.

Both human and animalistic beings are finite beings; thus, the metaphysical boundary is drawn within finitude (the first finite being obtains access to the moral law, the second – does not). In contrast, the difference between human beings and God concerns a distinction between a finite and a perfect being, which possesses "*holy* will" (KpV, 6: 32). Now how are we to understand the difference between a diabolical and a human being?[3]

3 Kant's determinations concerning the structure of the will of a finite rational being, of a divine being (on which I will elaborate more in the next section) and of an animalistic being constitute the most fundamental Kantian conceptual framework in this regard. In this paper, after unfolding some relevant subtleties within Kant's distinctions, I ask: *Given* this conceptual framework, what might be the meaning of diabolical being? Kant's conceptual framework itself might be criticized, of course. To give two examples: (i) In his 1795 *Philosophische Briefe über Dogmatismus und Kritizismus*, in *Sämmtliche Werke*, vol. 1 (Stuttgart/Augsburg: J. G. Cotta, 1856–61) Schelling vehemently rejected Kant's moral demonstration of the existence of God and Kant's determination that a divine will is essentially linked to morality (see e.g. p. 285: "The idea of a moral God […] does not even have a philosophical side […] it is empty as every anthropomorphic representation"); (ii) In his 1798 *Sittenlehte* Fichte rejects Kant's thesis, which I mention in Section 4, that the emergence of the consciousness of the moral law is a "fact" which cannot possibly be explained (Fichte holds that a "genetic [*genetisch*] explanation" in this regard is possible); see Fichtes *Das System der Sittenlehre nach den Prinzipien der Sittenlehre* (Hamburg: Meiner, 1995), 14.

The answer is not clear at first glance. Surely this difference is not to be conflated with the difference between human and divine, for a diabolical being is not a perfect (holy) being, whose will is determined "*of itself* […] *only* by the representation of the good" (GMS, 4: 414). Nevertheless, the difference between human and diabolical is not to be identified with the difference between human and animalistic, since a diabolical being obtains per definition a relation to the moral law[4] – it is defined as a being that *rebels against it*, thus essentially retaining a relation to the law.[5] This renders the task of drawing a metaphysical boundary *which concerns the formal structure of the will* to be highly challenging. Why?

It is unclear at first glance why it is impossible *on structural grounds* to ascribe to the human being the incentive of "evil as evil" (or "rebellion against the law", i.e. "collision with the law"), for 'rebellion' cannot mean that a diabolical being acts in light of an *evil law*; such a law is absolutely impossible in Kant's philosophy (even for God). It is e.g. possible not to pay back a loan; however, it is utterly impossible to render the rule 'you should not pay back a loan' to be a universal law, for generalization here, according to Kant, leads to contradiction. This is why Kant says in MS that such a being (a devil) "rejects the authority of the law itself, *whose validity he still cannot deny before his own reason*" (MS, 6: 320; my emphasis), and this is the ground of the distinction Kant makes in RGV between corruption of reason itself, which is "absolutely impossible" (RGV, 6: 35), and the incentive "collision [or conflict; *Widerstreit*] with the law", which is only impossible for human beings. So both human and diabolical beings acknowledge the validity of the moral law, and both can reject the law, however not in the name of a different – objectively valid – law; the difference is that it is claimed that a specific *incentive* ('rebellion', 'evil as evil') is essentially denied to an evil human agent.[6] However, what can be the *structural grounds* for such a denial? An evil human agent can reject the law on many grounds

4 Here I agree with Klinge 2018, 162: "malicious reason must have a relation to the moral law". This contradicts positions such as Formosa 2009, 191: "[…] a diabolical being cannot even formulate […] [categorical] imperatives […]". However, in the following sections I will unfold unique problems which cannot be found in Klinge's book.

5 Papish 2018, 106 writes: "Kant's stipulation against diabolical evil […] is meant only to eliminate the possibility that an agent can remove herself from the reach of the law". This position is misleading, for according to Kant even a diabolical being cannot completely release itself from the reach of the moral law; it only resists the law on grounds that essentially cannot be attributed to humans.

6 Kant argues precisely in the same manner in MS, 6: 321–322. There, Kant distinguishes between two incentives: (i) deviating from the law "only as an exception […] (exempting oneself from it occasionally"), and deviating from the law (ii) by "making it [the deviation] a rule". The second "is impossible for a human being". So the question is the following: Why can the second incentive not serve as an incentive for humans?

(see the discussion in Section 3); eliminating one of them *on principal reason* must be accounted for *structurally*.

I will continue as follows: In Section (2) I will focus on central aspects of the structural difference between human and divine beings. In the subsequent two sections I will address the issue of the possible structural difference between diabolical and human wills in two steps: in Section (3), which also entails a discussion of the literature on the issue, I will first explain why the task of finding a structural difference between human and diabolical will is *especially* difficult; I will then present a possible new solution to the problem. After explaining why I think this solution eventually fails, I will present in Section (4) a second new possible solution that works. Section (5) entails a short concluding remark.

2 Human and Divine Will

A rational finite will consists of two components; the first concerns the "will of a human being, insofar as it is affected by nature" (GMS, 4: 387). The will here is merely affected by nature, but "not determined" (MS, 6: 213) by it. The second component concerns the "consciousness of the moral law" (KpV, 5: 121), which ought to determine our will. Notice that for the sake of the possibility of freedom of choice none of these components *determine* the will; the first *affects* the will, the second *ought to* determine it. These two components serve as a basis for two incentives: 'self-love' (see e.g. RGV, 6: 26–27)[7] and "respect for the moral law" (see e.g. KpV, 5: 73; 75–76; 78; henceforth: 'respect').

In light of this, the 'power of choice' [*Willkür*] of a finite rational being has two tasks: first to elevate "both incentives in the same maxim" (RGV, 6: 36) and then to decide upon the inner "subordination" (RGV, 6: 36) within the maxim, meaning, to determine which incentive of the two will serve as the "highest maxim" (RGV, 6: 36). So rational finite beings – good or evil – always relate to *both* incentives (the moral and the sensual), and can "eliminate" (RGV, 6: 28) *neither* of them.

As a part of his discussion on "predisposition to personality" in RGV, 6: 27–28 Kant introduces a further subtlety, which will be crucial for my argumentation in the next section. According to Kant, since the incentive of 'respect' is ascribed to every human being, "even the basest" (RGV, 6: 36), then every human being must have receptivity for it. This receptivity has two senses:[8] the first is "the receptivity of mere respect for the

7 Kant distinguishes there between different senses of 'self-love'; however, they all belong to the same kind, as Kant explicitly says (see e.g. KpV, 5: 22; 34).

