

2014

## Transcendental Idealism as the Backdrop for Kant's Theory of Religion

Stephen R. Palmquist

*Department of Religion and Philosophy, Hong Kong Baptist University, [stevepq@hkbu.edu.hk](mailto:stevepq@hkbu.edu.hk)*

Follow this and additional works at: [https://repository.hkbu.edu.hk/hkbu\\_staff\\_publication](https://repository.hkbu.edu.hk/hkbu_staff_publication)



Part of the [Philosophy Commons](#)

This document is the authors' final version of the published article.

Link to published article: [https://doi.org/10.1007/978-1-137-33475-6\\_8](https://doi.org/10.1007/978-1-137-33475-6_8)

---

### APA Citation

Palmquist, S. (2014). Transcendental Idealism as the Backdrop for Kant's Theory of Religion. *The Palgrave Handbook of German Idealism*, 144-164. [https://doi.org/10.1007/978-1-137-33475-6\\_8](https://doi.org/10.1007/978-1-137-33475-6_8)

This Book Chapter is brought to you for free and open access by HKBU Institutional Repository. It has been accepted for inclusion in HKBU Staff Publication by an authorized administrator of HKBU Institutional Repository. For more information, please contact [repository@hkbu.edu.hk](mailto:repository@hkbu.edu.hk).

## Transcendental Idealism as the Backdrop for Kant's Theory of Religion

*Stephen R. Palmquist*

### *The partially hidden influence of transcendental idealism on Kant's "Religion"*

Recent work on Kant's *Religion within the Bounds of Bare Reason*<sup>1</sup> frequently begins by calling attention to the many unresolved interpretive problems that plague the careful reader of this influential yet perplexing book.<sup>2</sup> Because the characteristic features of transcendental idealism do not play an obviously constitutive role in the exposition of Kant's arguments in this text, for many years most Kant scholars simply passed it over as an optional appendix to his ethics, if not an outright aberration.<sup>3</sup> A central focus of the so-called "affirmative" trend in interpreting *Religion* that has come to the fore over the past 20 years<sup>4</sup> has therefore been to argue that these conundrums in Kant's text are largely resolvable. The present chapter will draw from that recent literature an account of how four key components of Kant's argument in *Religion*, often regarded as either incoherent or mistaken, actually make good sense if they are read against the backdrop of his transcendental idealism.

Part of the responsibility for the aforementioned tendencies, both that most twentieth-century Kant-scholars treated *Religion* as if it were irrelevant to the central concerns of Critical philosophy and that more recent interpreters focus so much attention on its apparently contradictory claims, rests squarely with Kant. For in the prefaces to the book's two editions Kant confesses motivations that would have pulled him in two conflicting directions as he contemplated what to write. In the first (1793) preface, he expresses the hope that this book might eventually come to be used as a "guideline" (presumably, therefore, as a *textbook*) for courses on philosophy of religion that theology students would be required to take before concluding their studies (Rel 6:10). In the second preface, by contrast, he responds to an anonymous reviewer of the first edition<sup>5</sup> by claiming that the book was *not* (as the reviewer had claimed) meant to be comprehensible only to those who are adept in the intricacies of the Critical system. The essential points

defended in *Religion* are “readily understandable,” Kant claims, because they are “contained, even if in different words, in the most popular instruction for children or in sermons” (Rel 6:14). Although this passage suggests he attempted to write the book *without* depending on the framework provided by his Critical philosophy, Kant admits that some occasional references to his technical terminology were necessary, though “only because of the school” (i.e., to please scholars). As an example, he cites the distinction between virtue considered as an action’s phenomenal “legality” (i.e., as a “duty-tempered” skill) and virtue considered as noumenal “morality” (i.e., “as a steadfast conviction”<sup>6</sup>) (Rel 6:14). Keeping in mind Kant’s self-confessed twofold concern, pleasing the schools and making his theory comprehensible to the general reader, I shall devote the remainder of this section to a broad overview of the key features of Kant’s transcendental idealism. This will set the context for my attempt, in the remaining sections of this chapter, to highlight the deep dependence of Kant’s key arguments in *Religion* on the transcendental backdrop provided by the Critical philosophy.

