RADICAL EVIL AND THE INVISIBILITY OF MORAL WORTH IN KANT'S *DIE RELIGION*

(EL MAL RADICAL Y LA INVISIBILIDAD DEL VALOR MORAL EN *DIE RELIGION* DE KANT)

Carlos Manrique[1]
University of Chicago
Chicago, USA
manrique@uchicago.edu

Abstract: There is an *aporia* in Kant's analysis of evil: he defines radical evil as an invisible disposition of the will, but he also demands an inferential connection between visible evil actions and this invisible disposition. This inference, however, undermines the radical invisibility of radical evil according to Kant's own definition of the latter. Noting how this invisibility of moral worth is a distinctive feature of Kant's approach to the moral problem, the paper then asks why, in the *Groundwork,* he nonetheless forecloses a question about evil that seems to be consistent with this approach. It is argued that to account for this *aporia* and this foreclosure, one has to interrogate the way in which the category of religion orients Kant's incipient philosophy of history in *Die Religion*.
Keywords: Kant, radical evil, philosophy of religion.

Resumen: Hay una aporía en el análisis kantiano del mal: Kant define el mal radical como una disposición de la voluntad invisible, pero también exige que esta disposición invisible se pueda inferir a partir de aquellas acciones visiblemente malas. Sin embargo, esta inferencia socava el carácter radicalmente invisible del mal radical según la definición que de éste da el propio Kant. Enfatizando la manera en que este carácter invisible del valor moral es una característica distintiva de la aproximación kantiana al problema moral, se plantea la pregunta de por qué Kant, no obstante, rechaza en la *Fundamentación* una pregunta sobre el mal que parece ser consistente con esta aproximación. Se argumenta que para dar cuenta de esta aporía y de este rechazo, es necesario interrogar la manera en la que la categoría de la religión orienta la incipiente filosofía de la historia esbozada por Kant en *Die Religion*.
***Palabras clave*:** Kant, mal radical, filosofía de la religión.

I. The *aporia* of radical evil

Following his distinctive way of inquiring, one of the central questions that Kant addresses in the first installment of his *Die Religion innerhalb der Grenzen der bloßen Vernunft* (1793), is about the *a priori* conditions of possibility that may account for the pervasive existence of evil observed throughout the spectacle of human existence; and hence, that may account for our naming this or that

[1] Candidato a Doctorado en filosofía de la religión.

action, person, or event, as "evil". As is always the case in his path of thought, the point of departure for the inquiry is a linguistic *factum* (a judgment) that calls for an elucidation regarding the transcendental (non-empirical) grounds that make this *factum* in the first place possible; or, in other words, an elucidation of the transcendental grounds that allow us to account for some of our judgments, not as random or arbitrary propositions that could well be absent or falsified given other circumstances, but rather as referring to certain "necessary" and constitutive characteristics of human existence[2]. In this text the question for Kant is not: *there are* synthetic judgments *a priori*, so how are they possible?; nor is it: *there are* moral judgments in which we say "this is good", or *there are* those other judgments in which we say "this is beautiful", so how are they possible?; the question now is rather: *there is* evil; we judge sometimes this or that action, person, or event, as evil, so how is such a judgment possible?; and what do mean when, in such instances, we say "evil"? Such is the question that the first part of *Die Religion* is concerned with, a question that for Kant becomes urgent given the overwhelming evidence he finds of a human propensity to evil, or, in his own words "the multitude of woeful examples that the experience of human deeds parades before us" (R 80 / 6:32)[3]. In virtue of this evidence, says Kant, "we can spare ourselves the formal proof that there must be such a corrupt [moral] propensity rooted in the human being"(R 80 / 6:32). Under the burden of this recognition, he distances himself

[2] This question about the *a priori* conditions of the possibility of evil is inextricably connected to the other central question that Kant addresses in Part I: since evil pertains to the moral realm of imputability its ultimate source must be a "free choice" of the will, and therefore the characterization of the *a priori* "necessary" conditions from which evil derives must be compatible with the understanding of evil as an outcome of human freedom. This results in the apparent antinomy that an evil disposition of the will must be at the same time the ground of all evil deeds and, at the same time, be itself a (freely chosen) deed, an antinomy that Kant resolves with the distinction between a *noumenal* and a *phenomenal deed.* Starting our discussion from a different *angle* and postponing for the moment the explicit mention of this problem concerning the relation between freedom and evil we are not, however, ignoring it. In this respect, it should however be kept in mind that when he refers to the *a priori* ground of evil as "necessary", Kant does not mean that the predicate "evil" can be inferred from the concept of a human being in general (in which case it would not be the outcome of *freedom*), but as he himself says, that "according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, we may presuppose evil as subjectively necessary in every human being" (R 80 / 6:32). And yet, as we shall soon see, Kant also insistently says, that the judgment that a human being is "evil" cannot be based in experience. This is one form of the *aporia* that this first part of the paper attempts to expose.

[3] *Cf.* Kant 1996. (All the quotations from Kant's "Religion" are taken from this edition; the quotes will be followed first by the page number of the English translation and then the original page number of the German standard edition).

both from the Rousseaunian nostalgia for the natural goodness of the "savage" corrupted by the advance of "civilization", as well as from the enlightened optimism in the triumph of "civilization" over the irrational perversity of the "savage". Kant encounters an overwhelming evidence of evil in both scenarios: on the one hand, as he calls them, "the vices of savagery", and on the other hand "the vices of culture and civilization" (R 80-81 / 6:33). The assessment of the universality of evil that this evidence entails, discredits in equal measure the nostalgic pessimism of the romantic as well as the naive optimism of the enlightened bourgeois.

It is in this manner that the first part of *Die Religion* is devoted to a detailed consideration of the question of evil. However, despite such experiential attestation of the existence of "evil", despite such "overwhelming evidence", what this rich and complex text attempts to provide is precisely an understanding of moral evil that displaces the criteria of moral worth from the phenomenal appearance of certain actions recognized and named as *evil*, to the invisible inward disposition from which these actions arise. In Kant's own words:

> We call a human being evil not because he performs actions that are evil (contrary to law), but because they are so constituted that they allow the inference of evil maxims in him. Now through experience we can indeed notice unlawful [*gesetzwidrig*] actions, and also notice (at least within ourselves) that they are consciously contrary to law. But *we cannot observe maxims*, we cannot do so un-problematically even within ourselves; hence *the judgment that an agent is an evil human being cannot reliably be based on experience*. In order, then, to call a human being evil, it must be possible to infer *a priori* from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particularly morally evil maxims. (R 70 / 6:20; my emphasis)