8 In KpV Kant also mentions this concept; see e.g. KpV, 5: 152, where he speaks of the "property of our mind [*Gemüth*], this receptivity for pure moral interest […]". See also KU, 5: 197, where Kant speaks of

moral law" (RGV, 6: 27); the second is "the receptivity for the respect for the moral law as a sufficient incentive of the power of choice" (RGV, 6: 27). This distinction is important, since obtaining a mere receptivity for 'respect' is not sufficient for the possibility of a moral action in the full sense of the word, in which 'respect' must also serve as the sufficient (the subordinating) incentive of the power of choice.

Now how are we to understand the structural difference between human and divine will? The following citation seems to capture the essence in this regard:

> An absolute good will [….] can be determined of itself according to its subjective constitution [*Beschaffenheit*] only through the representation of the good. No imperatives can therefore be applied to divine will, and generally to holy will; the 'ought' is out of place here, because the will already and of itself necessarily corresponds with the law. Imperatives, thus, are only formulas which express the relation of objective laws of willing in general to the subjective imperfection of the will of this or that rational being, e.g. the human will (GMS, 4: 414).[9]

The differences can be summarized in the following manner:

(i) Divine will is determined only by *one* component (the representation of the good); human will obtains in contrast always *two* incentives ('respect' and 'self-love'). Thus, no freedom of choice can be attributed to divine will.[10] However, according to Kant's celebrated determination that "a free will and a will under moral laws are one and the same [*einerlei*]" (GMS, 4: 447; see also KpV, 5: 29), 'perfect freedom' does not necessarily entail 'freedom of choice' (in fact, the first excludes the second), but rather: a 'perfectly free' will is one that is *necessarily* determined by reason. Put differently, the possibility of *evil* cannot be attributed to God.

(ii) Given that a divine will is determined *only* through the representation of the good, it follows that it has no relation whatsoever to sensuality;[11] it is not "tempted [or misled; *verleitet*] by sensual inclinations" (MS, 6: 380).

"the receptivity of the mind for moral feeling". However, only in RGV does Kant distinguish between two possible senses of receptivity.

9 Notice that in the present paper I am not interested in the question of whether Kant thought that God exists, or what it means that God is an 'idea', or what it means to merely postulate God's existence etc. Kant unfolds a *conceptual difference* between human and divine will which can be accounted for *independently* of the question of whether God exists. (In the citation above: A divine will is e.g. "determined of itself […] only through the representation of the good"; this determines the *concept* of a divine will, regardless of whether such a being genuinely exists or not).

10 See Insole 2013, 86: "only in the case of God could genuinely free actions be determined by perfect goodness in such a way that God could not do otherwise". However, in Prol, 4: 344 Kant says that "we cannot ascribe freedom to matter […] or to a purely intelligible being, e.g. God, *insofar as His action is immanent*" (my emphasis), but it is possible concerning "the beginning of the world" (ibid). Thus, concerning 'creation' things are more complicated. On this issue see Kain 2021.

11 Even when Kant clarifies the concept of 'creation' (see KpV, 5: 102), he maintains that it would be a "contradiction to say that God is the creator of appearances", for creation can only be related to "things in themselves" (KpV, 5: 102).

(iii) The moral incentive appears for a finite rational being in the form of imperative, precisely because the moral incentive does not *already* determine *of itself* the will: "All imperatives […] designate the relation of an objective law of reason to a will […] which is *not necessarily* determined by this law" (GMS, 4: 413).

Notice that it would be misleading to argue that these differences refer solely to the quantity and not the quality of the incentives, as if a human being obtains two incentives (sensual and moral), whereas a divine being obtains only one incentive (the moral). Given that divine will is not to be ascribed any additional determining component apart from the moral one, it cannot be ascribed '*respect* for the moral law' in the first place, for 'respect' is a predominant property of finitude:

> It ought to be noticed that respect is an effect on feeling, and thus on the sensibility of a rational being; thus, it presupposes the sensuality and the finitude of such beings on whom the moral law imposes respect. Therefore respect cannot be attributed to a supreme being, or even to a being which is free from any relation to sensuality […] (KpV, 5: 76).

Thus, divine will is not determined by the same component which ought to determine the will of finite rational beings ('respect'), but rather it is determined by a *categorically different component* ('representation of the good'). This is precisely the reason why even the concept of 'incentive' cannot be attributed to divine will:

> [O]ne can ascribe a divine will no incentives whatsoever […] [for] by *incentive* […] one understands the subjective determining ground of the will of a being whose reason does not already by its very nature correspond with the objective law (KpV, 5: 71–72).[12]

To illustrate: consider a finite rational agent who throughout its entire moral life with no exception has *always* chosen the good. Can it be said that its will is holy (divine)? No, on structural grounds: this agent must have always subordinated *anew* (for it was each time confronted with the possibility of evil) the sensual incentive to the moral incentive. In contrast, God does not subordinate incentives, and cannot be ascribed the possibility of evil (or 'freedom of choice'), and the moral component which already determines its will is of a different kind ('representation of the good' and not 'respect'). Due to *structural limitations* it is impossible for human will to be divine.

---

12 This holds true as well for the concepts 'maxim' and 'interest': "all three concepts – *incentive, interest* and *maxim* – can be applied only to finite beings" (KpV, 5: 79).

3 Human and Diabolical Will: The Second Receptivity

Now a diabolical will is ascribed an incentive ("evil as evil as an incentive", "collision with the law" etc.) which cannot possibly serve as an incentive for humans. This impossibility of the ascription of diabolical incentives to humans must be accounted for in structural terms.[13]

Before explaining in detail why this challenge is highly unique, we can already reject positions which treat this issue empirically. Silber writes e.g. that Kant's thesis ought to be dismissed because "man's free power to reject the moral law in defiance is an ineradicable fact of human experience";[14] elsewhere he determines in the same spirit that "Kant's ethics is inadequate to the understanding of Auschwitz".[15] These allegations are highly misleading; no historical (or future) event can change the formal structure of a finite rational being, just as the morally good deeds of the so called 'Righteous Among the Nations' in the same historical context which Silber mentions do not indicate that such peoples' wills became divine. The same holds true for Pasternack's position, according to which "the exclusion of diabolical evil may, however, seem naïve. Given the litany of horrors that humans have perpetrated, explanations that come down to mere self-interest may seem just too languid".[16] The possibility or impossibility of the ascription of an incentive to finite rational (human) beings concerns solely the structure of their will, and can never be decided upon on empirical grounds.