In Kant’s hands the term “idealism” takes on a unique connotation that is bound to be misunderstood by anyone who assumes he is following in the footsteps of Berkeley or Descartes.<sup>7</sup> For Berkeley, like Plato before him, idealism entails the belief that objects in the empirical world are not real but are merely illusions created by the human mind; the true reality of anything we identify as an object is *mental*. By contrast, Kant’s idealism is thoroughly *perspectival*: “The *transcendental* idealist,” he insists, “can be an empirical *realist*” (A370, emphasis added) – this being the only *genuine* type of realist, at that. Calling something “transcendental,” in other words, signifies for Kant a concern to identify a set of *boundary-conditions* that define necessary and universal features of a *perspective*; and this perspective is what *makes* an object “real,” whenever it meets the conditions so defined. Thus, when Kant says in the first *Critique* that space and time “exist . . . only in us” (A42/B59), he does not mean that the entire spatiotemporal world is actually a figment of our imagination and that only ideas actually exist, perhaps in a transcendent “world of forms” (as in Plato’s idealism). What he means is that space and time serve as the perspectival *boundary-conditions* that must be presupposed as having a determining role in the formation of anything we are to call an “object” – *as far as our empirical knowledge of the world* is concerned.

Any doubt that Kant’s understanding of transcendental idealism is thoroughly perspectival is laid to rest in the second *Critique*, when he takes the (for some, unexpected) step of reversing many of the

claims he had established in the first *Critique*. He insists, for example, that when practical reason (rather than theoretical reason) defines the philosopher's subject-matter, one must first examine the principles (as known directly to reason, through our awareness of the moral law), then proceed to examine their logical form, and only after that consider the role that any aesthetic/feeling-based "input" might play (see CPrR 5:16). The topics of the three chapters of the second *Critique*'s Analytic must therefore follow the reverse order of the Aesthetic, Analytic of Concepts, and Analytic of Principles in the first *Critique*. But more importantly for our purposes, Kant also insists in the second *Critique* that the *problem* transcendental idealism creates for our theoretical knowledge of the ideas of reason (God, freedom, and immortality – whose realization constitutes the whole aim of metaphysics, according to Kant) is *solved* by reason's ability to adopt the practical standpoint. Thus, immediately after claiming (CPrR 5:132–33) that the practical postulates of immortality, freedom (as a *transcendent* cause of actions), and God fill the gaps of theoretical ignorance left by the first *Critique*'s conclusions (as set out in the sections called the Paralogisms, the Antinomy, and the Ideal), Kant asks: "is our cognition really extended in this way by pure practical reason, and is what was *transcendent* for speculative reason *immanent* in practical reason?" (CPrR 5:133). Anyone who mistakenly takes Kant's transcendental idealism to imply that our metaphysical ignorance is absolute and insurmountable would surely expect a negative answer, given Kant's wording of this question. Yet Kant surprises such readers by answering: "Certainly, but *only for practical purposes*" (CPrR 5:134).

What Kant is claiming here, regarding the cognitive status of immortality, noumenal freedom, and God, is directly parallel to his claims in the first *Critique* regarding the reality of empirical objects. Even though a correct understanding of the boundary-conditions of human knowledge (i.e., an awareness of the fact that spatiotemporality is a mental condition that we place onto the empirical objects we come to know) requires us to accept that space and time are merely *in us* (and thus have no *empirical* reality in themselves), the very same fact enables us to affirm, without any fear of being outsmarted by the skeptic, that the *objects* of our knowledge are undoubtedly real and, imbedded as they are in the objective reality of space and time, really do exist *outside of us*. The same perspectival duality applies to the second *Critique*: even though a correct understanding of the boundary-conditions of human morality (i.e., an awareness of the fact that ethical obligation has its proper source in a mental condition that we place onto the practical choices we make) requires us to accept that incentives to action must be entirely *in us* (i.e., *autonomous*) in