Moral evil is, then, no longer to be considered as the transgression of the content of a specific moral law (do not lie, pay your debts, do not kill, etc.), but as a certain inward disposition, a certain inflexion of the will (which Kant here and elsewhere calls a *maxim*), that, though in itself invisible, must let itself be inferred on the basis of such visible transgressions. One of the main purposes of this first section of *Die Religion* is to determine and characterize the configuration of this inflexion of the will which itself constitutes the source of moral evil. When we say that this or that action is evil, what we mean is not that such an action transgresses some norm, or some prescription of conduct, but rather that it is done out of a certain disposition of the will, a certain maxim. To understand what moral

evil is, then, amounts to understanding what is the configuration of this inflexion of the will, which can alone be properly called "evil" in a moral sense. In a way, Kant is replicating here in his analysis of evil the same revolutionary movement introduced in the analysis of the question of moral goodness articulated in the *Groundwork*: what matters in relation to moral worth is not so much *what* is done, but rather the *how* of this doing. Nonetheless, one finds an intriguing asymmetry between the analysis of the constitution of a *good will* undertaken in the *Groundwork*, and the approach taken here in the analysis and understanding of an *evil will*. In the first case, it is impossible to make an *inference* from an apparently good action, i.e., from an action that conforms to the specific content of the moral law(s), or even from a whole series of this kind of lawful actions, to a good moral disposition from which these actions arise[4]. From the perspective of the phenomenal appearance of human conduct there *may well be* absolutely no difference between a good or an evil will. Moral worth is in this sense, for Kant, radically opaque, hidden, even inaccessible, from a phenomenological perspective. By its very definition, moral worth does not appear, does not show itself, it remains inescapably hidden in what in *Die Religion* Kant calls the "depths of the heart". In this text he emphasizes in several passages, once again, this point initially articulated in the *Groundwork*; among them the following where he establishes the distinction between *a person of good morals* and *a morally good person*:

> So far as the agreement of actions with the law goes, there is no difference (or at least there ought to be none) between a human being of good morals (*bene moratus*) and a morally good human being (*moraliter bonus*), except that the actions of the former do not always have, perhaps never have, the law as their sole and supreme incentive, whereas those of the latter *always* do. We can say of the first that he complies with the law according to the letter [*…er befolge das Gesetz dem Buchstaben nach*] (i.e. as regards the action commanded by the law); but of the second that he observes it according to the spirit [*…er beobachte es dem Geiste nach*]

[4] In the *Groundwork* Kant even stresses the impossibility of such an inference not only in the case of empirical observation, but also in the case of introspection or self-examination: "In fact it is absolutely impossible to settle with complete certainty through experience whether there is even a single case in which the maxim of an otherwise dutiful action has rested solely on moral grounds and in the representation of one's duty. For it is sometimes the case that with the most acute self-examination we encounter nothing that could have been powerful enough apart from the moral ground of duty to move us to this or that good action and to so great a sacrifice; *but from this it cannot be safely inferred* that it was not actually some covert impulse of self-love, under the mere false pretense of that idea, that was the real determining cause of the will […]" (G 23 / 4: 407; my emphasis).

> (the spirit of the moral law consists in the law being of itself a sufficient incentive). (R 78 / 6:30)

By the end of this paragraph Kant adds that in the case of "the person of good morals" (*bene moratus*) who complies with the *letter* of the law without being attuned with its *spirit*, "the human being, despite all his good actions is nevertheless evil" (R 78 / 6:31).

But if the *inference* from phenomenological appearance to the inflexion or disposition of the will is precluded in the case of the determination of moral goodness, it seems to be not only allowed but even more *required* in the case of the determination of moral evil: "In order, then, to call a human being evil, *it must be possible to infer a priori* from a number of consciously evil actions, or even from a single one, an underlying evil maxim" (R 70 / 6:20; my emphasis). Why is this *inference* from the sphere of phenomenological appearance to the inwardness and secrecy of the will *necessary* in the case of the determination of moral evil, but *impossible* in the case of the determination of moral goodness? What is the reason for this asymmetry? In the terms introduced in *Die Religion*, one could say that the necessary character of this inference from phenomenal appearance to the invisibility of the will in the case of moral evil forecloses the possibility of a transgression of the *letter* of the moral law that is somehow attuned with its *spirit*. There may well be an *evil will* hidden under the appearance of *good actions*, but there cannot be a *good will* hidden under the appearance of *evil actions*. The "good citizen" may well harbor an evil heart, but the "criminal" or the "outlaw" cannot be thought of as harboring a good one[5].

In the very typology, already mentioned above, in which Kant divides the examples gathered in his empirical attestation of the alleged universality of the propensity to evil in the human being, one already encounters a difficulty entailed in the presumed "visibility" of such examples. Kant divides his examples in two groups: the

[5] Although the description of "frailty" as the first degree of the propensity to evil in the human condition, indeed allows the thought of this possibility; Kant describes "frailty" in the following terms: "The frailty of human nature is expressed even in the complaint of an Apostle: 'What I would, that I do not' i.e., I incorporate the good (the law) into the maxim of my power of choice; but this good, which is an irresistible incentive objectively or ideally (*in thesi*), is subjectively (*in hypothesi*) the weaker (in comparison with inclination) whenever the maxim is to be followed" (R 77 / 6:30). This motif of "frailty", which opens the possibility of a visible transgression of the law nonetheless bound to a moral inflexion of inwardness, i.e., to the "incorporation of the good into the maxim of my power of choice", would certainly deserve closer examination in relation to the main argument of this paper. Nonetheless, the situation of "frailty" still seems to differ from the *aporia* in Kant's analysis of radical evil which this paper attempts to explore, but the relation between the former and the latter certainly requires further elaboration.

"vices" found in the uncivilized "state of nature" of certain "savages", and those he refers to as "the vices of culture and civilization". The former are visible in the form of explicit and excessive violence, of "unprovoked" and "never-ending cruelty" (R 80 / 6:33). But a quality of *hiddenness* seems to be a constitutive feature of the latter: "secret falsity even in the most intimate friendship", or "a propensity to hate him to whom we are indebted", or "many other vices yet *hidden* under the appearance of virtue" [*…vielen andern unter dem Tugendscheine noch verborgenen*] (R 81 / 6:33; my emphasis). If the "vices of culture and civilization" are hidden, one is then prompted to ask how is it possible for Kant to gather them here in the form of empirical evidence, as part of the "woeful examples that the experience of human *deeds* parades before us". If they are hidden under the appearance of virtue, this means precisely that they do not appear as "vices", even more, that they do not *appear* at all. How can this kind of "vices", then, be pointed out as part of the "multitude of woeful examples" that constitute the overwhelming empirical evidence of the existence of a morally "evil" disposition in human nature? How is this invisible "evidence" supposed to be *seen*?

Perhaps aware of precisely this difficulty Kant rectifies, then, his line of argument and leaves aside the *a posteriori* attestation of a propensity to evil in human nature based on the gathering of examples. He recognizes that these examples cannot "teach us the real nature of that propensity or the grounds of this resistance [of the human power of choice against the law]" (R 82 / 6:35), and states that such an elucidation requires rather the *a priori* articulation of the concept of moral evil, i.e., the identification of the so to say transcendental ground of the *phenomenon of evil*, and then provides such an *a priori* definition. This rectification in his argument is also in concordance with one of Kant's most important methodological principles: that mere empirical observation is always an insufficient source to determine the constitutive structure of a phenomenon, and much less this phenomenon's necessity or universality.

It is in this shift in his argumentation towards the *a priori* elucidation of the transcendental ground of moral evil, that Kant formulates his well known characterization of radical evil as a "reversal of incentives". Through this "reversal" the will subordinates the incentive of the moral law to the incentive of self-interest, or, in other words, regards the (external) conformity to the moral law as a means to securing the incentives of self-interest, instead of subordinating the latter to the unconditional compliance required by the moral law, that should always be an end in itself. In Kant's own words, moral evil is thus characterized as a movement through which the will "makes the incentives of self-love and their inclinations the condition of compliance to the moral law" (R 83 / 6:37).