Due to similar reasons I think that Bernstein's thesis, according to which Kant's claim must be rejected because it is rooted in Kant's "limited moral psychology",[17] is misleading, as well as Louden's determination that Kant excluded diabolical evil from the realm of human freedom because he seeks to avoid romanticizing evil – that is, that according to Kant "we must not be seduced […] into thinking that some human beings, by virtue of their 'strong' and 'potent' personalities, are somehow above the rest of us, and are not to be judged by the same laws and principles that apply to ordinary human beings".[18] Kant seeks metaphysical criteria to designate

13 I agree with Anderson-Gold 1984, 41 concerning the following determination: "the *metaphysical structure* of human will cannot be devilish". However and as I will show in detail, I reject her thesis that this issue must be located in "social context" (ibid., 36).

14 Silber 1960, cxxxviiif.

15 Silber 1991, 198; see a similar position by Klein 2008, 274.

16 Pasternack 2014, 121; see also Card 2002, 91;212.

17 Bernstein 2002, 36; 41–42. In Section 4 I will say something about psychological explanations in this regard.

18 Louden 2010, 107.

human beings *as (moral) species* – independent of the question of whether a *specific* agent is strong or weak, beautiful or ugly, powerful or impotent.

The main problem in such readings, I think, lies in the implied assumption that the meaning of 'in defiance' is already known, and on account of this preceding understanding, Kant's position is held to be naïve. But Kant's stance must be accounted for in Kant's own terms, i.e. as a *structural account* according to which an agent cannot *by definition* reject the moral law in defiance, *regardless of historical circumstances*. Thus, the interpretative task is to decipher what 'in defiance' for Kant *structurally* means, so that it would entail 'Auschwitz' as well. And as I will show, such an account exists.

Now what does the (exclusively diabolical) incentive 'evil as evil' (or 'collision with the law'), in which the validity of the moral law is nevertheless acknowledged, mean? A first possible explanation concerns the 'predisposition to personality' which was mentioned in Section 2. Notice that according to Kant, 'predispositions' constitute the *structure* of human will ("they are *original*, since they belong to the possibility of human [moral] nature"; RGV, 6: 28); thus, if one wants to find a structural explanation for the exclusion of diabolical incentive from the realm of human evil, 'predispositions' is a good place to start.

Recall that I mentioned that in RGV (but not in KpV or KU), as a part of his discussion on predispositions to the good, Kant introduces two senses of the receptivity for the moral law: the receptivity of *mere* respect for the moral law, and the receptivity for the respect for the moral law as a *sufficient incentive* of the power of choice. In fact, Kant describes here two different kinds (or levels) of relation (access) to the moral law. Let us raise now the following speculation, bearing in mind that 'rebellion' must still retain by definition a relation to the moral law: A diabolical being might obtain only the first but not the second kind of receptivity for the moral law. The reason for this possible state of affairs might be twofold: either (i) a diabolical being *intentionally (freely) gave up* the second relation to the law,[19] or (ii) a diabolical being is *defined* as having ab ovo only the first and not the second kind of access to the moral law.

(i) If renunciation as an outcome of a free accountable decision is the case, it must first be explained how such an (evil) act of renunciation can be thought of in Kant's philosophy.

To grasp this, it is crucial to distinguish between *form* and *ground* of an evil act. Here is a typical citation:

19 Formulations such as "*incorporating* [*aufnehmen*] evil as evil into the maxim as the incentive" (RGV, 6: 37; my emphasis), or speaking of the devil as "*making* it a rule to act against the law" (MS, 6: 320; my emphasis) seem to imply that the action ascribed to a diabolical being is rooted in freedom.

> The human being […] is evil only because it reverses the moral order of the incentives when incorporating them into its maxim: the moral law is elevated indeed to the maxim alongside the law of self-love, however when it becomes aware that the one cannot subsist alongside the other, but rather that one must be subordinated to the other as its supreme condition, it renders the incentive of self-love and its inclinations to be the condition of compliance with the moral law […] (RGV, 6: 36).

Notice that Kant suggests here merely a *description* of evil action, but not an account of the *determining ground* which generated this reversed order. The description seems clear enough: an evil action is an action in which the incentive of 'self-love' subordinates the incentive of 'respect' (whereas a morally good action is an action in which 'respect' is the subordinating and not the subordinated incentive). However the *ground* for why an agent brought about the evil (reversed) order instead of the morally good order can be diverse; one agent might decide to reverse the right (moral) order of the incentive due to boredom, a second agent out of curiosity to experience what an evil action feels like (since evil fascinates), a third due to weakness of the will, a fourth due to habit, laziness, spite,[20] and so on.[21]

Thus, when Kant asks "*which of the two [incentives] does he* [*er*; meaning, the free agent] *make the condition of the other*" (RGV, 6: 36), the word 'he' plays a crucial role; it means that *neither* 'respect' *nor* 'self-love' alone serves as the determining ground of the will. Rather, the determining ground of the will is the free *decision* of the agent to elevate 'self-love' or 'respect' to be the subordinating incentive. The ground for the decision concerning the order of the incentives is not to be identified with the two incentives themselves, or else the human agent would be a mere marionette, whose will is determined sometimes by 'self-love', sometimes by 'respect', *but never through its own free decision*.[22]

Now we can better conjecture what a (possible) diabolical renunciation of the second kind of receptivity for the moral law might mean. To grasp this, notice first

---

20 One might argue that 'out of spite' is precisely the incentive 'evil for evil' Kant refers to; see e.g. Michalson 1990, 73: "Devilishness […] would mean the rejection of the moral law precisely because it *is* the moral law." But no *structural ground* seems to prevent the ascription of such an incentive to humans. The explanation, as we shall see in the next section, lies elsewhere.

21 Linking 'evil' to 'freedom' is a special challenge in Kant's conceptual framework; given that "a free will and a will under moral laws are one and the same [einerlei]" (GMS, 4: 447), it is not clear in which sense precisely an evil action can be considered 'free' (see here also MS, 6: 227, where Kant says that the possibility to deviate from the law. i.e. to carry out an evil act, is 'incapacity [*Unvermögen*]'). On the specific problematic of linking 'evil' to 'freedom' in Kant's philosophy and the many attempts carried out by post-Kantian philosophers (e.g. Reinhold, Fichte, Schelling) to overcome Kant's maze in this regard see Noller 2015.

