

The Systematic Relation between the Notions of Predisposition to Good and Propensity to Evil. Some Remarks on an Asymmetry in Kant's Theory of Moral Evil

SEBASTIÁN CABEZAS¹

Abstract

Kant's *Religion Within the Boundaries of Mere Reason* has been a controversial text within Kant scholarship since its earliest reception. One of the main reasons for the critical reception relates to Kant's theory of radical evil and the notion of propensity to evil connected therewith. It has become a commonplace in Kant literature to attribute these Kantian doctrines to a sort of late concession to Christian theology, as can be observed in the tenor of the reception by e.g. Goethe, Schiller and Michalson (1990), among others. Against this view, I aim to show in this article that the most striking and controversial claims contained in Kant's doctrine of evil, namely the assumption of an intelligible deed, the universality of evil in humans as well as the inextirpability of the propensity to evil, can be accounted for by paying attention to the systematic relation between Kant's notions of *Anlage* and *Hang* and the key trait distinguishing them, namely practical objectivity. By doing so, it shall become clear that Kant's main contentious claims in *Religion I* can be sufficiently explained by systematic commitments of his own critical moral philosophy, thus rendering the references to Christian theology found in secondary literature superfluous.

Keywords: moral philosophy, moral evil, *Anlage* (predisposition), *Hang* (propensity), practical objectivity.

La relación sistemática entre las nociones de predisposición al bien y propensión al mal. Algunas observaciones sobre una asimetría en la teoría kantiana del mal moral

Resumen

El escrito *Religión dentro de los límites de la mera razón* ha sido un texto controvertido en la literatura sobre Kant desde su primera recepción. Una de las

¹ Eberhard Karls Universität Tübingen. Contact: cabezas.sebastian95@gmail.com. ORCID: <https://orcid.org/0009-0006-1998-5948>. I would like to express my deep gratitude towards the *Studienstiftung des deutschen Volkes* for the generous financial and ideal support I have been endowed with during my Ph.D. studies. This paper has arisen in the context of the scholarship provided by them.

principales razones de la recepción crítica yace en la teoría del mal radical y la noción de propensión al mal. Es común en la literatura secundaria atribuir estas doctrinas a una especie de concesión del viejo Kant a la teología cristiana, como se puede observar en la recepción de, por ejemplo, Goethe, Schiller y Michalson (1990). Contra esta posición quiero mostrar aquí que las tesis más provocadoras y controvertidas dentro de la doctrina kantiana del mal, esto es, la tesis de un acto inteligible, la universalidad del mal en los humanos junto con la inextirpabilidad de la propensión al mal, pueden explicarse prestando atención a la relación sistemática entre las nociones de *Anlage* y *Hang* y al rasgo central que las diferencia: objetividad práctica. Así, se dejará en evidencia que las principales tesis de *Religión I* se pueden explicar suficientemente a través de restricciones sistemáticas dadas por la propia filosofía moral crítica de Kant, convirtiendo así en superfluas las referencias a la teología cristiana que se encuentran en la literatura secundaria.

Palabras clave: filosofía moral, mal moral, *Anlage* (predisposición), *Hang* (propensión), objetividad práctica.

In 1793 Kant publishes *Religion Within the Boundaries of Mere Reason*, one of his major works after the three *Critiques* (hereinafter: *Religion*). However, it should be noted that in 1792 he had already published one part of it in the April issue of the *Berlinische Monatsschrift*, namely the so-called *Erstes Stück* entitled *Concerning the indwelling of the evil principle alongside the good or Of the radical evil in human nature (Von der Einwohnung des bösen Prinzips neben dem guten: oder über das radikale Böse in der menschlichen Natur)*.² The fact that this part of the book was published separately is already an indication that it can be considered as an integral whole whose understanding does not depend on the following book chapters. Here, Kant confronts us with a number of topics in moral

² In the separately published version of this text the title read only: *Über das radikale Böse in der menschlichen Natur*.

philosophy, with his notion of evil and the claim of a propensity to evil in human nature as some of the central elements in it. We can therefore agree with O. Höffe in his assessment that a comprehensive account of Kant's moral philosophy cannot be provided without taking into account this text, and even though the book's title may give rise to a different impression, *Religion* must not be seen only as a Kantian contribution to philosophy of religion, but also as central for his moral philosophy (see Höffe, 1995, p. 23f.).³

In the present paper, I want to focus on one of those topics central to Kant's moral philosophy, namely his notion of a propensity to evil and Kant's controversial claim of radical evil connected therewith. As is well known, Kant maintains that humans have a propensity to evil (*Hang zum Bösen*), which is itself said to be radically evil and considered to belong in some way to their nature. These claims are indeed striking and have prompted a rather critical reception of this Kantian work. More specifically, I shall draw attention to an equally striking asymmetry in Kant's moral theory, namely that between the already mentioned propensity to evil and the so-called predispositions to good (*Anlagen zum Guten*) which can be considered its counterpart in Kant's account. As the propensity to evil, this predisposition is also said to belong to human nature, but while the propensity to evil is attributed to an act of freedom performed by the agent himself (the so-called intelligible deed) and therefore seen as something *acquired* by him, the *Anlagen* are regarded as *original* (*ursprünglich*), i.e. as elements necessary for the very possibility of human nature. I shall argue that understanding the systematic relation as well as the asymmetry between *Hang* and *Anlage* can shed light on the source of some difficulties identified in Kant scholarship, such as the universality of evil and the inextirpability of the propensity to evil. By doing this, I show that the tensions in Kant's moral theory in *Religion* are *not*—as this text is sometimes read—due to a late concession made to Christianity, but can be explained *immanently*, i.e. with reference to philosophical reasons that lie in Kant's own (critical) practical philosophy.⁴

³ A good example of this approach can be found in Baxley's study *Kant's Theory of Virtue* (2010, esp. pp. 67–75). Here, she builds upon Kant's notion of a propensity to evil in order to develop a precise account of how Kant's concept of virtue as autocracy and defend it against common criticisms, see p. 68: "This doctrine of radical evil is the backdrop for understanding Kant's conception of virtue and the moral psychology associated with it".

In order to do so, I proceed as follows: I will start by dealing with Kant's notion of *Anlage zum Guten* as presented in the first part of *Religion* (I). In a second step I shall emphasise the large parallelism that can be drawn between the notion of *Anlage* and the idea of a propensity (*Hang*). This will lead us to the systematic question of why on Kant's view the propensity to evil cannot—in spite of those strong similarities—be considered a predisposition. I will argue that for this, one should assume that there is an objective side to evil (as there is in the case of moral goodness), which assumption, however, would go against Kant's concept of (practical) objectivity, which is closely linked to pure practical reason. Since Kant cannot accept such a claim, the *Hang zum Bösen* must reside in a free act performed by the agent himself and cannot, therefore, be thought of as a predisposition (II). Finally, I show how this two-sided relation of the notion of *Hang* to that of *Anlage* is at the basis of serious difficulties in Kant's text such as his doctrine of an intelligible deed, his claim of the universality of evil and the inextirpability of the propensity to evil (III). By doing this, it shall become clear that in order to account for the tensions present in Kant's treatment of moral evil we do not need to resort to the idea of late concessions made to Christian religion. Rather, they are indicative of *internal* systematic difficulties encountered by Kant when dealing with the problem of moral evil within his philosophical framework.

1. The counterpart of the propensity to evil: *Anlage zum Guten*

The reception of Kant's *Religion* seems to have been to a great extent under the influence of the first reactions to it. As mentioned in the introduction, the first part of the book was published separately in April 1792, while in 1793 the book appeared as we know it today with that first part being one of four parts of the work. In that very same year, the text was met with fierce criticism by the likes of Goethe and Schiller. In a famous letter to Herder from 7th June 1793, Goethe writes:

⁴ I thus agree with Allison's general view according to which Kant's account "constitutes a deepening of, rather than a break with, the moral psychology of the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*" (2002, p. 337). Nevertheless, my approach to the topic differs from his as will be evident from my elaborations in the following.