The will *decides* to conform to *the letter* of the moral law inasmuch as such conformity is conceived as a condition towards the securing of *happiness*, as a "good deal". This compliance to the *letter* of the moral law associated with the inner disposition that constitutes radical evil, is the reason why it is so crucial for Kant to stress the radical invisibility of moral worth, i.e., that from the perspective of the phenomenal appearance of human conduct there may well be absolutely no difference between a *good* or an *evil* will. Even more, the assumption of the contrary, i.e., that the empirical evidence of external conduct can be in itself an indication of moral worth is an unequivocal expression of an "attitude of mind" that he designates as the "radical perversity in the human heart":

> [E]ven though a lawless action and a propensity to such contrariety, i.e. *vice*, do not always originate from it, *the attitude of mind that construes the absence of vice as already being conformity to the disposition to the law of duty (i.e. as virtue) is nonetheless itself to be named a radical perversity in the human heart* (for in this case no attention at all is given to the incentives in the maxim but only to compliance with the letter of the law). (R 84 / 6:37; my emphasis)

Pressing this same point further Kant characterizes, some lines ahead, this "radical perversity in the human heart" as a *dishonesty* that "puts out of tune the moral ability to judge what to think of a human being, and renders any imputability entirely uncertain, whether internal or external" (R 85 / 6:38). In his examination of the Kantian doctrine of radical evil Henry Allison, reminding us that this "radical perversity of the human heart" has been previously identified by Kant as the third and highest degree of the propensity to evil in human nature (after fragility and impurity), gives us an accurate and succinct account of the dishonesty constitutive of this evil disposition of inwardness, in terms of self-deception:

> Kant suggests that a fundamental feature of this third stage is a kind of systematic self-deception. The idea here is that one tells oneself that one is doing all that morality requires *as long as one's overt behavior agrees with the law*. Accordingly, Kant suggests that this stage can coexist with a certain ungrounded moral self-satisfaction, which stems from the fact that one has simply been fortunate in avoiding those circumstances that would have led to actual immoral behavior. (Allison 158; my emphasis)

But put in these terms, the definition of *radical evil* could have a shocking consequence that is, however, not explicitly acknowledged by Kant (nor by Allison). If *radical evil* is defined in terms of this kind

of dishonesty (an evil moral disposition hidden under the appearance of virtue), then the blatant and un-hidden transgressions of the moral law cannot be regarded as expressions of that "attitude of mind" [*Denkungsart*] which constitutes *radical evil* insofar as, in such cases, the very explicitness and visibility of in-morality exhibited by such un-hidden transgressions, precludes the very possibility of evil *qua* dishonesty and self-deception, this is, of evil *qua* radical evil. In other words, visible evil cannot be the manifestation of radical evil, because according to its very definition the latter is one that *hides itself* under the appearance of virtue.

This is certainly not what Kant has in mind. As was already noted, concerning the question of evil his project is to determine the transcendental grounds that may account for the empirical instances of evil actions observed throughout human experience. He describes the project as that of making an inference into the evil maxim (the subjective ground of the will) that underlies the phenomenal appearance of evil. In this vein he states, in the very first paragraph of Part I of *Die Religion*, already quoted above: "We call a human being evil not because he performs actions that are evil (contrary to law), but because they are so constituted that they allow the inference of evil maxims in him" (R 70 / 6:20). Right away, nevertheless, he puts into question the very possibility of this inference by recognizing that: "we cannot observe maxims, we cannot do so un-problematically even within ourselves; hence the judgment that an agent is an evil human being cannot reliably be based on experience" (R 70 / 6:20). The tension is forcefully present in the phrasing of these opening sentences: moral evil must be identified in the inward maxim and not in this or that empirically observable action; but visibly evil actions may nonetheless "allow" the inference of the evil maxim from which they derive. But if we cannot rely on the empirical observation of actions in order to assess their moral worth, we have no criteria left for identifying certain actions as "evil" that would "allow" us to infer from them the invisible principle (maxim) from which they presumably derive. In order to infer an evil maxim from evil actions we have to be able to know that certain actions are evil, to recognize them as such; but we can only know whether an action is evil if we first know the maxim from which it derives.

In the very statement of the project of seeking the transcendental (*noumenal*) conditions of possibility of the *phenomenon* of evil there seems to be, then, an *aporia*. Perhaps this *aporia* could be further characterized in terms of the impossibility of reconciling, on the one hand, a direct connection or continuity between the phenomenal instances of evil observable in human experience and an inward evil disposition which "no one sees"; and, on the other hand, in an inescapable friction with this alleged continuity, the radical

incommensurability between the phenomenality of human conduct and the invisibility of the will that is so crucial in the articulation of Kant's understanding of the moral problem. The direct continuity between the *visible* and the *invisible* dimensions of evil requires a certain inference that the incommensurability between the visible "outside" and the invisible "inside", the inaccessibility from one to the other, precludes.

The role of this incommensurability in Kant's analysis of radical evil, has led some of the interpreters of this difficult text to claim that this analysis entails a series of consequences which are "morally scandalous". In this vein, for example, Richard Bernstein has pointed out to what he calls "a troubling consequence" of Kant's analysis of evil. He notes how, contrary to all expectations, in the characterization of wickedness [*Bosartigkeit*] as the third (and highest) degree of the "propensity to evil in human nature", "wickedness" is not conceived by Kant as "some horrendous type or form of evil" (Bernstein 71), but rather as a subtle and perhaps even unnoticeable arrangement of the will's incentives. An arrangement which, furthermore (as we have insisted above), could be in Kant's view accompanied by an irreproachably lawful conduct, a conduct that is, in all respects, correct. It is this radical invisibility of radical evil what Bernstein finds so troubling in Kant's analysis, insofar as it undermines the very possibility of establishing a moral distinction between the "good citizen" and the "criminal" (i.e., the very possibility of making an inference from the legal to the moral spheres). To express his indignation, Bernstein rhetorically pushes the point a bit further by depicting the "good citizen" as the "sympathetic person" who helps others out of a natural inclination (i.e., not by incorporating the moral law as the supreme and unconditional incentive of the will), and the criminal as the "mass murderer", and by then noting that:

> On the basis of Kant's characterization of wickedness, such a self-consciously motivated sympathetic person whose actions are 'lawfully good' is a paradigm of wickedness. He has a cast of mind that is corrupted at the root, and he must be 'designated as evil'. […] But to judge such a person to be an exemplar of wickedness; to judge his maxims –in respect to the degree of evil- to be in the same category as those of the mass murderer is much more than an awkward consequence; it is *morally perverse*. (Bernstein 71; my emphasis)

For Bernstein, then, it is morally perverse to regard moral worth as utterly invisible or, in other words, to establish a fracture between "legality" and "morality" in the way Kant does (i.e., between what Kant would also call the "letter" and the "spirit" of the law), such

that it is impossible to "see" or recognize any moral worth in the mere conformity to the law (only on the basis of such a recognizable visibility of moral value could one then sharply oppose the moral worth of the "sympathetic good person" and the "criminal"). For Kant, nonetheless, it is exactly the other way around: what he considers "morally perverse" is to assume that it is possible to discern moral worth on the basis of the mere conformity to the law:

> [T]he attitude of mind that construes the absence of vice [the mere conformity to the law] as already being conformity to the disposition to the law of duty (i.e. as virtue) is nonetheless itself to be named a radical perversity in the human heart. (R 84 / 6:37)