22 This was Karl Leonhard Reinhold's celebrated critique; this problem is designated by Allison 1996, 422 as "Reinhold's dilemma" and by Kosch 2006, 55 as "Reinhold's complaint".

that it is absolutely possible for a rational finite being to *want* to change the structure of its will, since no structural ground prevents it from wanting it. (No structural ground prevents us e.g. from *wanting* or longing to be divine, or from longing to be devoid of inclinations;[23] structural grounds only prevent us from *becoming* divine, or from becoming devoid of inclinations).

Now consider a being which is originally ascribed the same structural relation to the moral law like us – it obtains the first kind of receptivity for the moral law as well as the second kind and thus possesses the full possibility to obey the law. At some point it wants to release itself from the troublesome burden of the demands of the moral law which it loathes, i.e. it wants to alter the structure of its will. It still wants, however, to retain a vital relation to the moral law, for it wants to perpetuate and to steadily recall so-to-say its rejection of the law (that is, it does not want to become merely animalistic). This can be done by renouncing the second but not the first kind of receptivity for the law, wherefore this being still retains a relation to the law but renounces the possibility to render the moral incentive to be the sufficient incentive of its will. This is one possible structural meaning of 'rebellion' against or 'collision' with the law, which yields a genuine *structural difference* between human and diabolical will: Human will obtains two kinds of receptivity for the law; diabolical will originally obtained these two kinds of receptivity, but after renouncing the second, now obtains only the first kind.

However, the possibility of altering structures renders the very term 'structure' in the Kantian context empty. The terminology Kant uses time and again in this regard attests to it – it is *impossible* for a human evil agent to obtain a diabolical incentive, *impossible* for a human good agent to be divine, *impossible* for divine to be human (for the 'possibility of evil' cannot even be ascribed to God), *impossible* for a human agent to become animalistic (i.e. to renounce the consciousness of the moral law) and so on. The possibility of altering structures cannot be seen as a *structural component*; rather, it *negates the very concept of 'structure'* in the Kantian context.

(ii) If a diabolical being is *defined* as having only the first and not the second kind of access to the moral law, then the original impossibility to act morally yields the structural meaning of 'rebellion'. In this case a diabolical rebellion against the law is lawful, i.e. it serves *by definition* as a *determining* ground of a diabolical will.[24] Now 'lawful' is not to be confused with 'always'; if rebellion defines the

23 In KpV, 5: 118 Kant explains why inclinations steadily leave a greater and greater "void" and thus are "always *burdensome* [*lästig*]" for a rational being; according to Kant, this necessarily generates "the wish to get rid of them".

24 Wood 1970, 213 expresses this position by arguing that a diabolical rebellion is rooted in the devil's predisposition, that is, in a component that *defines* diabolical will (and not as something *contingent* that was acquired by means of free decision); see also O'Connor 1985, 290. Wood does not tackle the problematic I am about to present.

structure of a diabolical will, then refraining from rebellion must be considered impossible. In this case, however, we can ascribe to diabolical will neither 'freedom of choice' (human freedom), nor 'perfect freedom' (divine freedom). In Kant's philosophy there is only one case in which a will is considered free despite the fact that it is *defined* as *already* being *determined* by a certain component: a holy (divine) will. But if the devil cannot be attributed any sense of freedom whatsoever, what makes it *evil*? Such a view contradicts[25] the very concept of diabolical evil.[26]

4 Human and Diabolical Will: The Question Regarding Rationalization

In order to see if there is nevertheless a way to maintain the structural difference between human and diabolical evil without excluding 'diabolical evil' from the realm of freedom,[27] it might be useful to systematically examine sections in which Kant speaks of the primordial reaction of human beings *as a moral species* to the moral law;[28] for even if the concept of diabolical evil is not mentioned in these sections explicitly, it pertains directly to the issue at stake.

One revealing formulation in this regard can be found in Kant's discussion on the "origin of evil in human nature" (RGV, 6: 39ff.); notice that Kant stresses time and again that he is referring to the human "considered in its species" (RGV, 6: 32; see also RGV, 6: 20; 21; 25; 29). In his account on this issue Kant at some point describes the

25 Allison 1996, 176 writes: "Kant's denial of a diabolical will is […] an *a priori* claim about the conditions of the possibility of moral account". However, if the devil is absolutely out of the reach of moral account, it is not clear if its concept (as an *evil* being) is intelligible at all.

26 So despite the fact that Kant, indeed, implies that a diabolical being is devoid of sensuality, the question of whether a diabolical being has any relation to sensuality or not is not sufficient to constitute a *relevant* structural difference between human and diabolical evil (Klinge 2018, 165 thinks that this is the relevant structural difference).

27 Thus, I utterly agree with Samet-Porat 2007, 85–91, when she writes: "Satanic agents seem to know what they are doing […] satanic actions are, to a large extent, reflective actions […] In the case of the satanic agents […] the motivation is inversely built upon a proper judgment of our moral duty. A person who cannot distinguish between good and bad, who loses her capacity to understand the demands of morality, should be treated instead of punished […] It is a necessary condition for satanic wickedness that the agent accepts morality as binding and therefore recognizes his actions as morally bad". However, as I will now argue and contrary to her interpretation of Kant, one can elicit from Kant's account precisely such an account of diabolical being.

28 A *primordial reaction* to the moral law, which is thought of as designating *every* agent in the species without exception (i.e. human "considered in its species", RGV, 6: 32), is also termed by Kant "radical evil" (RGV, 38) – radical, for it concerns the "root of evil" (RGV, 6, 39Fn.)

"way of presenting [*Vorstellungsart*]" this issue in the 'Scripture' (RGV, 6: 41).[29] Before delving into this issue, let me say something about the systematic status of this presentation. Kant says that the Scripture describes the "*beginning* of the origin of evil in the human species" (RGV, 6: 41). By determining this, Kant is neither (i) interested in analyzing the falling into evil of a specific agent (the First Adam), nor (ii) does he raise a causal claim (as if the fall of the First Adam into evil caused the evil of all other free agents – each agent accounts for its own deeds according to Kant). Rather, the Scripture presents how finite agents *considered as a species* react to the moral law. *Considered as a species* does not mean (iii) that the species itself is reified and treated as an accountable agent; it means that Kant thinks that *all* agents *freely* react to the moral law in this primordial context in the same manner (radical evil), i.e. that "the grounds that justify ascribing it [freely acquiring an evil rather than a good fundamental attitude] to *one person* are constituted in a way that there is *no reason to exclude* any other human being from it" (RGV, 6: 25, my emphasis). So according to Kant the reaction to the law, though rooted in freedom (i.e. essentially contingent), can nevertheless be proven[30] to be attributed to the entire species, meaning, to every agent without exception.[31] Now how is this primordial reaction to the law (radical evil) described by Kant?