Kant required a long lifetime to purify his philosophical mantle 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 (1990, p. 270).⁵

One could also refer to K. Barth, who viewed Kant's doctrine of radical evil as "the last thing one would expect" (1969, p. 176) Kant to write on considering his previous works on moral philosophy. More recently, the difficulties identified in Kant's text have been seen as the expression of his efforts to render compatible two intellectual traditions that exerted a strong influence upon him: on the one hand, the Enlightenment reflected in his philosophy of autonomy and, on the other, the Christian tradition reflected in the Biblical allegories and notions such as the fall and the innateness of evil in human beings.⁶

As we can see from the passage by Goethe, the charge he makes against Kant concerns, at least in the first place, the notion of radical evil and he even claims to have identified the motivation for this notion, namely, a sort of concession made by Kant to Christian theology. Even though the tone adopted now cannot be compared to that of Goethe's or Schiller's polemical reactions and significant efforts are made in order to make sense of Kant's position, this notion continues to pose difficulties for Kant interpreters. In this sense, Allison (1990, p. 146) and Baxley (2010, p. 72) have called this doctrine "perplexing". In short, the radicality is a predicate Kant attributes to evil insofar as it is somehow rooted in human nature (see RGV, AA 06: 32). Therefore, it is to be understood in the etymological sense going back to the Latin noun *radix* and not as a particularly high degree of evil.⁷

What Goethe's reaction leaves unmentioned, however, is the fact that Kant is not only claiming that humans are radically evil, i.e. that they

⁵ *Goethe an Karoline und Gottfried Herder*, 7th June 1793 in *Johann Wolfgang Goethe. Gedenkausgabe der Werke, Briefe und Gespräche* (1949, vol. 19, p. 213). The English translation is taken from Allison (1990). Schiller, for his part, writes to his friend Christian G. Körner saying that Kant's claim of a propensity to evil is outrageous (*empörend*) for his feeling, see *Schiller an Koerner*, 28th February 1793 in *Schillers Werke. Nationalausgabe* (1992, vol. 26, p. 219) (at that point Kant's *Religion* was not published yet, but Schiller had access to it and announced its publication to his friend).

⁶ This is the general thesis defended by G. Michalson (1990).

⁷ This has already been noted by Allison (1990, p. 147; 1996, p. 170) as well as Baxley (2010, p. 68).

carry an evil which is rooted in their nature. At the same time, Kant attributes to human beings what he calls *Anlage zum Guten*, a predisposition to (moral) good which is also said to belong to human nature. In the following, I want to elaborate on the relation between these two notions in order to show how this helps to understand the source of major tensions in *Religion*.

To start with, Kant's conception of the moral *conditio humana* in *Religion* is not rightly described by any one-sided depiction such as the one suggested by Goethe's reaction. Instead, it is best characterized as an *ambivalent* view which ascribes to humans on the one hand natural predispositions to moral good, but, on the other, an innate propensity to evil.⁸ The notion of *Anlagen* is introduced by Kant in the first section of *Religion* I. His elaborations on this topic are rather sparse, but a closer look at it proves useful in order to understand in more detail his theory of moral evil. He defines *Anlagen* as "constituent parts [*Bestandstücke*] required for it [a being] as well as the forms of their combination that make for such a being" (AA 06: 28).⁹ Accordingly, a predisposition has a *constitutive* character with regard to the being in question, in this case, human beings.¹⁰ This is an important first insight insofar as it follows from this that the predispositions are not *acquired* by the beings possessing them. This is in line with what Kant says elsewhere about the notion of predisposition, namely, that they are to be understood as elements the species is equipped with (AA KU, 05: 420). Against this backdrop, we can understand why Kant, when introducing the concept of *Anlage* by dividing it into three classes, speaks of "elements of the determination of the human being" ["Elemente der Bestimmung des Menschen"] (RGV, AA 06: 26). By this, Kant seems to mean that the predispositions he will ascribe to human beings

⁸ This view is also suggested by Susan Shell's entry on "Anlage" (see *Kant-Lexikon*, 2021, p. 96f.).

⁹ The English translations from *Religion* are taken from A. Wood and G. di Giovanni's volume *Religion and Rational Theology* (1996). The translations from *Groundwork of the Metaphysics of Morals* stem from M. Gregor and J. Timmermann's edition (2011).

¹⁰ It should be noted that this definition concerns the concept of *Anlage* in general, not merely the more specific one of *Anlagen zum Guten*. Given this determination, however, it seems difficult to understand how we can think of *contingent* predispositions (*zufällige Anlagen*) as opposed to *original* predispositions (*ursprüngliche Anlagen*). Since the *Anlagen zum Guten* are indeed original, the possibility of contingent predispositions does not concern the topic of my elaborations in this work. However, I would like to draw attention to the fact that there seems to be a difficulty in understanding predisposition in general as a constitutive element of the being in question while allowing for the concept of contingent predispositions, i.e. predispositions without which the being in question can still be what it is (RGV, AA 06: 28).

are contained in the very concept of human. To put it in a more technical terminology: these predispositions are elements of the intension of the concept ‘human’. I will come back in section 2 to the difference *constitutive* vs. *acquired*. At this juncture, we must specify the conceptual determination of predisposition we have just reached.

They are contained in the concept of ‘human’ as a being belonging, on the one hand, to the sensible realm, and, on the other, as a being equipped with the faculty of reason and additionally with *pure* practical reason. The concept of human thus understood gives rise to the division of *Anlage zum Guten* proposed by Kant in *Religion*: i) the predisposition to the animality of the human being, ii) the predisposition to the humanity (*Menschheit*) and finally ii) the predisposition to his personality (*Persönlichkeit*). As it is evident from the very title of this section and the formulations used by Kant, the concept of *Anlage* entails the concept of purposefulness (*Zweckmäßigkeit*): they are predispositions *to* the (moral) good (*zum Guten*) and the very *principium divisionis* underlying the above tripartition is the respective purpose (“in Beziehung auf ihren Zweck” [RGV, AA 06: 26]). We can bring together both conceptual determinations of *Anlage* (the constitutive character and the purposiveness) by saying that the *Anlagen zum Guten* are a sort of ‘tool’ our species is equipped with in order to realise a certain end given by nature. In the case of the *Anlagen zum Guten* this end is, as indicated by the name, morality. It is therefore appropriate to call them *natural* dispositions,¹¹ and Kant himself confirms this point by speaking—in the case of the predisposition to personality—of *Naturanlage* (RGV, AA 06: 27). But also in the case of the predisposition to humanity he attributes its respective end to nature: “for nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture” (RGV, AA 06: 27).

Moreover, insofar as these predispositions are contained, as we saw, in the very concept of ‘human’, they are necessary for a being to be a human being. To make this point clearer, it is useful to specify the above-mentioned predispositions Kant ascribes to human beings. By the first one, the predisposition to the animality of the human being, Kant means the predisposition for self-preservation, for the propagation of the species and for the preservation of the offspring as well as the social drive, i.e. for living

¹¹ Also Ch. Horn (2011, p. 54) speaks in his treatment of this concept of ‘*naturale Disposition*’.

in community with other human beings. The second one, viz., the predisposition to humanity, refers to the physical self-love which entails comparison—and thereby reason. By virtue of this predisposition, we strive for being recognized as equals in the opinion of others. Finally, the third and last predisposition, which can be called moral in an eminent sense, consists in the “susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” [“Empfänglichkeit der Achtung für das moralische Gesetz, als einer für sich hinreichenden Triebfeder der Willkür”] (RGV, AA 06: 27), that is, the capacity by virtue of which we can be determined to action by the sole respect for the moral law.