Bernstein's scandalized indignation, hence, is symptomatic of the profound and disturbing displacement in the very approach to the moral question that is effected in Kant's analysis of radical evil: the establishment of an insurmountable rupture between the visible surface of conduct and the invisible depth of a moral inflexion of the will in which, then, the entire question of moral worth is situated. This radical displacement effected by Kant's moral philosophy is not exclusive of this later text, but rather, informs his very formulation of the moral problem since as early as the *Groundwork*. One could not, as Bernstein would pretend to do, keep Kant's moral philosophy and get rid of his perplexing analysis of evil, because both are ultimately articulated on the basis of the same fundamental intuition: the displacement of moral worth from the visible surface of conduct to the invisible depth of the will. Instead of a scandalized indignation that is ultimately grounded in nothing else but an unexamined "common sense" and "moral sensibility", a "common sense" which Kant's philosophical analysis attempted precisely to disqualify as a source of moral judgments, one should rather ask why Kant retracted from the radical displacement distinctive of his own approach to the moral problem; and, furthermore, ask not only what are the conceptual consequences of this retraction, such as the *aporia* in which his analysis of radical evil is entangled as we have tried to show, but also what are the historico-political "hidden springs" and implications of this retraction, which, in its turn, might well approximate to the "hidden springs" that underpin the scandalized moral sensibility of interpreters like Bernstein.

II. A question regarding the foreclosure of a question in the *Groundwork*

Kant's understanding of radical evil in *Die Religion*, then, opens the same *fracture* between phenomenal visibility and hidden inwardness

that was already established in the *Groundwork* in his understanding of moral goodness. One is intrigued, then, by how this fracture between visibility and invisibility, effected with extreme thoroughness by his analysis of the question about 'moral goodness' (a fracture which in its most radical formulations is presented as irreparable), tends nonetheless to be either passed over or finally repaired when it is a matter of thinking about 'moral evil'. One is intrigued by why the destabilizing and disturbing putting into question of the 'goodness' of what appears as 'good behavior', is not replicated with the same strenuousness when it is a matter of putting into question the 'evilness' of what appears as 'evil behavior'. One asks why the affirmation of the at once poignant and elusive *inflexion of the will* which Kant calls 'respect for the moral law', or 'freedom', is so firm in making tremble what one would call (with the necessary precaution that such encapsulations demand) a certain bourgeois moral self-complacency, but is at the same time so wavering when it is a matter of undermining and putting into question the repulsion that the uncivilized-criminal-unrest has always inspired in the 'civilized world', then and now. In this section of the paper, I will attempt to retrace this intrigue back to the opening formulations of Kant's analysis of the moral problem in the first pages of the *Groundwork*, in order to identify a certain question regarding moral evil that remains foreclosed in this analysis, and to identify, as well, the argumentative devices through which such a foreclosure is sealed. In the third and last section of the paper I will come back to examine what this foreclosure might entail and how can one account for it, in the context of other important aspects of Kant's philosophical analysis of religion in *Die Religion*.

In order to formulate this intrigue from within the Kantian text itself one needs to read it very closely. In the opening pages of the *Groundwork*, for instance, when what is at stake is the definition of the concept of a "good will" through the notion of *duty*, we encounter the denial of a question, the passing over [*übergehen*] a question:

> I pass over all actions that are already recognized as contrary to duty […] for with them the question cannot arise at all whether they might be done from duty, since they even conflict with it. (G 13 / 4:397)[6]

[6] (All the quotes from the *Groundwork* are from Yale edition, and the English translation's page number is followed by the page number of the *Akademie Ausgabe*, from which the German original is also quoted occasionally). [*ich übergehe hier alle Handlungen, die schon als pflicht-widrig erkannt werden (…) denn bei denen ist gar nicht einmal die Frage, ob sie aus Pflicht geschehen sein mögen, da sie dieser sogar widerstreiten*].

And yet, one wonders what does it finally mean for an action to be 'contrary to duty' [*Pflicht-widrig*], and is lead to ask if the distinctive articulation of this matter in Kant's argument makes it indeed so un-problematic to recognize [*erkennen*] such actions; if it makes indeed of this recognition a procedure in which "there is not even once the question […]" [*nicht einmal die Frage…*]. Because, even granting that if there were actions which could be un-problematically recognized as contrary to duty (let us say "*evil*" actions), then there would be no question about whether if such actions have been performed out of duty [*aus Pflicht*], the question still remains: is it the case that an action can be so un-problematically recognized as contrary to duty, as *Pflicht-widrig*, and how so? With this in mind, if one follows the definition of the concept of duty given by Kant as "the necessity of an action from *respect* for the law" (G 16 / 4:400) [*aus Achtung fur Gesetz*], one has to conclude that an action contrary to duty would be one which is not necessitated by this peculiar inward disposition that Kant calls *respect*. Defined in this manner *duty* is, then, not a specific behavior, not a specific action or set of actions, but rather an *inflexion* of the will. If this is the case, though, to recognize an action as contrary to duty would amount to probe the depths of inwardness, to measure its inflexion, its tonality, a procedure of probing and measuring which Kant himself, very soon in his argument, will explicitly regard as impossible. In fact, after having effected this radical displacement of the *center of gravity* of the moral problem from the (visible) appearance of any action or practice to the (invisible) inner disposition that is somehow *connected* to it, Kant's line of argument arrives to the conclusion that moral worth can never be *seen*: "because when we are talking about moral worth, it does not depend on the actions which one sees, but on the inner principles, which one does not see" (G 17 / 4:407) [*wenn vom moralischen Werte die Rede ist, es nicht auf die Handlungen ankommt, die man sieht, sondern auf jener inneren Prinzipien derselben, die man nicht sieht*]. How is it then, that certain actions can be so un-problematically *recognized* [*erkannt werden*] as immoral, and then passed over, if when it comes to the thinking about moral worth it all depends on an inflexion of the will that cannot be *recognized* in what shows itself to be seen? How could this inflexion be so easily grasped, decided upon, on those cases that appear as transgression of the moral law, and nevertheless be un-recognizable, un-decidable, on those cases that appear to conform to it? Why does the inwardness of 'the criminal', of the 'outlaw', remain so unquestionably transparent, while that of the 'good citizen' becomes so drastically opaque, so that it is so easy to decide (to see and to recognize by seeing) the 'evilness' of those practices that appear to be 'evil', and yet so difficult, even more,

impossible[7], to decide upon the 'goodness' of those that appear to be 'good'?