The point of departure must be thought of as the "stage of innocence [*Unschuld*]. The moral law preceded, as a prohibition, the human being, which is not a pure being

29 It is worth noting that Kant's take on positive religions, inherited representations obtaining ethical significance etc. is highly complicated. Kant says reading the Bible in light of morality should be carried out "as far as it is possible" (RGV, 6: 110), demonstrating by this awareness that the content of historical religions, which by definition contain the "consciousness of [their] contingency" (RGV, 6: 115) *cannot* be fully reduced to morality. On this issue see e.g. Stephen R. Palmquist, "Does Kant reduce religion to morality?" *Kant-Studien* 83, 1992: 129–148; John E. Hare, *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance* (Oxford UK: Clarendon Press, 1996); Peter Byrne, *Kant on God* (Aldershot UK: Ashgate, 2007). Given Kant's determination that "morality leads *inescapably* to religion" (RGV, 6: 6; see also KpV, 5: 129), it seems that humans will always stand in some relation to the realm of positive religions, inherited representations etc.; on this issue see Kravitz 2022a.

30 Kant explicitly uses "to prove" here; see RGV, 6: 25, 30; there is an ongoing debate in the literature regarding Kant's argument, which lies outside the scope of this article. For some examples see O'Connor 1985; Michalson 1990, 14ff.; Allison 1990, 154–157; Buchheim 2001; Morgan 2005; Formosa 2007; Palmquist 2008; Muchnik 2010; Papish 2018, 117–153. Examples of anthropological readings of Kant's proof are e.g. Wood 1999, 288–289; Sussman 2001, 18.

31 Recall that Kant's entire discussion concerns the *primordial reaction* to the moral law, a reaction which includes *every* agent in the species. This context is categorically distinct from *ordinary moral actions*, in which agents differ from each other (one agent chooses the good; the second chooses evil on grounds which has nothing to do with self-deception, and so on). Thus, the fact that Kant "proves" that the fundamental attitude of every agent is evil and not good, and further that this primordial reaction is linked to self-deception, does not mean that every ordinary evil action must be described in terms of 'self-deception' (or that every ordinary free action must be evil at all).

but a being which is tempted [*versucht*] by inclinations" (RGV, 6: 42). So no determination of the will has taken place yet. What happens next? How does the agent react to the demand of the law? Kant writes (I have marked the three relevant verbs):

> Instead of simply to follow this law as sufficient incentive (which is alone unconditionally good, thus there is no place for further deliberations [or qualm; *Bedenken*], the human being [i] has looked around [*umsehen*] for other incentives […] which can be good only conditionally […] and has made it – if one thinks of the action as arising consciously from freedom – a maxim to follow the law of duty not out of duty but at best [*allenfalls*] out of consideration for other aims. As a consequence the human being began to doubt the strictness of the command which excludes the influence of any other incentive, and afterwards [ii] to rationalize-in-order-to-downgrade [or to quibble; *herabvernünfteln*] the obedience to be a mere (under the principle of self-love) conditional obedience as a means, and eventually the preponderance of the sensual impulses over the incentive rooted in the law [iii] was elevated [*aufgenommen*] in the maxim and thus sin came about […] (RGV, 6: 42).

The three verbs Kant uses here correspond with the two components of an evil act; the third verb (to elevate [*aufnehmen*] 'self-love' over 'respect') is a *description* of an evil act. The first two constitute the *ground* of the evil act.[32]

RGV is not the first work in which Kant uses the verb '*vernünfteln*' to describe a similar mechanism; In GMS, 4: 405 for instance, one can find an almost identical formulation (however, without distinguishing 'looking for' as preceding 'rationalizing'). Kant speaks there of a "natural dialectic", i.e. of the

> propensity [*Hang*] to quibble [*vernünfteln*] against these strict laws of duty and to cast doubt upon their validity, or at least upon their pureness and strictness and to make them, where possible, more suitable to our wishes and inclinations, that is, to corrupt them in their ground […].

It seems that the mechanism Kant describes ('rationalization') has something to do with self-deception;[33] self-deception, however, is not that easy to conceptually account for, as Kant well knew (see MS, 6: 430).

In the last few years there has been increasing interest in this issue in the literature.[34] To give a few representative examples: Green thinks that all three kinds of "natural propensity of humans to evil", and not only the third kind, represent a form of self-deception.[35] In contrast, Rukgaber argues that the first kind, i.e. "Kantian

32 In this section Kant actually unfolds the mechanism that leads to the third kind of the "natural propensity of humans to evil" (RGV, 6: 28ff.): "the propensity to adopt evil maxims, i.e. the viciousness [*Bösartigkeit*] of human nature or of the human heart".

33 Sticker 2017 simply calls the 'natural dialectic' the "natural propensity to self-deception".

34 For the question of how *vernünfteln* can be overcome according to Kant see Noller 2021, 47–50.

35 Green 1992.

frailty [*Gebrechlichkeit*] is a form of practical irrationality *without* self-deception".[36] Welsch, as well, rejects Green's position and holds that only the third kind of original evil can be identified with lying to oneself.[37] Papish says that this original act cannot be fully identified with self-deception, but nevertheless a "dissimulation" is involved.[38] Callanan says that the natural dialectic mentioned in GMS is "a kind of self-deception whereby the subject undermines morality while still paying lip service to it"[39] and concludes that according to Kant, and contrary to Rousseau's position, reason is not responsible for this self deception.

Now the question concerning the adequate *positive* understanding of rationalization in this context is of no systematic relevance to the issue at stake; much more important is the *negative* aspect of Kant's description. For whatever rationalization means, it does *not* concern a direct rejection of the law; rather, what the agent does is a bit different: facing the demand of the law, the agent "has made it […] a maxim *to follow the law of duty* [however] not out of duty" (I am interested here solely in the outcome of the process Kant describes in the long citation above from RGV, 6: 42, and not in the stages leading to it, such as beginning "to doubt the strictness of the command" etc.). From this we learn that the primordial confrontation with the law consists of two possibilities: the agent either (i) follows the law (the moral possibility) or (ii) begins a process of rationalization.

Notice that these two possibilities of the primordial reaction to the moral law described by Kant are not constituent components of the structure of the will; it is not the *structure* which is at stake here, but the free (contingent) *usage* of it. However, the formal structure of the will and possible free usages of it shed light on each other. That we can infer from the structure of the will to possible usages of it is clear enough, I think; for instance, based on the structure of the will of a finite rational being as obtaining per definition two incentives (sensual and moral) we know that neither a divine nor an animalistic usage of the will is possible for human agents. But a structural conclusion can be inferred from possible (or impossible) usages of the will as well. If the primordial reaction to the moral law – besides fully adhering to it – is *rationalization* and not *rejection*, it must be rooted in an *essential limitation* that is attributed to us. Which limitation might Kant be implying here?