It should be noted that Kant is not providing any derivation of this tripartition of the *Anlage zum Guten* in humans. This is remarkable given Kant’s own theoretical requirements for systematicity. In fact, he takes the deduction, i.e. deriving any division from a principle, to be equivalent with the proof of the completeness (*Vollständigkeit*) of that division (see MS, AA 06: 218).¹² In view of this requirement for systematicity one could claim at this point that Kant’s tripartition of *Anlagen zum Guten* does not guarantee that this division is exhaustive. For our present purposes, we can put aside this internal difficulty of Kant’s account of *Anlagen zum Guten* and focus on the third trait just identified, that of necessity. Why does Kant view these predispositions as being necessary for a being to be a human being?

This point can be best illustrated by focusing on the third propensity, the predisposition to personality. As stated above, Kant conceives of this predisposition as the capacity within us to give the moral law a sufficient motivational force in determining our will. As we know from Kant’s previous critical writings on moral philosophy, an agent who *actualises* this capacity, is considered morally good: the determination of his will takes place solely on the basis of respect for the moral law. Kant expresses this point in the terminology of *Religion* by saying that “a power of choice [*Willkür*] so constituted is a good character” (RGV, AA 06: 27). Since moral assessment strictly speaking rests on the moral quality of the agent’s

¹² See also the *Critique of Pure Reason*, where within the context of his account of the categories his main criticism against Aristotle is directed precisely at the fact that he did not derive the categories from a principle and can therefore not guarantee the completeness of the table of categories (KrV, A80f./B106f.). See also GMS, AA 04: 387, where he likewise requires the establishment of a principle in order to guarantee the completeness and continuity (*Stetigkeit*) of the division of philosophy into logic, physics and ethics. By failing to deduct the tripartition of the *Anlagen zum Guten* in human nature Kant is, then, lagging behind his own standards.

character (RGV, AA 06: 23, 47) we can simply say that an agent who has actualised the capacity in question is a morally good agent. The original question now can therefore be restated as follows: Why does a capacity to moral good belong to the very concept of ‘human’? The answer to this question rests upon the faculty of pure practical reason. As we saw above, the predisposition to personality depends on the faculty of pure practical reason.¹³ According to Kant’s critical philosophy, a morally good will is a will whose determining principles are adopted solely out of respect for the moral law, i.e. with all other incentives playing no motivational role in the determination of the will. Given that for Kant respect for the law is a feeling that arises from pure (practical) reason (GMS, AA 04: 401),¹⁴ the determining principle that moves a morally good will to action is in the last analysis nothing other than pure practical reason. To be sure, the mere fact of possessing the faculty of pure practical reason does not imply that the will is *indeed* determined by it, but it does entail that it *can* be determined by it. To put it another way: the concept of human being, insofar as it is represented as a being endowed with pure practical reason, implies the possibility of his actions being determined by pure practical reason. As we have just seen, a will determined by pure practical reason is precisely what a morally good will consists in. Therefore, we can conclude that—from a Kantian point of view—the *possibility of moral good* is implied in the concept of pure practical reason. If this is the case, then, the possibility of moral good is contained in the concept of human being as a being equipped with the faculty of pure practical reason, which is equivalent to claiming that the predisposition to personality is contained in the concept of human being, i.e. is necessary for a being to be a human being. It is in this sense that Kant takes the *Anlagen zum Guten* to be necessary for human beings. This aspect of necessity is expressed by him with the term *ursprünglich* (original). As he explains: all these three predispositions are “*original*, for they belong to the possibility of human nature” (RGV, AA 06: 28). From

¹³ See further RGV, AA 06: 28: “If we consider the three predispositions just named according to the conditions of their possibility, we find that the *first* does not have reason at its root at all; that the *second* is rooted in a reason which is indeed practical, but only as subservient to other incentives; and that the *third* alone is rooted in reason practical of itself, i.e. in reason legislating unconditionally”.

¹⁴ Respect is a feeling “self-wrought by a rational concept and therefore specifically different from all feelings of the former kind [those received by influence, S. C.], which come down to inclination or fear” (GMS, AA 04: 401). Even though at this place Kant is only speaking of “rational concept” (*Vernunftbegriff*), the origin of such a *Vernunftbegriff* is precisely *pure* practical reason, given that *Vernunftbegriffe* are those representations of reason which do not borrow anything from experience (see KrV, A310–314/B366–370).

this feature, viz., the necessity of these *Anlagen* for human nature, Kant derives, as it were, a corollary, namely their inextirpability or ineradicability (RGV, AA 06: 28). Indeed, since these predispositions belong to the very nature of human beings, they cannot be removed from a human agent without depriving him *ipso facto* of his human nature.

The conceptual analysis of the notion of predisposition to the good I have just advanced has brought to light the following features of it: the *Anlagen zum Guten* are i) *constitutive* elements of human nature which are thought of as a sort of ii) *means to realising certain ends* enjoined on us by our very nature (in the case of *Anlagen zum Guten* the end is morality). By virtue of this constitutive character, iii) these predispositions are considered to be *contained in the very concept* of ‘human’, for which reason iv) they *cannot be eradicated*. Let us now turn to the competing concept of propensity to evil introduced by Kant in the following sections of *Religion*.

2. *Hang zum Bösen* and its conceptual parallelism to the notion of *Anlage*

As I pointed out in the previous section, Kant does not provide any deduction of his tripartition of the *Anlagen zum Guten* in humans. For this reason, the question whether he takes his list of three predispositions to be exhaustive remains open from a systematic point of view. Nevertheless, if I interpret Kant correctly, he does seem to assume that these three *Anlagen* he has just presented are *all* predispositions *to good* found in human beings, and even more: they seem to be *all morally relevant* predispositions in human nature, that is, either to good or evil. In my view, this is suggested by Kant in the concluding remark of this section of *Religion*: “It should be noted, finally, that there is no question here of other predispositions except those that relate immediately to the faculty of desire and the exercise of the power of choice” [“die sich unmittelbar auf das Begehrungsvermögen und den Gebrauch der Willkür beziehen”] (RGV, AA 06: 28). At this juncture, while restricting his account of predispositions to those connected with the faculty of desire (and thus excluding, for instance, biological predispositions), he is explicitly saying that the treatment given refers to the predispositions that are relevant *for the use of freedom*. Accordingly, Kant’s account of predispositions seems not to be limited to a particular use of freedom (to moral good), but rather to include *any* use of freedom, therefore also to evil. For this reason, based on this preliminary observation, one

might already claim that on Kant's proposal there are no *predispositions* to evil. And indeed, the concept he will introduce now is that of a *propensity* (*Hang, propensio*), more specifically a propensity to moral evil. At a later place in *Religion*, he will explicitly distinguish it from a predisposition by saying that this propensity cannot be considered a "natural predisposition" (*Naturanlage*) (RGV, AA 06: 32). The question now is thus how Kant conceives of this propensity as compared to the previous concept of predisposition.

After the first section of *Religion* I was concerned with the predispositions to good, the second one addresses the most commented topic of *Religion*, namely that of a propensity to evil.¹⁵ Kant introduces the notion of *propensio* in general (*überhaupt*) in the following manner: "By propensity (*propensio*) I understand the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*), insofar as this possibility is contingent for humanity in general" (RGV, AA 06: 28). And at this point, Kant expands on this notion: "*Propensity* is actually only the *predisposition* to desire an enjoyment which, when the subject has experienced it, arouses *inclination* to it" (RGV, AA 06: 28). We can see from the second passage that Kant seems to place the notion of *propensio* very close to that of *Anlage* by calling it also a predisposition (*Prädisposition*). As we know, *praedispositio* is usually the Latin equivalent of the German *Anlage*, which is already reflected in the English translation of the latter as *predisposition*. This terminological remark is a first indication of a large parallelism that we can observe between the notions of *Anlage* and *Hang*. In the following, I want to elaborate on this point by building on the conceptual determination advanced in the previous section.

As we have seen, we can capture the core of the concept of *Anlage zum Guten* through the following features: constitutive character (belonging to the very nature of humans), purposefulness, necessity and inextirpability. Surprisingly, after a first reading of the relevant passages, we can find most of these features in Kant's concept of *Hang zum Bösen*.