order for the resulting action to be genuinely moral, nevertheless this same fact enables us to affirm, without any fear of being outsmarted by the atheist, that the objects of metaphysics (God, freedom, and immortality) are *realities* that we can stake our lives on, in spite of our theoretical ignorance of their precise mode of existence. Even though “no human understanding will ever fathom” their theoretical possibility, practical reason assures us that “no sophistry will ever wrest from even the commonest human being the conviction that they are not true concepts” (CPrR 5:133–34, translation modified). In other words, as Kant goes on to explain, the application of transcendental idealism to morality – the recognition that pure moral principles serve as *boundary-conditions* for all our practical choices – provides “objective reality” to the three ideas of reason (i.e., it assures us that the objects that such ideas point to are *possible*), even though “speculative reason . . . could not assure” this result for them (CPrR 5:134).

Theoretical reason alone is incapable of even *thinking* the reality of God, freedom, and immortality, because from that standpoint intuition is required in order for something to be declared real; but from the practical standpoint, these same objects acquire “significance” (CPrR 5:136) – that is, reference to a real object – enabling us to apply the categories to them in producing genuine cognition, though only for practical purposes. This argument, elaborated in §VII of the second *Critique*’s Dialectic, shows that Kant intends his moral philosophy to be part and parcel of his transcendental idealism. Indeed, he had already clarified this point toward the beginning of the Dialectic, by arguing that “the antinomy of practical reason” can be resolved in precisely the same way as “the antinomy of pure speculative reason” – namely, through an appeal to the distinction between “appearance” and “noumenon,” this being the key perspectival distinction that underpins his transcendental idealism (CPrR 5:114–15). Likewise, Kant appeals to the crucial role this distinction plays in the argument of the second *Critique* (CPrR 5:104–6): without it, Kant’s unshakable faith in freedom would crumble.

Kant wrote the second *Critique* in 1788, shortly after publishing the second edition of the first *Critique* in 1787; so in concluding this initial overview of transcendental idealism, it is relevant to recall that the second preface to the first *Critique* introduces a new metaphor that explicitly focuses on the perspectival character of transcendental idealism. His revolutionary approach to philosophy, Kant there claims, is based on a Copernican “hypothesis” (Bxxii, note), which he initially describes as follows:

Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them *a priori* through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try [*versuchen*] whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition . . . (Bxvi)

The fundamental hypothesis that guides transcendental idealism, therefore, is that the philosopher's task is to locate *a priori* boundary-conditions that the mind places onto empirical objects in order to make them knowable. With this basic definition in mind, let us examine in the remainder of this chapter how four of Kant's central arguments in *Religion* are bound to seem perplexing (if not utterly incoherent) until we read them against the backdrop of this perspectival distinction between empirical knowledge and the transcendently ideal conditions that we must presuppose in order to explain the possibility of the former. As we shall see, Kant's arguments turn out to carry considerable force once we interpret them in this light.

### *Evil as the transcendental problem that religion solves*

Perspectival distinctions abound in the text of Kant's *Religion*. One need go no further than the first paragraph to find a prime example: Kant's opening claim, that we can cognize our moral duty without making any reference to God ("the idea of another being above [us]") or happiness ("an incentive other than the [moral] law itself"), gives many readers the impression that Kant is reducing religion to morality, especially when he goes on to conclude that "on its own behalf morality in no way needs religion" (Rel 6:3). Interpreting this claim against the backdrop of the transcendental ideality of space and time in the first *Critique* and of the moral law in the second *Critique*, however, enables us to recognize that Kant's point is actually perspectival: from the transcendental perspective, morality is justifiable independently of all empirical religious traditions, even though it may well turn out (as the careful reader of *Religion* eventually discovers) that from the empirical perspective religion provides a necessary means of *propagating* morality universally. Indeed, precisely this dual perspective ends up being the focus of the second half of the first preface, where Kant contrasts "philosophical theology" (i.e., theology that takes morality as its core concern) with "biblical theology" (i.e., theology that takes some scripture-based empirical religion as its core concern) (Rel 6:9). There Kant argues not that philosophical theologians have *no need* of biblical

theologians, nor (certainly!) vice versa, but rather that these two distinct types of theologian should “be at one,” as far as their ultimate aims are concerned (Rel 6:10).