The answer lies in the manner in which the law is given to humans. Kant's systematical position is that in the realm of morality explanations concerning *emergence* count as transgressing the limits of cognition. This is why Kant terms the

36 Rukgaber 2015, 235.

37 Welsch 2019, 53: "the ascription of self-lying of the third kind to all kinds […] is an interesting, however untenable interpretation".

38 Papish 2018, 117–153.

39 Callanan 2019, 16.

consciousness of the moral law an "*inexplicable* fact [*unerklärliches Factum*]" (KpV, 5: 43, my emphasis). Similarly, Kant says that reason "imposes itself" (KpV, 5: 85) upon us, and determines further: "how the consciousness of the moral law […] is possible cannot be further explained" (KpV, 5: 46; see also KpV, 5: 72). Elsewhere Kant maintains that "the law finds *on its own* entrance [or access; *Eingang*] into the mind [*Gemüth*]" (KpV, 5: 86; my emphasis). Thus, any explanations concerning the emergence of moral consciousness are impossible; the awareness of the moral law *without knowing its origin* is an original irreducible fact of our existence as rational beings and serves as an ultimate point of departure in this regard. The point of departure of morality for finite rational beings entails therefore an essential *epistemic limitation*.

Now let us reconstruct a possible hypothetical being which is, just like a finite human being, confronted with the same law, but is epistemically superior in comparison to humans in the sense that it can cognize the way of emergence of the moral law – i.e. it sees where the law comes from so-to-say, its ultimate origin (God; see my discussion in Section 5). Such a being, if it decides to reject the law, *cannot rationalize* (in the sense mentioned above), for we can rationalize and deceive ourselves only concerning something which we cannot tell its ultimate origin; only a being which is epistemically inferior in this regard might try to convince itself that the law cannot come into force without its subjective agreement. For this is what the mechanism of rationalization is actually all about; it relates the unconditional to one's own subjective will, i.e. it interprets the demand of the law by means of one's subjective inclinations. The essential lack of knowledge concerning the origin of the law enables such a mechanism. A being which is epistemically superior in this sense can only reject the law directly (call it 'collision' or 'rebellion') and not by means of rationalization and quibbling. The impossibility to attribute diabolical evil to humans is rooted, thus, in the essential epistemic superiority[40] of a diabolical being.[41]

Such a hypothetical being can also be ascribed freedom of choice, but it is categorically different from humans. This holds true for the (i) primordial decision for evil, i.e. before evil took place; both kinds of beings are confronted with the law; the human can either follow the law (then it is morally good), or set up a rationalization process which ends up in the reversal of incentives (then it is evil). The

40 That the difference between human and diabolical evil is rooted in an essential *cognitive* limitation can be found in other philosophers in Kant's time. To give a less known example: Erhard 1795, 125: "the human in its theoretical cognitions is too limited […] thus, the ideal of malice cannot be thought of as a possibility for humans".

41 So here we see why Kant's discussion is not to be understood as a mere psychological one; the fact that 'rationalization' designates our primordial relation to the law might have some interesting *psychological consequences*, but this fact is not a *consequence of psychology*; rather, it is the outcome of an *essential epistemic limitation*. This limitation is just as little a psychological claim as Kant's contention that 'things in themselves' are beyond the reach of human cognition.

diabolical, due to its superior epistemic status concerning the origin of law, can either rebel against it (in this case it is diabolical), or adhere to it (then it is a good angel). But no less important is that (ii) this holds true for the possibility of restoration of the good as well. For both kinds of beings restoration is possible, but categorically distinct. As for humans, they are "beings-in-the-world [*Weltwesen*]", thus subjected to "continually becoming" (RGV, 6: 74–5). This is why the departure from original evil for humans concerns an endless, ongoing process. A diabolical being on the other hand has a direct relation to the law; for such a being, thus, rehabilitating the good (if it freely decides to alter its rebellious relation to the law) does not involve a process at all; rather, its goal is immediately achieved, with no need of mediating links whatsoever.[42]

Let me address five possible misconceptions.

Firstly (i), the human agent recognizes the *authority* of the law as well as the *validity* of the law *independently* of the question concerning its origin. (For instance, the moral law binds atheists as well,[43] and the validity of the moral law concerns the concept of contradiction and not knowledge of its origin). Not knowing the origin of the law is simply another aspect of human finitude (along with possessing sensual incentive next to the moral incentive), an aspect which changes *nothing* regarding Kant's familiar account of moral action: a morally good action still means subordinating the sensual to the moral incentive *without knowing the last metaphysical origin of moral consciousness*, and a morally evil action still means subordinating the moral to the sensual incentive *without knowing the last metaphysical origin of moral consciousness*.

Secondly (ii), it is not claimed here that a human agent would be *morally better or worse* if it knows the origin of the law. Knowing the origin does not render a being better or worse, but categorially different; to wit, it enables either a rejection of the law *in a way which is impossible for humans*, or an acceptance of the law *in a way which is impossible for humans*. Put differently, humans would not earn or lose moral points if they obtained cognition of the origin of the law; they would simply be different *kinds* of beings.

---

42 To illuminate with an example: according to Kant, even if a specific agent has decided to break off with evil, other subjects still serve as "seductive examples"; subjects "mutually corrupt" each other. Thus, some "means" must be found to prevent it, a "society" whose aim is to "preserve morality" (RGV, 6: 94). Since a diabolical being has a direct relation to the origin of the law, it cannot be corrupted by others and does not need to establish a "republic according to the laws of virtue" (RGV, 6: 98) to maintain its (possible) moral turn.

43 See e.g. Kant's clear words on the moral demonstration of the existence of God in KU, 5: 450–1: "This proof […] does not mean that it is just as necessary to assume (annehmen) the existence of God as it is to acknowledge the validity of the moral law, hence that whoever cannot convince himself of the former can judge himself to be free from the obligations of the latter. No!".