To begin with, in one of the most famous and controversial places of *Religion*, Kant claims, quoting Horace's dictum *vitiis nemo sine nascitur*, that the human being is evil *by nature* (RGV, AA 06: 32). On Kant's

¹⁵ See, for example, Allison (2002) and more recently Forschner (2011) as well as the above-mentioned Horn (2011).

account, this natural character refers to the generality (*Allgemeinheit*) of evil among humans: “‘He is evil *by nature*’ simply means that being evil applies to him [the human being, S. C.] considered in his species” (RGV, AA 06: 32).¹⁶ It is this propensity to evil which Kant calls a radical evil insofar as it is somehow rooted in human nature (RGV, AA 06: 32), which claim has been the main target of the fierce reception by Goethe and Schiller, among others. Accordingly, Kant goes as far as to speak of a *natural* propensity to evil (“natürlicher Hang zum Bösen” [RGV, AA 06: 32]).¹⁷ Hence, we see at this point that both the *Anlagen zum Guten* and the *Hang zum Bösen* are considered to belong in some way to human nature and are in this regard general, i.e. determinations applying not merely to some individuals, but to the whole species.¹⁸

The next parallel between both notions can be found when looking at the modality Kant ascribes to the propensity to evil. Indeed, he claims that “we may presuppose evil as subjectively necessary in every human being, even the best” (RGV, AA 06: 32). Hence, the propensity to evil is in some sense also necessary for human beings. I will come back later in the text to the specification of ‘subjective’ in this context. For now, however, we should keep track of the apparent parallelism in the conceptual construction found in the notions of predisposition and propensity. In addition to both belonging in some way to human nature and being necessary, the probably most striking parallel between the *Anlagen zum Guten* and *Hang zum Bösen* is the one concerning the inextirpable character of it. Indeed, not only are the *Anlagen zum Guten* ineradicable (see first section), but the propensity to

¹⁶ Kant had anticipated this claim earlier in the text (see RGV, AA 06: 25).

¹⁷ See also two occurrences of “natural propensity” in RGV, AA 06: 29.

¹⁸ Given that according to the doctrine of the *Critique of Pure Reason* strict generality can only stem from *a priori* knowledge and this Kantian claim cannot be considered to be *a priori*, Horn (2011, p. 66) proposes a distinction (known from the *Groundwork*) between empirical and strict generality (*generalitas* vs. *universalitas*) and takes the generality of radical evil to be empirical. By contrast, Allison (1990, pp. 154–157) argues for the apriority of Kant’s claim. Kant seems very ambiguous on this issue. On the one hand, he suggests that he is in a position to give a *formal* proof (*förmlicher Beweis*) of the rootedness of the propensity to evil in human nature, but takes this to be unnecessary (RGV, AA 06: 33). He even suggests that he has provided an “appropriate proof” (*eigentlicher Beweis*) in section 2, as distinct from mere confirmation through experience (RGV, AA 06: 39). On the other hand, earlier in the text he held that the proof of such a generality is to be delivered by *anthropological* research (RGV, AA 06: 25), thus suggesting an *empirical* path (for Kant uses the term *practical anthropology* to refer to the empirical part of Ethics [see GMS, AA 04: 388]). This impression is reinforced by a remark in the very same passage where Kant is defending his claim of human beings being evil by nature when he says that “according to the cognition we have of the human being *through experience*, he cannot be judged otherwise [than evil, S. C.]” (RGV, AA 06: 32; my emphasis).

evil, too, “cannot be eradicated [*ausgerottet*] (for the supreme maxim for that would have to be the maxim of the good, whereas in this propensity the maxim has been assumed to be evil” (RGV, AA 06: 31f.), which is why it is said to be inextirpable (*unvertilgbar* [RGV, AA 06: 51]). It should be observed that Kant is using in the last case the same German term (*vertilgen*) to describe the quality of the propensity to evil as he did when dealing with the *Anlagen zum Guten* (RGV, AA 06: 28).¹⁹

At this place, it should be noted that this systematic correspondence between the notions of *Anlage* and *Hang* extends beyond the core features I have just laid out. This becomes evident when we look closer at Kant’s elaborations on the predispositions to good and the definition of propensity quoted above. As we saw above, Kant has determined the concept of propensity as distinct from an inclination (*Neigung*) and adds the following example to illustrate it:

Thus all savages have a propensity for intoxicants; for although many of them have no acquaintance at all with intoxication, and hence absolutely no desire for the things that produce it, let them try these things but once, and there is aroused in them an almost inextinguishable desire for them (RGV, AA 06: 28).

As I see it, this passage contains the core of the distinction Kant is trying to draw between *Hang* and *Neigung*. In this context, Kant uses the term desire (*Begierde*) and inclination interchangeably. As we can learn from this passage, inclination or desire presupposes acquaintance (*Bekanntschaft*) with the object in question, while a propensity does not. The whole context of Kant’s treatment here amounts to a sort of stepladder of the appetitive faculty consisting of four stages (bottom up): propensity, instinct, inclination and passion. As I have just outlined, the central difference between these levels of the appetitive faculty rests on the agent’s epistemic relation to the object in question. While in the case of a propensity the subject does not have any acquaintance with the object, instinct and inclination do presuppose a certain acquaintance with the object of the appetite (the former, however, not having a *concept* of it). Finally, the

¹⁹ At this point, I cannot elaborate on the implications and difficulties of this Kantian doctrine. I will take up this controversial claim again in the following section and propose a systematic justification for it.

highest stage is passion (*Leidenschaft*), which is itself a kind of inclination²⁰ and therefore also presupposes acquaintance with the object, but differing from *Neigung* in that in the former case the agent is no longer master over himself (RGV, AA 06: 29). What is relevant in this account for our present purpose is the fact that the propensity is thought of as the origin of a *Neigung* or *Begierde*. In the passage just quoted, Kant seems to point out that the experience of the object in question is a sufficient condition for a transition from *Hang* to *Neigung*. This quality, i.e. being the origin of an inclination, is also present in Kant's account of *Anlagen zum Guten*. If we look at Kant's description of the second predisposition, viz. that to humanity, we will see that this is thought to be the origin of an inclination:

The predispositions to humanity can be brought under the general title of a self-love which is physical and yet *involves comparison* [...]. Out of this self-love originates the inclination *to gain worth in the opinion of others*, originally, of course, merely *equal worth* (RGV, AA 06: 27).

As we see from this passage, the predisposition to humanity is on Kant's view the ground of an inclination, namely, the inclination to be recognised as equal by others. This is reinforced by a subsequent remark according to which an unjust desire (*Begierde*) can arise from this (remember that in the section on the notion of *propensio* Kant uses the term *Begierde* as synonymous with *Neigung*). In light of these observations, we can argue that Kant seems to have on the side of the *Anlagen zum Guten* a similar idea of gradual stages of the appetitive faculty. Both an *Anlage* and a *Hang* can give rise—under certain conditions—to a *Neigung*.²¹

²⁰ See RGV AA 06: 29, where Kant calls it explicitly *Neigung*.

²¹ From these elaborations it follows that *propensity* and *inclination/desire* must be distinguished from one another. As we saw above, the reason for this distinction lies in the agent's epistemic relation to the object in question (the propensity is referred to an object the agent has not experienced yet, while an inclination presupposes acquaintance with its object). In the text of *Religion I*, this distinction is clearly made by Kant. For this reason, it is not correct to speak of the propensity to evil as an inclination, let alone attribute this to Kant himself, as is suggested by Horn (2011, p. 65). Nevertheless, it should be noted that at some points Kant's use of his own terminology is, as in some other cases in both his theoretical and practical philosophy, somewhat inaccurate. For instance, in the *Metaphysics of Morals* (1797) Kant will translate the Latin term *propensio*, which is the equivalent of *Hang* in *Religion*, as *Neigung* (MS AA, 06: 213). However, this is rather a *terminological* inaccuracy than a *conceptual* modification.