Any doubt that Kant’s justification for this claim rests squarely on his transcendental idealism is dispelled in the second preface, where Kant distinguishes even more explicitly between the same two perspectives by comparing their respective concerns to two “concentric” circles, whose relation gives rise to the two “experiments” (or “attempts [*Versuchen*]”) that Kant conducts throughout the pages of *Religion* (Rel 6:12). The first experiment (i.e., the attempt to identify a system of “religion within the bounds of bare reason”) corresponds to the task of proving the transcendental ideality of space and time in the first *Critique* and the pure ideality of the moral law in the second *Critique*. Likewise, the second experiment (i.e., the attempt to ascertain whether or not the empirical elements of a particular historical religious tradition – and here Kant takes Christianity as his test case – succeed in preserving moral religion as their rational core) corresponds to the first *Critique*’s task of showing that empirical objects really do exist in time and space and to the second *Critique*’s task of showing that moral judgments really do need to exhibit respect for the moral law in order to be called “good.” This much, however, is not the source of significant controversy among interpreters of Kant’s *Religion*: no commentators, prior to a recent (and fairly minor) debate,<sup>8</sup> have regarded Kant’s distinctions between philosophical and biblical theology or the first and second experiments as the source of apparently self-contradictory perplexities.

When we come to Kant’s arguments in the First Piece,<sup>9</sup> this situation changes dramatically: his account of human nature, as consisting of a threefold “predisposition [*Anlage*]” that is inevitably corrupted by a “propensity to evil [*Hang zum Böse*],” has been the source of countless expressions of frustration (Rel 6:26–32).<sup>10</sup> Among the various inconsistencies that commentators claim to detect in Kant’s argument, by far the most intractable has been the fact that he claims that an *a priori* proof of the propensity to evil is required, yet he never seems to provide such a proof in the text. As I have previously proposed a detailed solution to this problem (see note 10), I shall not rehearse my full argument at this point. Instead, I shall focus my overview on how and why this problem persists only as long as interpreters fail to read *Religion* against the backdrop of Kant’s transcendental idealism.

When we *do* acknowledge this backdrop, Kant's repeated references to the "possibility" of evil immediately stand out as the central focus of his argument, for Kant's special form of "transcendental argument" always functions as an attempt to identify the necessary conditions for the *possibility* of experience (or of whatever subject matter may be under investigation) – these being the *boundary-conditions* that make empirical knowledge (or moral action, etc.) what it is. Read in this light, the First Piece is not Kant's attempt to demonstrate that human beings *really are* (empirically) evil. Even his impressive list of the "long melancholy litany of charges against humanity" (Rel 6:33), which so many readers mistakenly take as Kant's excuse for *not* providing the required *a priori* proof, is not intended to *demonstrate* that human beings are evil (any more than the arguments in the first *Critique* demonstrate that human beings *have* knowledge, or those in the second *Critique*, that we *are* morally good). Kant's list of evils does suggest, of course, that he never seriously doubted the empirical reality of evil. However, as a transcendental idealist, his core (philosophical) question is not "Are the empirical choices made by human beings evil?" but rather, "What must hold true for our minds (i.e., what must we presuppose as transcendently ideal) *if* empirical evil is to be possible at all?" This, I have argued, is the focus of Kant's arguments for the propensity to evil in the First Piece. The basic steps of his argument are difficult to detect because they are so obvious, being "hidden" in the very headings of the subsections of the First Piece. In a nutshell, his argument is that any being with a predisposition to good (Section I) would need to *presuppose* the existence of a propensity to evil in order to explain how any evil action (or choice)<sup>11</sup> would be *possible* (Section II); we *do* observe evil actions (or choices) in the world (Section III); so human beings must possess this propensity that therefore makes them radically evil (Section IV).