Still, when this "passing over" takes place in the opening pages of the *Groundwork* one certainly follows the logic of Kant's argument in all its apparently unquestionable transparency. We all know the story well. The point is to distinguish in the sharpest possible way those actions done out of duty [*aus Pflicht*], from, on the other hand, those actions whose subjective ground is what Kant calls "inclination" [*Neigung*]. He wants to reduce his analysis to the most critical *type* of actions, those in which this distinction is the most difficult to establish since they are *at the same time* in conformity with duty and also the object of an 'immediate' inclination. Also, it should be noted that at this point it is a matter of a conceptual differentiation and not a question of whether one is able to recognize this difference in experience. First we establish the conceptual difference, first we define in all its rigor the concept of a 'good will' through the concept of duty, and then we ask if it is possible to recognize a 'good will' in experience or not. Perhaps, then, the reasoning in the previous paragraph was too hasty and, perhaps more, it reflects an incompetence to follow the logic of Kant's argument. Perhaps it is not legitimate at all to point at this previously quoted passage with suspicion and claim that there is a certain *gap*, a certain omission, since the matter could not be more clear: when he talks about "passing over" those actions already recognized as "contrary to duty", for with them "there is not even once the question" if they could be done "out of duty", Kant is working out the definition of a concept, the concept of a "good will", a definition for which he needs to clarify what does duty mean, since a "good will" is precisely that which performs *dutiful* actions *out of duty alone*, and not *out of* an immediate or a mediated inclination. If a good will is that which performs *dutiful actions out of duty alone*, to understand what is at stake in this '*out of duty alone*' on which all the definition relies at this point, it is useless to ponder on those undutiful actions since, being *contrary to duty*, it is impossible (there is no way, there is no chance) that they could be done *out of duty alone*. In view of this impossibility, there is "not even once the question". Perhaps this is all there is to it, all that is at stake in what was previously called with unjustified and perhaps premature suspicion, the "passing over" [*übergehe*] a question.

---

[7] "In fact it is absolutely impossible to settle with complete certainty through experience whether there is even a single case in which the maxim of an otherwise dutiful action has rested solely on moral grounds […]" (G 23 / 4:407) ["*In der Tat ist es schlechterdings unmöglich, durch Erfahrung einen einzigen Fall mit völliger Gewissheit auszumachen, da die Maxime einer sonst pflichtmässigen Handlung lediglich auf moralischen Gründen und auf der Vorstellung seiner Pflicht beruhet habe*"].

But perhaps not. In order to decide the issue it would be necessary to make a pause, to allow oneself to be captured by these actions 'contrary to duty', to not pass them over, even if one follows the clarity and transparency of the logic (and the strategy) of Kant's argument. Make a pause and ponder on the impossibility prescribed by the strategy of this logic (and the logic of this strategy): It is impossible that an action *contrary to duty* could be *done out of duty*. What is the logic that grounds this premise? What is the necessity of this logic? Is it a purely logical necessity, such that "an undutiful action done out of duty" would be a self-contradictory statement, a proposition that annuls itself in its absurdity, in its impossibility? Is the thought of an "undutiful action done out of duty" as impossible and self-annulling as the thought "not x and x"? This would be so only if "duty" had the same meaning in the two terms of the statement: "an action contrary to duty" ("an undutiful action"), and "an action done out of duty" ("a dutiful action"). But one soon realizes this is not the case in the thread of Kant's argument, because in the first term of the statement "duty" is meant in the sense of the specific content of *a* "moral law", whereas in the second the term "duty" is meant in the sense of "*respect* for the moral law". Consequently, the apparently contradictory statement "an unditiful dutiful action" (meaning "an action contrary to duty done out of duty") could be translated for "an action contrary to the content of the moral law done out of respect for the law". Still, though, the statement appears to be a contradiction: is it possible to think of "an action against the content of the law done out of *respect* for the law?" But now, nevertheless, it is clear that at least the contradiction could not be a logical one (such as "not x and x"), because there is a crucial difference between the two terms: "conformity to the (content) of the moral law" and "respect for the moral law", such that the opposite of the former does not amount to the negation of the latter (as in "not x and y"). Even more, if the thought of an "unlawful action done out of respect for the law" were a logical contradiction ("not x and x"), then the proposition obtained from the replacement of the first term for its opposite ("x and x"), would be a tautology. Hence, it would be a tautology to say: "an action in conformity to the moral law done out of respect for the law"; but it is precisely the *entire* attempt of Kant's moral philosophy to show that this statement *is not* a tautology because there is an *abyss*, a fundamental difference, the difference that in its subtlety makes the whole difference in relation to the moral problem, between "conformity to the law" and "respect for the law". If it is, then, not a logical necessity that which precludes as impossible the thought of "an action contrary to duty done out of duty" [*ist gar nicht einmal die Frage*], then what kind of *necessity* is it, and why should we be bound by it? Should we be bound by it?

It is, then, important to note that as far as inwardness (i.e., the invisibility of the will's inflexion, or as Kant calls it, of the will's maxim) becomes more and more the center of gravity of the moral problem, the distinction between conformity to duty *without respect* and contrariety to duty *without respect* (a distinction with which Kant is operating when he "passes over" the question about those actions 'contrary to duty', in order to examine rather those that are 'in conformity with duty') becomes less and less relevant. It tends to efface itself, insofar as 'contrariety to duty' *is* itself defined as the absence of *respect* ("duty is the necessity of an action from *respect* to the moral law"); hence, to be 'contrary' to duty, to be against duty, is to lack that peculiar inflexion of the will called *respect*, in which case "conformity to duty without respect" would make no sense, and "contrariety to duty without respect" would be a mere tautology. But still, up to a certain point in the argument the distinction makes sense, and it does so only in virtue of a certain ambivalence that haunts not only the concept of "duty", but also the concept of the "moral law". The distinction makes sense if "duty" and "law" refer to *the specific content* of certain prescriptions for conduct: do not lie, pay your debts, do not kill yourself or anyone else, etc. In that case, one can act in conformity or in contrariety to *the content* of these prescriptions, and although one can still be 'immoral' in both cases if one does not act 'out of respect', it still makes a difference for Kant insofar as in the case of the behavior in contrariety to the *specific content* of these prescriptions there can be no question about its 'immorality', whereas in the case of the behavior in conformity to it the question remains open. But, if "duty" and "law" do not mean *the content* of a prescription for conduct but rather *an inflexion* of inwardness referred to as *respect*, and the whole gravity of the moral problem relies on the extent to which the will is attuned or not with this inflexion of inwardness, the difference between external conformity or transgression of *the content* of certain prescriptions for conduct tends to become more and more irrelevant.

In the *Groundwork*, Kant's text oscillates between these two connotations of the concept of duty: "duty" as *the content* of a prescription for conduct, or "duty" as *an inflexion* of inwardness, of the will. One should note precisely this oscillation operating in the passage previously quoted where we pointed out the foreclosure of the question whether an action 'contrary to duty' [*Pflicht-widrig*] could be 'done from duty' [*aus Pflicht geschehen*]: "I pass over all actions that are already recognized as contrary to *duty [content of a prescription for conduct]* […] for with them the question cannot arise at all whether they might be done from *duty [inflexion of the will]*, since they even conflict with it". The same semantic ambivalence operates throughout the text with the concept of the "moral law", which sometimes

means the specific *content(s)* of the prescription(s) of certain actions (as when Kant speaks of 'conformity to the law'), and sometimes it means rather the mere *form* of the law devoid of any specific *content* (as when Kant speaks of '*respect* for the law'). The following passage, which comes right after the first formulation of the categorical imperative, indicates these two possible meanings of the law [*Gesetz*]: "Here it is merely lawfulness in general (without grounding it on any law determining certain actions), that serves the will as its principle […]" (G 18 / 4:402) ["*Hier ist nun die blosse Gesetzmässigkeit überhaupt (ohne irgend ein auf gewisse Handlungen bestimmtes Gesetz zum Grunde zu legen)* […]"].