Finally, at this stage we have to observe how this large correspondence I have developed is reinforced by the systematic function that the predisposition and the propensity are meant to fulfil. I take this to be one of the key points to coming closer to an understanding of Kant's enigmatic claims about moral evil in *Religion*. The focus here lies again in the third predisposition, the predisposition to personality. When dealing with the question of why Kant assumes that the predispositions to good belong to human nature I illustrated this by elaborating on how this works in the case of the predisposition to personality. One of the results of this was that this predisposition is conceived of by Kant as the *capacity*, that is, the *possibility* of moral goodness. Indeed, the *Anlagen* are not themselves morally good, since being *morally* good or evil is a quality whose *actual* possession depends on it being ascribable to the agent's *free* action.²² The predispositions are, however, as we have seen, *Naturanlagen*, and therefore not attributable to the agent's freedom. Thus, these predispositions are not themselves good, but—as indicated by the name itself—to the good. As Kant puts it: “All these predispositions in the human being [the three mentioned above, S. C.] are not only (negatively) *good* (they do not resist the moral law) but they are also predispositions *to the good*” (RGV, AA 06:28). In this sense, the predisposition to personality—to come back to our previous example—accounts for the mere *possibility* of moral goodness, which seems to be explicitly stated by Kant himself (RGV, AA 06: 27f.). Similarly, the notion of *propensity* explicitly includes the same modal aspect: “By *propensity* (*propensio*) I understand the subjective ground of the *possibility* of an inclination” (RGV, AA 06: 28; last emphasis mine). In the specific case of a propensity to moral evil, this means that such a propensity accounts according to this definition for the *possibility* of evil, which Kant formulates in the following terms:

Here, however, we are only talking of a propensity to genuine evil, i.e. moral evil, which, since it is only possible as a determination of a free power of choice [*Bestimmung der freien Willkür*] and this power for its part can be judged good or evil only on the basis of its maxims, must reside in the subjective ground of the possibility of the deviation of the maxims from the moral law (RGV, AA 06: 29).

²² “The human being must make or have made himself into whatever he is or should become in a moral sense, good or evil. These two [characters] must be an effect of his free power of choice, for otherwise they could not be imputed to him and, consequently, he could be neither *morally* good nor evil” (RGV, AA 06: 44).

Kant is explicitly speaking of ground of the *possibility* of moral evil, which is said to be a quality of maxims. The systematic functions Kant attributes to the predispositions to good and the propensity to evil seem, then, at least at this stage, to coincide with one another. In view of these observations, one might be tempted to express the general structure of Kant's account in *Religion I* by saying that the *Anlage zum Guten* and the *Hang zum Bösen* are meant to perform an equivalent systematic function differing only as to the side of the moral spectrum they occupy respectively: what the predisposition to good is in the case of moral goodness corresponds to what the propensity to evil is in the case of moral evil.

On the basis of the above elaborations, it seems plausible to assume that Kant introduced the notion of a propensity to evil in a clear correspondence to that of predisposition to good, particularly to the predisposition to personality *qua* possibility of moral goodness. Arguably, the pursuit of this systematic correspondence is at the basis of the core features of the notion of propensity to evil as set out in this section. Yet the establishment of this large parallelism between these two notions leaves open an obvious question. Despite the fact that both belong in some way to human nature and are as such a *general* determination (not merely of some individuals), that both are said to carry a certain necessity in the case of human beings, that they are both inextirpable and can give rise to an inclination, and finally, in spite of both being used to account for the possibility of the respective moral determination (either good or evil), it is still true that the propensity to evil is *not a predisposition* to evil. There must be a reason why Kant views the *Hang zum Bösen* as a propensity rather than as a predisposition. The question is, then: why does Kant still think—in spite of this evident systematic parallelism he has established—that the propensity to evil cannot be considered a *predisposition*? This problem is all the more pressing given Kant's statement in the context of the predisposition to personality:

a power of choice so constituted [one that has incorporated the moral feeling into its maxim, S. C.] is a good character, and this character, as in general every character of the free power of choice, is something that can only be acquired; *yet, for its possibility there must be present in our nature a predisposition* (RGV, AA 06: 27; my emphasis).

Here, Kant claims that while the *actuality* of the good character depends on our own freedom, its *possibility* must be grounded in our nature, more specifically in a predisposition. As we know from the first section, speaking of a morally good character amounts to saying that the agent in question is morally good given that moral assessment is dependent on the character's moral quality. Therefore, Kant's claim at this point is that for the possibility of moral goodness there must be a predisposition to it in our nature. In view of this, one could expect the same on the other side of the moral spectrum: the *possibility* of moral evil should be accounted for by a predisposition.²³ We have just seen in this section that the propensity to evil is precisely supposed to account for the possibility of evil (RGV, AA 06: 29). For this reason, it would seem natural to assume that the *Hang zum Bösen* should be understood as a *predisposition*, but this is exactly not how Kant sees it. In this sense, we find here a striking asymmetry in Kant's moral theory in *Religion I* which should be accounted for.

When addressing this question, Kant's use of the terminology might be confusing. Indeed, it should be noted that in the Preface to the first edition of *Religion*, he speaks of human nature as containing 'both good and evil predispositions' ("die menschliche, theils mit guten theils bösen Anlagen behaftete Natur" [RVG, AA 06: 11]), by which he is clearly referring to the doctrine of *Anlagen zum Guten* and *Hang zum Bösen*. At this point, then, Kant uses the term *Anlage* to refer to both the predispositions and the propensity. Nevertheless, this simply seems to be yet another example for the inaccuracy Kant exhibits sometimes with his own terminology. Indeed, as noted at the beginning of this section, in the main text of *Religion* Kant clearly wants to distinguish the notion of propensity from that of predisposition (RGV, AA 06: 32).

As I see it, the key to dealing with the issue just proposed lies in the concept of (practical) objectivity, which is at the basis of the predisposition to good, but cannot be found in the notion of a propensity to evil. In the following, I will disclose this asymmetry in Kant's *Religion* and show how

²³ Surprisingly, this is Allison's view (2002, p. 339): "Significantly, the predisposition to personality is just the capacity to be motivated by respect for the law; and it is in virtue of this capacity that we are moral agents, *capable of either good or evil*" (my emphasis). As it shall become clear from the following elaborations, I cannot agree with Allison on this point. Grounding the possibility of moral evil in the predisposition to personality seems to disregard Kant's remark according to which onto this predisposition "nothing evil can be grafted" (RGV, AA 06: 27).

it is this point which represents the source of the difficulties present in the moral theory of *Religion*.

The crucial point for an adequate understanding of Kant's positions concerning moral evil in *Religion* lies in the determination of the *propensio* as the *subjective* ground of possibility of an inclination. In fact, when introducing the notion of propensity, Kant is cautious in providing this additional determination ("subjectiver Grund der Möglichkeit einer Neigung" [RGV, AA 06: 28]). And when applying the general notion of propensity to the case of moral evil, he reiterates this addition ("subjectiver Grund der Möglichkeit der Abweichung der Maximen vom moralischen Gesetze" [RGV, AA 06: 29]). According to this, the propensity to evil is not said to ground the possibility of evil in general, but only *subjectively*.²⁴ It must be noted that the term *subjective* also plays a role in determining the necessity ascribed to this propensity. Kant's claim here is, again, *not*—as it was in the case of the predisposition to good—that the propensity is necessary for a being to be a human being, but rather that it is "subjectively necessary" in every human being (RGV, AA 06: 32). To my knowledge, Kant scholarship has not given this determination the required attention when providing accounts of Kant's theses on moral evil. What does Kant mean in this context by 'subjective' and what would be the systematic alternative?

According to the previous elaborations, one meaning of *subjective* can be excluded with certainty. By *subjective* Kant cannot understand what we usually mean when speaking of *subjective*, namely, something belonging to the individual and thus not being able to claim general validity. We can call this the *individualistic* meaning. For we have seen that on Kant's view the propensity to evil is not the trait of an individual, or even several individuals of the human species. Rather, as we have seen, he takes it to belong to the species.²⁵ 'Subjective' in this meaning is thus compatible with a form of generality. In his writings on practical philosophy, Kant uses the term *subjective* to designate the opposite of *objective* (see e.g. GMS, AA 04: 400, 421; KpV, AA 05: 20).²⁶ As is evident from these references, practical

²⁴ "And yet by the concept of a propensity is understood a *subjective* determining ground of the power of choice [*subjektiver Bestimmungsgrund der Willkür*]" (RGV, AA 06: 31; my emphasis).