Once we recognize the perspectival focus of Kant's theory of the propensity to evil, whereby his only defense is a *transcendental* argument that focuses on the *ideality* of this principle (i.e., its status as a boundary-condition for religious experience – a component of our *noumenal* character that must be *presupposed* if we are to understand how evil is possible and how religion arises therefrom), all of Kant's (apparently conflicting) statements about the need for an *a priori* (or "formal") proof can be rendered self-consistent. Thus, as he states in the opening section of the First Piece:

In order to call a human being evil . . . one would have to be able to infer *a priori* from a few consciously evil actions, or indeed from a single one, an evil maxim lying at their basis, and from

it again a basis, itself in turn a maxim [and] lying in the subject universally, of all particular morally evil maxims. (Rel 6:20)

*A priori* inferences that lead to a universal conclusion are the task of Kant's transcendental arguments. When he goes on to say, while introducing the examples he lists in Section III (entitled "The Human Being is Evil by Nature"), that "we can spare ourselves the formal proof" (Rel 6:32–33), he is not announcing that he has *given up* the task of providing the required proof. Rather, his point is the same as the equivalent step in his other transcendental arguments: just as even Hume did not doubt that "we have experience" of empirical objects – the only question being to identify what that experience actually *is* – so also nobody seriously doubts that human beings sometimes commit evil acts.

The confirmation of the empirical reality of the subject matter under consideration is an *uncontroversial* step in Kant's transcendental arguments; so his point in Section III of the First Piece is only that this premise of his argument is secure, as the examples he provides poignantly illustrate. That is, *from the empirical perspective*, nobody seriously doubts the reality of evil. (Precisely the same thing can be said about the spatiotemporal nature of empirical objects and about the moral status of human choices: ordinary, non-philosophical persons feel no need whatsoever to provide formal proofs of the reality of what they can see before their own eyes.) Thus, after completing his list of illustrations, Kant goes on to insist that, "even if the existence of this propensity to evil can be established through experiential proofs," this still leaves open the question of "the actual make-up of that propensity and the basis of" the empirical opposition to the moral law that the examples illustrate (Rel 6:35). In other words, just because this force within us, opposing our good predisposition, is "real in time" (i.e., *empirically* real), this alone does not determine whether the *status* of our propensity is empirical or transcendental. For the latter status to be established, the propensity to evil "must be cognized *a priori* from the concept of evil, as far as it is possible according to laws of freedom" – a task whose "development" Kant says he will now proceed to clarify (Rel 6:35).

Once Kant has restated his theory, in an attempt to clarify what he means by "the propensity to evil," Section III ends with a footnote (attached to a paraphrase of Romans 3:9–12, affirming the universality of sin) that begins:

The actual proof of this judgment of condemnation pronounced by [the court of] morally sentencing reason is contained not in this section but in the previous one. This section contains only the confirmation of the judgment through experience; but experience can never uncover the root of evil in the supreme maxim of free volition in reference to the law, the root which, as an *intelligible deed*, precedes all experience. (Rel 6:39n)

This passage unambiguously confirms that Kant does think he has provided the required proof. Yet it will surprise anyone who reads the previously quoted statement as indicating that, at the midpoint of Section III (Rel 6:35), the proof had not yet been presented. This, however, is not what Kant wrote in that earlier passage. Rather, he stated that, having *confirmed* the reality of evil by reference to various examples that illustrate the transcendental condition under consideration in the First Piece, he would complete Section III by clarifying the *apriority* of the proved concept. What he now explains, as a postscript to the completed argument, is that, with Section I having provided the argument's necessary precondition, the predisposition to good, Section II argued that human beings *can* be evil only if our nature has a necessary and universal propensity *to* such evil (i.e., only if this propensity serves as a transcendental boundary-condition for evil), and Section III then confirmed the empirical reality of that otherwise merely hypothetical situation by showing how obvious it is that evil really exists. The conclusion of the overall argument, therefore, must be that the propensity to evil is not *itself* an empirically real constituent of our nature, but is a *transcendentally ideal* boundary-condition that makes evil actions *possible*. Indeed, this is precisely what we find in Section IV (minus the baggage of the transcendental-empirical distinction): Kant explains the implications of his argument for the proper interpretation of the Christian doctrine of original sin, this being his first major application of the second experiment in *Religion*.<sup>12</sup>