The subjective imprint of the moral law in its purely formal sense as "mere lawfulness in general", devoid of any specific content, is the inflexion of the will characterized as "respect". But even if Kant establishes this distinction between the "law" as *content* and the "law" as *form* of inwardness (a distinction which later in *Die Religion* will be formulated as that between *the letter* and *the spirit* of the law [R 78 / 6:30]), the persistent semantic ambivalence in his text that makes the argument oscillate between these two meanings almost inadvertently, reveals that despite the distinction, the two meanings tend to be conflated and regarded as inescapably bound to each other. In this sense, the *form* of inwardness shaped by the "mere lawfulness in general", i.e., by a moral law that cannot be identified with any specific content(s), is nonetheless assumed to overlap with *the content* of a specific set of moral law(s), of moral prescription(s), clearly conditioned (as they always are) by a particular socio-historical *topos*. In virtue of this implicit demand, it is then unconceivable to think of a moral law as *form* of inwardness which manifests itself in the transgression of the content of the moral law(s) / prescription(s) of this specific socio-historical *topos*. Even when a radical distinction is established between the 'letter' and the 'spirit' of the law, and even when the latter is constituted as the sole center of gravity of the moral problem, the possibility of the 'spirit' of the law transgressing the 'letter' of the law remains foreclosed. But what kind of necessity or authority dictates this foreclosure? And, should not this authority and necessity be put into question?

It is a question, then, about the consistency with which Kant carries out in his texts the revolutionary claim that the moral problem is not about *what* is done, but about the *how* of what is done; his claim that moral worth rests entirely upon an inner *how*, a very peculiar and complicated inflexion of inwardness, and not in the performance of a prescribed course of actions. It is a question about the implications of this radical movement, a question of whether all these implications are followed all the way through by Kant's

thought or not, and if they are not, a question about how could one account for the restrain of his thinking from doing so.

Strictly speaking, once such a fracture has been opened between the invisible inner disposition and the visible material content of a course of conduct, and once all the emphasis is put on an elusive inner *how* which is inaccessible starting from any *what*, from any empirically observable practice, the same fracture, and the same inaccessibility, would have to be acknowledged in the case of the practices that conform to the (contents of the) law, and those that transgress the (contents of the) law. If what ultimately matters in relation to moral worth is the inflexion or disposition of the will, the same has to be true for what concerns moral goodness, and what concerns moral evil. Which is to say that, since the ground of moral worth is the pure inner *how* (what Kant later in *Die Religion* will repeatedly refer to as "the bottom of the heart" (R 92 / 6:48; 95 / 6:51) [*die Tiefe des Herzens*]), and not *what* conduct is followed, there are no longer any external actions which are *in themselves* morally *good*, since the most dutiful [*pflichmässigen*] conduct (i.e., a conduct that entirely conforms to what the moral law(s) prescribes, to its content), may still be grounded on an in-moral (evil) inflexion of the will. In the same way, it should have to follow that there are no actions which are *in themselves* morally *evil*, since even the most evident transgression of *the content* of *certain* (culturally and historically circumscribed) law(s) *could* be connected to a morally good inner inflexion of the will. A surprising statement for which, nonetheless, one can unexpectedly find a certain support within the Kantian text itself, here another text, a footnote in Part One of *Die Religion*:

> Thus the perpetual war between the Arathapescaw Indians and the Dog Rib Indians has no other aim than mere slaughter. In the savage's opinion, bravery in war is the highest virtue […] That a human being should be capable of adopting as his goal something (honor) which he values more highly still than his life, and of sacrificing all self-interest to it, this surely bespeaks a certain sublimity in his predisposition. (R 80 / 6:33) […*beweist doch eine gewisse Erhabenheit in seiner Anlage*]

We should certainly not fall into the temptation of over-emphasizing the import of this surprising, even if highly qualified, gesture of deference from Kant's part towards the "slaughter with no other aim" of certain tribes of "savages" in "the wide wastes of Northwestern America". It is a marginal footnote. Still more, this footnote appears in the context of an *excursus* in Kant's argument intended to point out how a mere glance throughout the spectacle of human experience confronts us with such a "multitude of woeful examples" (R 80 / 6:32), that perhaps a formal (*a priori*) proof that there is a

propensity to evil in the human condition is not even necessary. There is no question, then, that the blatant cruelty and violence exhibited by these "savages" is regarded by Kant as an undeniable manifestation of moral evil. Nonetheless, it is highly intriguing that such blatant cruelty and violence "with no other aim" can be, nonetheless, associated with a disposition of inwardness in which Kant recognizes a "certain sublimity"; i.e., with a certain inflexion or tonality of the will that at least *in some sense* reveals an striking affinity with the configuration of a good will. This affinity results from the fact that the "savage's" *purposeless evil* operates a disruption of the logic of calculating self-interest similar to the disruption of this logic required by pure practical reason, by the unconditional character of the moral imperative. Certain instances of blatant transgression of the moral law in which the transgressor's well-being and life itself are risked or injured, wounded, instances of what one would call *self-destructive evil*, on the one hand, and pure respect towards the moral law, on the other, both converge in breaking the logic of rational calculation by means of which *economic reason* seeks to secure the attainment of happiness, the satisfaction of self-interest, through the most intelligent means.

This footnote, then, could nevertheless serve us in the manner of a hint, an indication. It allows us, at least, to reopen the question foreclosed in the opening pages of the *Groundwork* regarding the character of those actions "contrary to duty" [*Pflicht-werdig Handlungen*] that are *recognized* as "evil" because they transgress the content (*the letter*) of certain specific moral law(s). Even more precisely, it allows us to reopen the question of whether these actions recognized as "evil" could be connected to a *good* moral disposition or not, or in Kant's own terms, the question of whether an action *contrary to duty* could be done *out of duty*. But the question now should be, then: why does Kant foreclose this question by repairing the fracture between the visible and invisible dimensions of human conduct in the case of moral "evil", as we have already shown by pointing out the *aporia* in which his analysis of radical evil is, in virtue of this repairing, entangled?

III. The opacity of moral worth and the instability of the distinction between "moral" and "cultic" religions

A tentative answer to this question will allow us to move further into the final stage of our argument: the question regarding the possibility of an empirically observable "evil" action, a transgression of the content of the law, somehow connected to a *good* inflexion of the will is foreclosed, because ultimately the fracture between these visible and invisible dimensions of human conduct must be

repaired, not only in the case of moral evil, but also, and perhaps more importantly, in the case of moral goodness. And this fracture needs to be repaired, so that the history of the world can be understood and told in a certain way, according to a narrative in which the main role is played by the crucial distinction between two different families of religions, sharply drawn by Kant in *Die Religion*: the "moral religion" and the "cultic religions". If the "Arathapescaw and the Dog Rib Indians" cannot be said to be *good*, despite a certain "sublimity in their disposition" that can be glimpsed in the midst of their terrifying and self-destructive violence, it is ultimately because their religion is not "moral", because it is not the *true* religion which alone is a "rational one", in sum, because they are "savages", they are behind in the march of history. This manner of telling the history of the world by means of a distinction between a "true" and a "false" modalities of religion (so typical, by the way, of the main figures of the Enlightenment), requires that one can identify which is the only "moral religion"; and this requires, nothing more and nothing less, that one can be able to demarcate, in history, which people are "moral" and which people are not. If the very possibility of such a demarcation is precisely what has been profoundly problematized and undermined, as we have insistently noted, by the fracture opened between the phenomenological appearance of conduct and the hiddenness of the will in the most radical moments of Kant's analysis of moral goodness and radical evil, this fracture has to be repaired, so that the "good religion" can prevail over the "bad religions", in history.