²⁵ As has already been described, Kant's thesis is that man is evil by nature, and this refers precisely to that generality, see again RGV, AA 06: 32: "'He is evil *by nature*' simply means that being evil applies to him considered in his species".

²⁶ Cf. also the entry on '*subjektiv*' by D. Sturma, *Kant-Lexikon* (2021, pp. 2203–2205).

objectivity is a trait of *practical laws*, that is, of practical principles valid for every rational being as such. More specifically, objectivity in its practical sense can be described as the feature of a particular kind of necessity, namely *unconditioned* necessity: “For only the *law* carries with it the concept of an *unconditional* and indeed objective and hence universally valid *necessity*” (GMS, AA 04: 416). This is crucial to understanding the source of practical objectivity. In fact, if something claims unconditionality—as moral requirements do according to Kant—it can never come from experience, but must stem from pure reason. Kant formulates this point already in the preface of the *Groundwork*:

Everyone must admit that a law, if it is to hold morally, i.e. as the ground of an obligation, must carry with it absolute necessity [...]; hence that the ground of the obligation here must not be sought in the nature of the human being, or in the circumstances of the world in which he is placed, but *a priori* solely in concepts of pure reason (GMS, AA 04: 389).

One can word this point also in a different way: if the validity of practical laws is said to hold for every rational being *as such*, then these laws can be grounded only on reason itself, since otherwise the claim of universal validity could not be upheld. Practical objectivity is, then, the quality of practical laws by virtue of which they hold universally for every rational being as such and therefore of practical principles taken from pure (practical) reason alone.²⁷ Accordingly, the source of (practical) objectivity can be nothing other than pure practical reason.²⁸ Where does, however, the relevance of this notion lie with regard to Kant’s theory of evil in *Religion*?

After these considerations, we can see that there is an objective side to moral goodness, namely its being grounded in pure practical reason. As has been stated earlier, an agent is morally good according to Kant if and only if his will is sufficiently determined by pure practical reason, that is, through the command of the moral law. Moral goodness thus consists in the quality of a will and its maxims insofar as these make pure practical reason their sufficient ground of determination. This shows that pure practical

²⁷ See also GMS, AA 04: 413, where “objectively” is paraphrased as follows: “from grounds that are valid for every rational being, as such”.

²⁸ It is for this reason that, for example, H. Klemme (2006, p. 121) can use both terms interchangeably when he speaks of “objective, i.e. rational ground of willing” (my translation).

reason is presupposed in the very determination of the concept of moral good. Indeed, Kant reinforces this point by claiming in the second *Critique* that “moral good and evil must always be judged through reason and thus through concepts that can be universally communicated” [“und verlangt, daß Gutes und Böses jederzeit durch Vernunft [...] beurteilt werde”] (KpV, AA 05: 58).

In the text of *Religion*, Kant is careful in speaking of the propensity to evil as *subjective* ground of possibility of evil. In the case of the predisposition to good, there is no such addition. Admittedly, here he simply uses the predisposition to personality to account for the possibility of moral goodness, without specifying whether he takes it to account for the *subjective* or the *objective* possibility of moral goodness. Yet it is clear that this predisposition is to be understood as the *objective* ground of possibility of the moral good. This becomes manifest if we pay attention to the kind of necessity Kant is attributing to the *Anlage* and the *Hang*, respectively. I have already pointed out that Kant takes the propensity to evil in humans to be *subjectively* necessary (“*subjektiv notwendig*” [RGV, AA 06: 32]). This necessity must be strictly distinguished from the necessity ascribed to the predispositions to good. According to Kant, these predispositions are necessary for the very possibility of the being in question, in this case, of human beings.²⁹ This point has been described above as a containment relationship (see section 1): the predispositions to good are contained in the very concept of human being. To stick to the example of the propensity to personality, which is to be seen as the equivalent of the propensity to evil on the other side of the moral spectrum: the possibility of moral goodness is implied in the concept of human as a being endowed with the faculty of pure practical reason. On the contrary, the necessity Kant assigns to the propensity to evil has another status. A containment relation, as has been identified here for the case of the predisposition to good, is explicitly rejected by Kant for the *Hang zum Bösen*:

“He is evil *by nature*” simply means that being evil applies to him [the human being, S. C.] considered in his species; not that this quality may be inferred from the concept of his species ([i.e.] from the concept of a human being in general, for then the quality would be necessary) (RGV, AA 06: 32).

²⁹ As we saw, Kant takes these predispositions to be *original* (*ursprünglich*), which means nothing but that “they belong with necessity to the possibility of this being” (RGV, AA 06: 28).

The predispositions to good are grounded on an *objective* necessity insofar as they result from the very nature of human beings. To put it in other words: they are *not contingent* (*zufällig*). This means that it is not possible for a being to be a human being without these predispositions (RGV, AA 06: 28). It is in this sense that they can be called *constitutive* of human beings. By contrast, the propensity to evil is said to be *contingent* (RGV, AA 06: 32).³⁰ Kant expresses this crucial difference when speaking of the predispositions to good as *conatural* (*anerschaffen* [RGV, AA 06: 51]), whereas a *propensio* is to be thought of as brought by the agent upon himself, i.e. self-acquired (RGV, AA 06: 29). Therefore, in order for the *Hang zum Bösen* to be a *predisposition*, one must assume that there is an objective side to moral evil—as there is in the case of moral goodness. In the latter, this objective side is given through the very condition for the possibility of the predisposition to good, namely, pure practical reason.³¹ There cannot be, however, such an *objective* side concerning moral evil, for, as we have seen, practical objectivity is closely linked to pure practical reason, but evil consists precisely in the subordination of the incentives given by reason (RGV, AA 06: 36). In addition, since on Kant's view moral evil must always refer, *qua* moral quality, to the agent's use of freedom and therefore to the adoption of maxims and maxim-making is a process which involves reason, the alternative to this account—that is, viewing the *Hang zum Bösen* as a natural *predisposition*—would require, analogously to the predisposition to personality, grounding the possibility of evil in practical reason itself, a claim Kant would like to avoid.³²

As I see it, this is the reason why the *Hang zum Bösen*—despite the large systematic parallelism we have disclosed between the notions of

³⁰ To be sure: at this juncture there seems to be a tension in Kant's account. For as we have seen, the propensity to evil is said to account for the *possibility* of evil. Therefore, by saying that it must be considered as *contingent* for humanity, Kant is claiming that it is possible to be a human being without having this propensity (for that is the definition of contingent, cf. RGV, AA 06: 28). This claim seems to amount to holding that it is possible for a human being to not even have the very *possibility* of moral evil. If this is the case, however, it becomes difficult to see how Kant's distinction between holiness and morality could still remain in place.

³¹ For the claim that pure practical reason grounds the possibility of this predisposition cf. section 1 (RGV, AA 06: 28).

³² If I understand Allison correctly, his construal of Kant's doctrine implies this thesis. He claims: "But since unlike the purely sensuous side of our nature it [the predisposition to humanity] is connected with a use of reason and, therefore, freedom, it is subject to a misuse or perversion that does generate evil" (Allison, 2002, p. 339). Insofar as he is placing the possibility of evil in the (misuse of the) predisposition to humanity, one should assume, according to my interpretation, that on Kant's view moral evil has an objective side, a claim I have argued against here.

Anlage and *Hang*—cannot be a predisposition. Claiming this would amount to grounding the possibility of moral evil in pure practical reason, a thesis which would go against central commitments of Kant’s moral philosophy. The possibility of moral evil—unlike the predispositions to good—lacks any grounding in original faculties of the human being and can, accordingly, only be traced back to his own use of freedom—hence not *ursprünglich*, but *zufällig*.³³ The propensity to evil, then, is conceived of by Kant as a moral quality acquired by the agent himself, i.e. through a free act, which for that reason cannot be a *predisposition*.