Thirdly (iii), my account explains what 'rebellion' (or rejection of the law 'in defiance') in Kantian terms means. Rejection of the law *while knowing its origin* (the diabolical case) is 'rebellion' according to Kant (this, and *not* our preconception of this concept). Rejection of the law while *not knowing its origin* (the human case) is 'rationalization' (this, and *not* our preconception of it). Now it is clear why assertions such as Silber's ("Kant's ethics is inadequate to the understanding of Auschwitz") or Pasternack's ("the exclusion of diabolical evil may, however, seem naïve, given the litany of horrors that humans have perpetrated") and other positions I have cited are misleading. *In Kantian terms*, Auschwitz changes nothing in the definition of 'rebellion' against the law, for no one would seriously argue that Goebbels for instance[44] was *not* a finite human agent, i.e. that he possessed knowledge of the origin of the law. Kant could not possibly have known many things about future developments (the industrialization of death like that which took place in Auschwitz, for instance, was surely beyond his intellectual and scientific horizons); however, this has nothing to do with Kant's *structural claim* concerning the epistemic limitations of finite beings as such.

Fourthly (iv), one might claim that of all things, knowing the origin might render 'rationalization' easier rather than impossible. But this cannot be the case here, for two reasons.

The first reason (a) concerns the inability of a diabolical being to deceive itself by rationalization; a diabolical being is *defined* by the knowledge of the divine origin of the law, i.e. it is defined by *not* being able to ignore it (and by ignoring it – to downgrade it); if a diabolical being starts ignoring the origin of the law, this strictly amounts to altering the very structure of its will. Put differently, knowing the origin of the law is not a temporal achievement or an accidental property of a diabolical being; it is a constitutive component of its being. Thus, it obtains by definition no leeway which can pave the way for 'rationalization'.

The second reason (b) concerns the inability of a diabolical being to deceive God by rationalization. To illustrate: Say that my wife is worried about our not saving enough for retirement and thus opposes my buying a new car, and say that I know the *origin* of her anxiety in this regard (her family was poor). Knowing the origin of her anxiety is precisely what *enables* me to rationalize buying a new car – say, by pointing out that we have enough savings, or by deceiving her about our financial state. However, Kant's point about diabolical beings and God is categorically different, for it does not concern the relation between two finite beings; a diabolical being cannot deceive *God*, for God knows *by definition* that the devil knows the origin of the law (omniscience is a divine attribute; see e.g. KU, 5: 444), and knows further

44 To take Samet-Porat's example (Samet-Porat 2007, 91).

that the devil cannot ignore this knowledge. The distinct metaphysical context makes all the difference.

Given these two points, it is clear why *on Kant's account* 'rationalization' can be ascribed only to humans.

Fifthly (v), one might ask what is the *ground* according to which diabolical beings act. To wit: A human rejection of morality involves reference to some end that the will aims to promote (rejecting 'respect for the law' in order to promote 'self-love'). However, what is the end which a diabolical being, defined purely negatively (in Kant's formulations, "evil as evil as an incentive", "rebellion against the law", "collision with the law" etc.), wishes to promote by rejecting morality? Recall that, as I have indicated at the outset, diabolical evil is according to Kant "an unreachable ideal" (V-MS/Vigil, 27: 691), i.e. the "idea of extreme evil" (MS, 6: 321–322); and 'ideas' in Kant's philosophy are not incoherent, but only unreachable for human beings. Thus, if rejection of morality involves reference to some end, this must be applied to a diabolical rejection as well.

Notice that unlike human beings, a diabolical being does not obtain a sensual aspect (see footnote 26) – meaning, it is not a being which is already embedded per definition *in* a sensual 'worldly' context (it is not *Weltwesen*; see e.g. KU, 447), a context in which, among other things, *other* humans play a constitutive role as well ("*comparing* [*vergleichende*] self-love", RGV, 27). Therefore, whatever the *non* moral aim a diabolical being wishes to promote, it cannot be thought of as preferring 'self-love' *in the human (sensual or inter-subjective) sense*; a diabolical rejection of morality must be thought of as devoid of reference to human 'self-love' interests of any kind. It follows that the utter destruction of morality – an aspiration that can be ascribed *only* to a being which is confronted with the *origin* of the law – serves as its ultimate incentive.

5 Concluding Remark

The challenge of drawing a metaphysical boundary between human and diabolical evil discloses, I think, a very deep feature of Kant's moral theory which is often ignored or misunderstood. Kant's moral theory is usually praised (or condemned) for its formalism, which is understood, among other things, as a claim that morality is essentially independent of the concept of God. However, the precise sense of this independence is not always clear. The fact that according to Kant finite rational beings are expected to obey the moral law independently of their belief in God stands beyond any doubt, just as Kant's claim that the touchstone of whether a certain practical rule might serve as law is not the question concerning its conjectural divine origin, but the possibility of its universalization without contradiction. However,

taking the general framework of Kant's morality into account and the metaphysical differences constituting it sheds a more subtle light on the inherent relation between morality and divinity.

In a somewhat paradoxical manner, morality as an independent (autonomous) project of finite rational beings is only possible given an essential cognitive limitation, namely that knowledge of the ultimate origin of the point of departure of morality (the consciousness of the moral law) is not accessible to them. This limitation constitutes the horizon to which we are unavoidably led from morality itself, as Kant well knew: "morality leads *inescapably* [*unumgänglich*] to religion" (RGV, 6: 6; my emphasis) (or: "the moral law [leads] […] to religion"; KpV, 5: 129), i.e. to "observing [*Beobachtung*] all human [moral] duties as divine command [*Gebot*]" (RGV, 6: 84). This determination is not meant to designate contingent, weak agents, who are in need of some relation to divinity in order to grasp (or to implement) their moral duty; rather, this is an essential contention about the *situatedness* of morality itself, i.e. about the inescapable context in which morality sprouts. In this sense morality is not just about *formal duties*, but rather about the "inescapable" transition from 'morality' to 'religion',[45] i.e. from the consciousness of the moral law to its (divine) origin, to which we can only practically (endlessly) approach but never theoretically cognize. More than the structural differences between human and animalistic will and human and divine will, which are relatively easy to grasp, it is the unfolding of the intricate structural background that renders diabolical evil impossible for us which ultimately reveals the essential framework of Kant's moral theory as a whole.

References

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

Allison, Henry E. 1996. *Idealism and Freedom. Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press.

Anderson-Gold, Sharon. 1984. "Kant's Rejection of Devilishness: The Limits of Human Volition." *Idealistic Studies* 14: 35–48.

45 This transition from "morality" to "religion" (i.e. God) is usually linked to Kant's treatment of the "highest good" (namely, to the idea of God as guarantor of the highest good). However, Kant explicitly links the practical need to assume the existence of God to the issue of "radical evil" as well. To elaborate a bit: After proving that the point of departure of humans, considered as species, is "radical evil", i.e. a state of affairs in which "evil has already taken its place [schon Platz genommen hat]" (RGV, 6: 57Fn.), Kant argues that humans cannot "destroy [vertilgen]" this "radical evil" *alone* (RGV, 6: 31) – i.e. that a "divine aid [übernatürliche Mitwirkung]" (RGV, 44) is inescapable. On this issue in detail see Kravitz 2022b.