If, as we said above, there is an asymmetry in the manner in which the inferential movement from actions to maxims, or from observable conduct to the invisible inflexion of the will, is regarded as impossible in the case of moral goodness, but necessary in the case of moral evil, this is only because the impossibility of such inferential movement in the first case, despite being the distinctive mark of Kant's analysis of the moral question, is only a provisional impossibility. And this is so, because the very possibility of somehow discerning on the basis of this inferential movement a progression towards moral goodness is going to be central to the distinction between "moral religion" and "cultic religions" (or "religions of rogation"). As we will see in this last section, if such a discernment of "moral improvement" is put into question or rendered problematic, the very distinction between a "moral religion" and a "cultic religion" becomes deeply dubious, or even more, impossible. In order to understand why this is so, we should briefly retrace the manner in which Kant conceptualizes what a "moral religion", or what is the same thing, what a purely rational religion, a religion within the limits of reason alone, consists of.

In the introduction to *Die Religion* Kant postulates roughly the same conception of the relation between morality and religion, which he had previously articulated in the Second Critique. Such a conception is summed up in the double edge formula: "morality in no way needs religion" (R 57 / 6:3), but "morality inevitably leads to religion" (R 59 / 6:6). It does not need religion because the moral law demands an unconditional compliance, this is, a compliance which makes abstraction of the consequences or of the end towards which the action might be oriented. This abstraction of consequences and ends defines the moral character of a decision, and it is utterly incompatible with a "moral" behavior grounded on the purpose of pleasing God. But Kant also thinks that reason is incapable of renouncing completely to the representation of an end towards which the action is oriented. If such a representation cannot be the ground of one's actions (in which case such actions could no longer be moral), the possibility must still remain that there is a representation of an end that does not precede the action (as its ground or subjective incentive), but rather follows it. After the question "What should I do?" has been answered: "obey the moral law out of pure respect for it and nothing else", and after such a behavior has been adopted and exercised, the inevitable (and legitimate) question: "What can I hope for?" arises as a consequence of this behavior. For Kant, a "rational religion" is precisely one which, not being in any sense the ground or condition for moral actions, nonetheless arises as a consequence of the exercise of pure practical reason, in the form of this question concerning the end or consequences of this exercise as such: "What will come out of all this pure respect for the law?".

In the Second Critique Kant had interpreted "immortality" and "God" as practical postulates, precisely in this respect. Although "immortality" as such must be postulated not in order to hope for a happiness in correspondence to moral worth, but rather to warrant an infinite progression that the always impossible and unattainable adequacy to the requirement of the law demands, "God" is in fact postulated as the condition for an (otherwise impossible) equilibrium between morality and happiness. But in *Die Religion* this relation between religion and morality appears to be somewhat different. At the outset of the Third Book which is concerned with tracing the developments by which the "good principle" overcomes the "evil principle" in the human condition, Kant asserts (in opposition to the double edge formula stated in the introduction to which we have just referred) that the human being needs an "ethical community" in order to overcome this "evil principle". Without an "ethical community", says Kant (rehearsing a certain dose of fatalism and misanthropy that had already transpired in some passages of his analysis of "radical evil" in the First Book), any human association

is inevitably bound to corrupt the moral disposition of its members, and inevitably leads them to their doing each other "evil" ("envy, addiction to power, avarice, and the malignant inclinations associated with these, assail the person's nature, as soon as he is among other human beings" R 129 / 6:93). Thus, a community of people committed to the pure respect for the moral law and thus to the cultivation of a good moral disposition is required for the overcoming of the propensity to "evil" in human nature[8]. This "ethical community" is distinguished from a "juridico-political community" in that the former is bound to moral laws whereas the latter is bound to the public laws of the state and, consequently, whereas the compliance to the laws of the juridico-political community can be determined empirically, through mere observation of external behavior, the compliance to the moral laws (which requires goodness in the inner disposition) cannot be determined by the observation of the human eye. Thus, such an "ethical community" cannot be thought of without the idea of God as the supreme legislator of this community, who is the only one capable of fathoming the "depths of the heart". In such an "ethical community" the moral laws are, then, in this respect, regarded as divine commandments.

This is, argues Kant, what defines a rational-moral religion in sharp contrast to its antipodes, which he calls throughout this text in several ways: "ecclesiastical faith", "cultic religions", "religions of rogation", "counterfeit service", etc. In these immoral (non-rational) religions the order of the relation between "moral law" and "divine commandment" is reversed: what are first recognized as divine commandments (through revelation or tradition) are *then* considered moral duties. The problem with this inversion, for Kant, is that certain actions (ritualistic, cultic) that are not "*in themselves*" moral, are considered as moral duties. But, as we have tried to explain above, it is precisely this idea that there are certain actions which are "in themselves" moral what Kant's analysis of the moral problem (in the *Groundwork*) and of "radical evil" (in *Die Religion*) has rendered deeply problematical, and has put into question; when the idea that an action has "in itself" moral worth independently of the inflexion of the will from which it arises is rendered problematical, the very ground if not of the difference itself, at least of the ability of recognizing this difference between a "rational-moral religion" and all the other "bad" religions, is also seriously destabilized.

[8] "[I]nasmuch as we can see, therefore, the dominion of the good principle is not otherwise attainable […], than through the setting up and the diffusion of a society in accordance with, and for the sake of, the laws of virtue" […] "an association of human beings merely under the laws of virtue can be called an *ethical community*" (R 130 / 6:95).

If the "moral religion" is defined as that in which moral worth becomes the condition *sine qua non* for the *hope* of a happiness that corresponds to this worth (a happiness of which God is the condition of possibility), a certain consciousness of this worth becomes a necessary criteria for distinguishing a moral ("true") religion from a cultic ("false") one. Kant states the distinction formulating the principle that "It is not essential, and hence not necessary, that every human being know what God does, or has done, for his salvation; but it is essential to know what a human being has to do himself in order to become worthy of this assistance" (R 96 / 6:52). But it seems to be essential to the distinction not only to know "what a human being has to do himself in order to become worthy", *but also that the self-consciousness of this worthiness is somehow possible*. For if such consciousness of the progress towards the good (or as Kant himself says, of the process of "becoming a better person") is not possible, then the awareness of the difference between a moral and an immoral religion is not possible either, since the moral religion is that which alone is conducive to moral improvement. Now, how is this consciousness of moral worth possible if external actions in themselves cannot tell us anything about the inner principles (the "bottom of the heart") from which they arise, and if this inner principles cannot ultimately be determined through self-examination either[9], and thus remain utterly opaque, utterly inaccessible? If both empirical observation and introspection are discarded, what is, then, the criterion upon which it is possible to determine whether there is progress towards the good or not, and consequently whether a form of religiosity is "moral" or rather "superstitious" (cultic)?