3. The source of major systematic difficulties in *Religion*’s moral theory

We have now established both the conceptual parallelism and the distinguishing feature between the notions of *Anlage* and *Hang* and systematically reconstructed why the propensity to evil cannot be thought of as a predisposition, namely, because of the concept of practical objectivity presupposed in the very notion of *Anlage*. The challenge now is to show why disclosing this conceptual relation is relevant for providing a *philosophical* (rather than biographical) account of the tensions present in *Religion*, i.e. without the need to resort to external, rather polemical explanations.

Relying on external grounds for accounting for Kant’s views in *Religion* is, as I pointed out earlier in the text, not only the main path taken by the first reception, but it is still present in more recent literature. In general lines, this is how, for instance, the above-mentioned Michalson (1990) and Horn (2011) frame the issue. See e.g. the following formulation:

The theory of radical evil is finally symptomatic of the fact that Kant has not totally thrown off the habits of mind produced by Christian culture, yet these

³³ It is thus wrong, or at least terminologically inaccurate and confusing, to call the propensity to evil an “original tendency” (*ursprüngliche Tendenz*) as Horn does (2011, p. 44). As a side note: this is also the reason why Kant’s expression *natural* (*natürlich*) or *by nature* (*von Natur*) must be understood differently when applied to the propensity to evil (cf. e.g. RGV, AA 06: 29, 32) than in the case of the *Anlagen*. In the first case it is referred to the fact that the origin of the *Hang zum Bösen* does not lie in time (see RGV, AA 06: 25), while in the latter it refers to *objective* determinations without which the being in question would not be that being.

habits of mind are in many ways antithetical to his deepest philosophical instincts (Michalson, 1990, p. 9).³⁴

To be more precise, my aim here is not to deny any kind of effort on Kant's part to make philosophical sense of theological doctrines. This is, indeed, suggested by the very title of the work. However, I do claim that these efforts take place within the framework of his critical philosophy and are not to be seen as a shift or even abandonment of Kant's previous philosophical positions. Instead, the theses presented in *Religion*, however striking they seem, as well as the difficulties connected therewith, can be accounted for immanently, i.e. with recourse to the resources offered by Kant's critical moral philosophy. Here I cannot claim exhaustiveness as this would require a comprehensive analysis of all Kantian theses in *Religion I*, which is beyond the scope of this paper. Nevertheless, I want to provide an overview that relies on what has been identified in the literature as the most striking positions in Kant's text, namely the notion of intelligible deed and the universality of evil.³⁵

First, in *Religion* Kant resorts to the notorious doctrine of an intelligible deed (*intelligibele Tat* [RGV, AA 06: 31]).³⁶ As we have seen, the propensity to evil as opposed to a predisposition cannot be conceived of as *anerschaffen*, but is to be thought of as *acquired* by the very agent. For Kant, this points to an act of freedom by virtue of which the agent has brought this propensity upon himself. Yet since this propensity is understood by Kant as the supreme maxim (*oberste Maxime* [RGV, AA 06: 31, 39]) and must as such lie at the basis of every use of freedom, it cannot be acquired in time. For this reason, he maintains that the act constituting that propensity is an *intelligible* act which precedes all experience. This notion of an atemporal act grounding in some way our empirical conduct

³⁴ To be sure, Horn's and Michalson's positions are not identical. Horn (2011, p. 68) explicitly rejects Michalson's assessment according to which the difficulties found in Kant's treatment of moral evil are a productive crisis. Yet he still sees the source of the difficulties in Kant's *Religion* in what he calls a theoretical bridging (*Brückenschlag*) between Kant's own moral philosophy and Christian dogmatics. On his view, it is this attempt that puts at risk Kant's moral philosophy (Horn, 2011, p. 43).

³⁵ For this, I build upon Horn's (2011, pp. 64–68) critical overview of *Religion I*. This, however, does not mean that I agree with the content of his criticism against Kant. Rather, I take this simply because it seems to offer a standard reading of what many take to be some of the major problems in Kant's text.

³⁶ For a more detailed treatment of this notion, see among others Brown (1984).

has been the target of fierce criticism. Horn calls the intelligible deed downright absurd and says:

However, an action of that kind which the agent does not and cannot have any knowledge [*wissen*] of appears as an absurd construct [*absurdes Konstrukt*]. It is essential for the concept of an imputable decision that it was knowingly and willingly taken at a definable moment in time (2011, p. 66, a. trans.).

Unfortunately, I cannot address here the systematic points made by Horn at this place. For the present purpose, however, it suffices to observe the target of Horn's critique. If the interpretation I have set forth here is correct, then it seems natural for Kant to resort to such a notion. In fact, we have seen that in the absence of an *objective* side to moral evil, Kant cannot account for its possibility other than drawing on an act of freedom performed by the agent himself. And if we pay attention to the systematic function the propensity to evil is meant to fulfil, we will also realise why such a free act must precede all experience, i.e. why the propensity to evil must be present in human beings from birth. As pointed out in section 2, the propensity to evil is used, analogously to the predisposition to personality, to account for the *possibility* of moral evil. If we were to assume, then, that such a propensity is acquired by the agent within his empirical life, then it follows from this that at some point in time such an agent did not even have the very *possibility* to transgress the moral law. Indeed, this possibility would be, in this case, something the agent brings upon himself in the course of his life. Such an assumption would, however, clearly blur the fundamental distinction between the moral condition of human beings and that of holy beings, as Kant developed it in *GMS* and the second *Critique*, for this distinction rests upon the fact that imperfect wills—as human wills are—do not only have the incentives of reason, but also those of sensibility and can, thus, also be determined by the latter. Since such a determination is exactly what Kant means by evil, an agent lacking the very possibility of evil would *eo ipso* also have to lack the possibility of such a determination of his will, thus not corresponding to what Kant takes to be the condition of an imperfect will.

As far as the second point, i.e. the universality of evil amongst humans is concerned, my interpretation allows for an immanent explanation

of this controversial claim as well. The reason for this lies in what has been developed in the previous point. Again, if we pay attention to the large systematic parallelism between the notions of *Anlage* and *Hang* and thereby see that the propensity to evil is thought of as the (subjective) ground of *possibility* of evil, it becomes manifest why a *Hang* must be found in every human being. As in the case of the previous point, this claim, too, functions as guarantor for the Kantian distinction between morality and holiness. The possibility of evil—and therefore the propensity to evil—belongs to the very notion of an imperfect will, namely, a will that beside the rational incentive *can* also be determined by sensuous incentives, which is, therefore, not a pure, but rather a *hybrid* will. Excepting some individuals of the species from this propensity would amount to removing in them the very *possibility* of evil, thus compromising, again, the fundamental distinction between morality and holiness drawn by Kant's critical moral philosophy. This definition of *Hang zum Bösen* as ground of *possibility* of evil is, as we can see from the elaborations in this text, conditioned by the path Kant takes to determine his notion of *Hang*, namely building on a strong analogy to the notion of *Anlage*.