Bernstein, Richard J. 2002. *Radical Evil: A Philosophical Interrogation*. Cambridge: Cambridge University Press.

Buchheim, Thomas. 2001. "Die Universalität des Bösen nach Kants Religionsschrift." In *Kant und die Berliner Aufklärung*, edited by V. Gerhardt, R. Horstmann, and R. Schumacher, 652–61. Berlin: De Gruyter.

Byrne, Peter. 2007. *Kant on God*. Aldershot: Ashgate.

Callanan, John J. 2019. "Kant on Misology and the Natural Dialectic." *Philosophers' Imprint* 19: 1–22.

Card, Claudia. 2002. *The Atrocity Paradigm: A Theory of Evil*. Oxford: Oxford University Press.

Erhard, Johann Benjamin. 1795. "Apologie des Teufels." In *Philosophisches Journal einer Gesellschaft Teutscher Gelehrten I*, edited by Johann Gottlieb Fichte, and Friedrich Immanuel Niethammer, 105–40. Neu-Strelitz: Michaelis.

Fichte, Johann Gottlieb. 1995. *Das System der Sittenlehre nach den Prinzipien der Sittenlehre*. Hamburg: Meiner.

Formosa, Paul. 2007. "Kant on the Radical Evil of Human Nature." *The Philosophical Forum* 38: 221–45.

Formosa, Paul. 2009. "Kant on the Limits of Human Evil." *Journal of Philosophical Research* 34: 189–214.

Green, Michael K. 1992. "Kant and Moral Self-Deception." *Kant-Studien* 83: 149–69.

Hare, John E. 1996. *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance*. Oxford: Clarendon Press.

Insole, Christopher J. 2013. *Kant and the Creation of Freedom: A Theological Problem*. Oxford: Oxford University Press.

Kain, Patrick. 2021. "The Development of Kant's Conception of Divine Freedom." In *Leibniz and Kant*, edited by Brandon C. Look, 295–319. Oxford: Oxford University Press.

Klein, Patrick. 2008. *Gibt es ein Moralgesetz, das für alle Menschen gültig ist? Eine Untersuchung zum Faktum der Vernunft bei Immanuel Kant*. Würzburg: Königshausen & Neumann.

Klinge, Hendrik. 2018. *Die moralische Stufenleiter. Kant über Teufel, Menschen, Engel und Gott*. Berlin: De Gruyter.

Kosch, Michelle. 2006. *Freedom and Reason in Kant, Schelling, and Kierkegaard*. Oxford: Clarendon Press.

Kravitz, Amit. 2022a. "Revelation's Entrenchment in Pure Reason in Fichte's *Versuch einer Kritik aller Offenbarung*." *Kant-Studien* 113: 299–329.

Kravitz, Amit. 2022b. "The Reason for Miracles and the Miracles in Reason: Kant's Conception of Practical Miracles." *Kantian Review* 27: 237–56.

Louden, Robert B. 2010. "Evil Everywhere. the Ordinariness of Kantian Radical Evil." In *Kant's Anatomy of Evil*, edited by Sharon Anderson-Gold, and Pablo Muchnik, 93–115. Cambridge: Cambridge University Press.

Michalson Jr, Gordon E. 1990. *Fallen Freedom. Kant on Radical Evil and Moral Resignation*. Cambridge: Cambridge University Press.

Morgan, Seiriol. 2005. "Kant's Missing Formal Proof of Humanity's Radical Evil in Kant's *Religion*." *Philosophical Review* 114: 63–114.

Muchnik, Pablo. 2010. "An Alternative Proof of the Universal Propensity to Evil." In *Kant's Anatomy of Evil*, edited by Sharon Anderson-Gold, and Pablo Muchnik, 116–43. Cambridge: Cambridge University Press.

Noller, Jörg. 2015. *Die Bestimmung des Willens. Zum Problem individueller Freiheit im Ausgang von Kant*. Freiburg: Karl Alber.

Noller, Jörg. 2021. "Logik der Schein. Kant über theorethische und praktischen Selbsttäuschung." *Kant-Studien* 112: 23–50.

O'Connor, Daniel. 1985. "Good and Evil Disposition." *Kant-Studien* 46: 288–302.

Palmquist, Stephen R. 1992. "Does Kant Reduce Religion to Morality?" *Kant-Studien* 83: 129–48.

Palmquist, Stephan. 2008. “Kant’s Quasi-Transcendental Argument for an Necessary and Universal Evil Propensity in Human Nature.” *The South Journal of Philosophy* 46: 261–97.

Papish, Laura. 2018. *Kant on Evil, Self-Deception, and Moral Reform*. New York: Oxford University Press.

Pasternack, Lawrence R. 2014. *Kant on Religion Within the Boundaries of Mere Reason*. London: Routledge.

Rukgaber, Matthew S. 2015. “Irrationality and Self-Deception Within Kant’s Grades of Evil.” *Kant-Studien* 106: 234–58.

Samet-Porat, Irit. 2007. “Satanic Motivations.” *The Journal of Value Inquiry* 41: 77–94.

Schelling, Friedrich Wilhelm Joseph. 1856–61. “Philosophische Briefe über Dogmatismus und Kritizismus.” In *Sämmtliche Werke*, Vol. 1, 283–341. Stuttgart: J. G. Cotta.

Silber, John R. 1960. “The Ethical Significance of Kant’s Religion”. In Immanuel Kant, *Religion Within the Limits of Reason Alone*, trans. by Theodore M. Greene, and Hoyt H. Hudson. New York: Harper & Raw.

Silber, John R. 1991. “Kant at Auschwitz.” In *Proceedings of the Sixth International Kant Congress*, edited by Gerhard Funke, and Thomas M. Seebohm, 177–212. Washington: Center of Advanced Research in Phenomenology and University Press of America.

Sticker, Martin. 2017. “When the Reflective Watch-Dog Barks. Conscience and Self-Deception in Kant.” *Journal of Value Enquiry* 51: 85–104.

Sussman, David G. 2001. *The Idea of Humanity: Anthropology and Anthroponomy in Kant’s Ethics*. New York: Routledge.

Welsch, Martin. 2019. “Kant über den Selbstbetrug des Bösen.” *Kant-Studien* 110: 49–73.

Wood, Allen W. 1970. *Kant’s Moral Religion*. Ithaca: Wiley-Black.

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