In several passages Kant argues that a certain "confidence" on the gradual realization of moral progress is a defining characteristic of the operation of a "moral religion". Although absolute certainty is precluded from such a confidence, this *confidence* can nevertheless "legitimately" rely on an inference based on the observation of one's conduct:

> Without any confidence in the disposition once acquired, perseverance in it would hardly be possible. We can, however, find this confidence, without delivering ourselves to the sweetness or the anxiety of enthusiasm, by comparing our life conduct so far pursued with the resolution we once embraced. For,

[9] On the issue of the incapability of recognizing the deepest motivations of one's own actions through mere self-examination or introspection Kant is, at least, as Freudian as Freud himself: "Indeed, even a human being's inner experience of himself does not allow him so to fathom the depths of his heart as to be able to attain, through self-observation, and entirely reliable cognition of the basis of the maxims which he professes, and of their purity and stability" (R 106 / 6:63).

> take a human being who, from the time of his adoptions of the principle of the good and throughout a sufficiently long life henceforth, *has perceived the efficacy of this principles on what he does, i.e., on the conduct of his life as it steadily improves, and from that has cause to infer, but only by way of conjecture, a fundamental improvement in his disposition* […] *on the basis of what he has perceived in himself so far he can legitimately assume that his disposition is fundamentally improved.* (R 110 / 6:68; my emphasis)

It is hard to think how can this passage be at all compatible with the radical external and internal opacity and invisibility of moral worth as inscribed in the depths of will (and there alone), an invisibility that has otherwise been stated by Kant in such a radical manner. For example: "it is absolutely impossible to settle with complete certainty through experience whether there is even a single case in which the maxim of an otherwise dutiful action has rested solely on moral grounds" (G 23 / 4:407). This radical invisibility predicates precisely as inescapably *illegitimate* any conclusion about the moral disposition that starts from the mere "perception" of what one or anybody else does, i.e., from the "perception" of certain empirically observable course of conduct or set of actions. It seems that the transparency (however partial) of moral worth that this confidence distinctive of a "moral religion" implies, can only arise if the mere conformity to the moral law(s), or what Kant in other passages denigrates as a mere compliance to "the letter of the law", acquires in itself the status of a positive and valid criteria of moral value. Only if *what one does* becomes intrinsically (i.e., in itself) valid independently of the inner disposition (*the how*) of that doing, can the inference from a "good conduct" to the inner maxim acquire any legitimacy. But the very assumption that the "what" of human conduct is morally worthy independently of the inner "how", has been defined by Kant as the unequivocally distinctive sign of a morally unworthy attitude (R 84 / 6:37; quoted above)[10]. In other

[10] It will certainly be worthwhile to articulate a connection between the most radical formulations in Kant's moral theory of the rift between "the what" and "the how" of morality, and Kierkegaard / Climacus's famous formulation of the distinction between "the what" and "the how" in the relation to God. Echoing Kant, Kierkegaard places all the worth in "the how": "one [person] prays in truth to God although he is worshiping an idol; the other prays in untruth to the 'true' God and is therefore in truth worshiping an idol" (Postscriptum VII 168). Kierkegaard's claim renders the "what" (what God?) completely irrelevant, and so precludes as inconsequential any distinction between a "true" and a "false" religion on the basis of the content assigned to "the what". In the same way, Kant's moral theory in its most radical formulations renders the "what" of moral conduct irrelevant, placing all the weight on the inner "how". Thus, these formulations preclude any distinction between a "true" and a "false" morality based on certain specific content assigned to the "what" of moral conduct. Nonetheless, Kant retreats from the radical consequences of his moral theory.

words, the confidence on moral improvement that alone is capable of establishing the distinction between the "moral religion" from the "bad" ones is only possible if moral worth is no longer invisible. But Kant's understanding of the moral problem asserts precisely that if "moral worth" is regarded as something visible it is no longer neither "moral" nor "worthy".

We have attempted throughout this paper to emphasize how Kant's line of argument in *Die Religion* is marked by a profound tension and ambivalence: on the one hand, the delineation of a fracture and incommensurability between the visibility of human conduct and the invisibility of the will which is decisive in Kant's analysis of the moral problem, and which alone enables him to effect the revolutionary displacement of the center of gravity of moral worth to a hidden and inscrutable inflexion of the will; and on the other hand, a certain necessity to repair this fracture, to replace the incommensurability between the visible and the invisible, the observable and the hidden, with some kind of continuity and commensurability; the necessity of substituting the radical invisibility and inaccessibility of moral worth with some degree of transparency; the necessity, in sum, to make *good* and *evil* somehow visible. It is this necessity what we have tried to interrogate throughout this paper by showing: i) how it leads Kant's analysis of radical evil to an *aporia*; ii) how it leads him, in the *Groundwork*, to foreclose a question that is in principle consistent with his radical and revolutionary reformulation of the moral problem; and finally, iii.) by showing how the "hidden springs" of this necessity might well be explained by the manner in which this reversal or withdrawal from a radical invisibility towards an urgently needed transparency of moral worth, is the necessary condition for telling in a certain way the history of the world. One should not be surprised that religion (in its internal tearing between a "good one" and a "bad one") is the main character in this story of history, even in the case of Kant. This privileged role of the distinction between a "true" and a "false" religion in the construction of a narrative of the history of the world is a common trait that marks the emergence of the Philosophy of Religion in modern western philosophy. But, with the appearance of this main character in the late Kantian text examined here, one could at least be alerted against the facile opposition between religion and reason that is sometimes hastily understood to be the crux of the legacy of the Enlightenment and modernity; and, thus, one could be alerted against a certain religion that might well hide itself under the appearance of "pure reason" in a manner, at least structurally analogous, to that in which

The distinction between a "moral (true) religion" and "immoral (untrue) religion(s)" is the clearest indication of this retreat.

radical evil, according to Kant, hides itself under the appearance of virtue. On the other hand, if one notes how this religion hidden behind "pure reason", amounts to the surreptitious transformation of Kant's anti-economical *respect* for the law, anti-economical in the way in which it interrupts the means-end logic of instrumental rationality, into the law of the economy (i.e., into history) and the entire system of hegemonies and exclusions which such a "law" administers, then one might want to attempt to cling to this anti-economical side of Kant's pure practical reason, but running then the inevitable risk of confronting the strange structural affinity that this pure goodness might have with certain anti-economical manifestations of "evil".

Bibliography

Allison, Henry E. *Kant's theory of freedom*. Cambridge, NY: Cambridge University Press, 1990.

Bernstein, Richard J. "Radical Evil: Kant at war with himself", *Rethinking Evil*. Ed. by Maria Pia Lara. Berkeley: University of California Press, 2001.

Kant, Immanuel. [R]. *Die Religion Innerhalb Der Grenzen Der Blossen Vernunft*. Hamburg: Felix Meiner, 1966. Trans. by Allen W. Wood under the title "Religion Within the Limits of Reason Alone"; *Religion and Rational Theology*. Cambridge University Press, NY, 1996.

_____. [G]. *Grundlegung Zur Methaphysik der Sitten*. Berlin: Elibron Classics, 2005. Trans. by Allen Wood under the title *Groundwork for the Metaphysics of Morals*. Yale University Press, New Haven, 2002.

_____. *Kritik der Praktischen Vernunft*. Leipzig: Felix Meiner, 1974. Trans. by Mary J. Gregor under the title "Critique of Practical Reason", *Practical Philosophy*. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996.

Kierkegaard, Sören. *Concluding Unscientific Postscriptum to the Philosophical Fragments*. Trans. by Edna and Howard Hong. N.J: Princeton University Press, 1992.

Artículo recibido: 13 del Agosto 2007; aprobado: 14 de Septiembre del 2007