Similarly, we can also address a further, equally striking claim, that of the inextirpability of this *Hang*.³⁷ As has been discussed in section 2, Kant holds that the propensity to evil cannot be eradicated (RGV, AA 06: 31, 51). However provocative this thesis may appear to our eyes, we do not need to refer to the Christian doctrine of original sin to understand Kant's motivation at this place. As is well known, Kant is committed to the ever-present possibility of moral betterment and to the claim that such a betterment, if it is to be called *moral*, must be traced back to the use of our own freedom (see RGV, AA 06: 44–53). He justifies this through the indelible consciousness of the moral law and the principle *ultra posse nemo obligatur*. With the inextirpability of the propensity to evil Kant is thus by no means saying that humans are, quasi as a result of their incurrance of this propensity, doomed to be evil for all days to come. Rather, if we understand this among the lines previously developed, this, too, is a claim he must make if he is to save the fundamental distinction between morality and holiness when dealing with moral evil. Indeed, the claim according to which the propensity to evil cannot be eradicated must be understood as saying

³⁷ Horn (2011, p. 65) addresses this topic under the title *Paradox einer Schuld ohne Schuldfähigkeit*.

nothing more than: every human agent, even after the successful completion of what Kant calls moral *revolution*, is *qua* finite being still *capable* of transgressing the requirements of pure practical reason.³⁸ In other words: even after the firm commitment to morality undertaken by an agent sensibility is still the source of practical incentives which *can* determine his will. Michalson seems to overlook this point when he claims that “the ever-present possibility of moral regeneration entailed by Kant’s rejection of devilishness does not have a counterpart in the case of the moral agent who has in fact undergone regeneration” (1990, p. 82).³⁹ Thus we see here, once again, how the determination of the propensity to evil as accounting for the possibility of moral evil lies at the heart of the issue. Such a determination is, as it were, systematically imposed on Kant by the way he uses to establish his notion of *Hang*, id est, taking as its basis the concept of *Anlage*.

With the aid of these elaborations, we can see how Kant’s major controversial claims in *Religion I* can be sufficiently explained by referring to the framework of his own critical moral philosophy of the 1780s, thus abandoning the explanatory model based on alleged concessions to Christian theology.⁴⁰ It is worth observing that my view also does justice to a statement Kant made in a letter to the Göttingen theologian C. F. Stäudlin. Kant sent him a copy of *Religion* and added in a letter:

The plan I made already long ago to cover the field of pure philosophy was referred to the solution of the three questions: 1) what can I know? (metaphysics.) 2) What must I do? (morals.) 3) What may I hope? (religion) [...]. With the work attached I have sought to carry out the third section of my plan [...]. (Br, AA 11: 414, a. trans.).

³⁸ For a thorough treatment of the notion of moral revolution in Kant see Papish (2018, pp. 177–201).

³⁹ Following this, he rightly notes that “the impossibility of a relapse appears to compromise his [Kant’s, S. C.] theory of freedom, by implying a way in which freedom cannot be exercised” (Michalson, 1990, p. 82). Indeed, this would not only be at odds with Kant’s theory of transcendental freedom, but it would also endanger his distinction between morality and holiness, as I have just said. However, this is—to my mind—not how Kant’s doctrine in *Religion* is to be read.

⁴⁰ Similarly, the mere use of theological language cannot be invoked as an unmistakable sign for these alleged concessions, for, as G. F. Munzel (1999, p. 134) rightly notes, Kant draws upon biblical motifs and language also in his formal moral philosophy. From this observation, she concludes that “[t]he use of the theological language and the parallels drawn in 1793 do not, then, necessarily signal a belated turn to admit orthodox, theological conceptions of human nature, nor is the significance of these allusions self-evident” (Munzel, 1999, p. 134).

With this confession, Kant is explicitly presenting the content treated in *Religion* as the answer to the third of his famous questions posed in the framework of his critical philosophy (KrV, A805/B833). Since Kant himself saw *Religion* within this critical framework, it seems advisable to treat the challenges presented by this text within this context.

To conclude, it should be noted here that my aim was not to show that Kant's doctrine of moral evil in *Religion* contains no difficulties or to provide a systematic solution to them. I also did not defend Kant's position by saying that he is right in establishing this asymmetry between *Anlage zum Guten* and *Hang zum Bösen*. Rather, I merely confined myself to disclosing this striking asymmetry between these two notions within the context of the large systematic conceptual parallelism that guides Kant's theory of evil. Not only did I not support this asymmetry, but I even think that one might raise justified objections against it. As I have already noted (section 2), the Kantian conception of the propensity to evil as being *contingent* (as opposed to *original*) seems to imply that it is possible for human beings not to have the very *possibility* of moral evil, which seems to compromise his own depiction of an imperfect will. Furthermore, Kant's asymmetry here seems to face the question of why the possibility of moral evil cannot be grounded objectively (as it is in the case of moral goodness), namely in the concept of human being as a *hybrid being*, that is, as possessing, as it were, a two-sided nature. Indeed, one could argue that the possibility of evil is given in humans already insofar as they are beings endowed with pure practical reason who, at the same time, can act upon incentives taken from sensibility. Why do we need, in addition to this, a free action in order to ground that possibility? This question is not addressed by Kant's text and is certainly one which deserves a thorough treatment. In this paper, however, my attention was directed only at showing that there is no need to resort to external, to some extent psychologising explanatory attempts to illuminate Kant's striking positions in *Religion*. Indeed, as has been laid out in this work, they can be sufficiently accounted for by philosophical commitments grounded in Kant's critical moral philosophy.

References

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

- Allison, H. E. (1996). *Idealism and Freedom. Essays on Kant's Theoretical and Practical Philosophy*. Cambridge University Press.
- Allison, H. E. (2002). On the Very Idea of a Propensity to Evil. *The Journal of Value Inquiry*, 36, 337–348.
- Bacin, S., Mohr, G., Stolzenberg, J., and Willaschek, M. (Eds.) (2015). *Kant-Lexikon*. Walter de Gruyter.
- Barth, K. (1969). *Protestant Thought from Rousseau to Ritschl*. Simon and Schuster.
- Baxley, A. M. (2010). *Kant's Theory of Virtue*. Cambridge University Press.
- Brown, R. F. (1984). The Transcendental Fall in Kant and Schelling. *Idealistic Studies*, 14(1), 49–66.
- Forschner, M. (2011). Über die verschiedenen Bedeutungen des »Hangs zum Bösen«. In O. Höffe (Ed.), *Immanuel Kant, Die Religion innerhalb der Grenzen der bloßen Vernunft* (pp. 71–90). Akademie Verlag.
- Goethe, J. W. (1949). *Gedenkausgabe der Werke, Briefe und Gespräche* (vol. 19: *Briefe der Jahre 1786–1814*). Artemis Verlag.
- Höffe, O. (1995). Ein Thema wiedergewinnen: Kant über das Böse. In O. Höffe and A. Pieper (Eds.), *Über das Wesen der menschlichen Freiheit* (pp. 11–34). Akademie Verlag.
- Horn, C. (2011). Die menschliche Gattungsnatur: Anlagen zum Guten und Hang zum Bösen. In O. Höffe (Ed.), *Immanuel Kant, Die Religion innerhalb der Grenzen der bloßen Vernunft* (pp. 43–69). Akademie Verlag.
- Kant, I. (1900ss.). *Gesammelte Schriften*. Herausgegeben von der Preußischen Akademie der Wissenschaften, der Deutschen Akademie der Wissenschaften zu Berlin und der Akademie der Wissenschaften zu Göttingen. (AA).
- Kant, I. (1996). *Religion and Rational Theology*. Cambridge University Press.
- Kant, I. (2011). *Groundwork of the Metaphysics of Morals. A German-English Edition*. Cambridge University Press.
- Klemme, H. (2006). Praktische Gründe und moralische Motivation. Eine deontologische Perspektive. In H. Klemme, M. Kühn, and D. Schönecker

(Eds.), *Moralische Motivation. Kant und die Alternativen* (pp. 113–153). Felix Meiner.

Michalson, G. E. (1990). *Fallen Freedom. Kant on Radical Evil and Moral Regeneration*. Cambridge University Press.

Munzel, G. F. (1999). *Kant's Conception of Moral Character. The "Critical" Link of Morality, Anthropology, and Reflective Judgment*. The University of Chicago Press.

Papish, L. (2018). *Kant on Evil, Self-Deception, and Moral Reform*. Oxford University Press.

Schiller, F. (1992). *Schillers Werke. Nationalausgabe* (vol. 26: *Briefwechsel. Schillers Briefe 1.3.1790–17.5.1794*). Hermann Böhlau Nachfolger.

Received: 12/01/2024

Accepted: 12/11/2024