Once we recognize that Kant's whole argument in the First Piece is a (partially hidden) application of transcendental idealism to the question of why human beings are religious, the various confusing features of Kant's exposition fall into place. By contrast, if an interpreter takes Kant's attempt to prove that a propensity to evil exists in human nature to be equivalent to proving the *empirical reality* of the Christian doctrine of original sin (a theological concept that belongs to Kant's second experiment), then perplexity abounds.<sup>13</sup> If the backdrop for this claim is Kant's assumption that all genuine philosophy takes the form of a system of perspectives, whereby some transcendently ideal features always serve as (mentally-imposed)

boundary-conditions for whatever empirically real features may be under consideration, then his claim makes good sense. For, having completed his (first experiment) argument that human beings possess an evil propensity only insofar as we find it necessary to *presuppose* such a transcendental feature as present at the ideal boundary of our character, he naturally goes on to *criticize* the Christian doctrine of original sin. His criticism, like the foregoing argument itself, is two-sided: *if* Christians interpret the doctrine of original sin as a *transcendentally ideal* feature of human nature, then the doctrine is fully in conformity with the religion of bare reason; however, if Christians interpret the doctrine to mean that *original sin* is *empirically real* in a form that can be *inherited* from one human being (such as Adam) to the next, then this doctrine is “a contradiction” and not worthy of serious consideration by any theologically minded philosopher (Rel 6:40).

### *Two perspectives on conversion: Kant’s ethics of grace*

Next to Kant’s theory of evil, the aspect of *Religion* that interpreters have most often regarded as a perplexing source of apparently self-contradictory claims is his theory of the “change of heart” that converts a person’s moral character from evil to good. Kant develops his theory primarily in Part One of the Second Piece, after having offered a lengthy preview in the General Comment to the First Piece. Much debate has focused on whether Kant leaves any room for *grace* in his philosophical theology, and if so, whether his theory makes sense or is merely a poorly digested jumble of unorganized assertions. Many interpreters have taken it as granted that Kant’s philosophy in general, and his philosophy of religion in particular, leave no room for grace – or at least, no grace that bears any resemblance to the Christian doctrine of vicarious atonement. Yet, as we shall see, once we read *Religion* against the backdrop of transcendental idealism, a revitalized theory of grace turns out to be Kant’s second major application of his second experiment.

The first of several ambiguities that plague the interpreter of Kant’s theory of conversion and grace is that at times he appears to affirm the possibility (if not even perhaps the necessity) of divine assistance, in order to explain how we can counteract the apparently insurmountable influence of radical evil and enter once again upon the path of goodness, while at other times he writes as if everything depends on human effort alone. Those who highlight the latter tendency typically interpret Kant as a moral

reductionist, who lacks any serious belief in religion as anything other than morality in disguise, while those who highlight the former tendency, adopting an “affirmative approach” to interpreting Kant’s theory of religion (see note 4), typically see him as seeking not to destroy but to reform empirical religious traditions.<sup>14</sup> This interpretive conflict parallels the debate over whether the Kant of the first *Critique* is best regarded as the “all-destroyer” of metaphysics<sup>15</sup> or as attempting to set metaphysics on its true and proper path. Whereas very few serious Kant scholars nowadays adopt the former position with respect to Kant’s views on metaphysics (i.e., his transcendental idealism is now widely recognized as an attempt to *reform* metaphysics, not to destroy it), the reductionist interpretation of Kant’s theory of religion still attracts considerable support. What we find, when we come to consider Kant’s theory of conversion and grace, is that the only way to make sense of the fact that two such varied interpretations are still prevalent (especially for such a significant figure as Kant) is to interpret his intentions as perspectival.

As I have argued elsewhere in greater detail, Kant’s theory of grace is so frequently misunderstood because readers expect him to be providing a *theology* of grace – that is, a concrete explanation of *how* divine assistance happens (if it happens) or of why it does *not* happen (if it does not). Reductionist interpreters rightly recognize that to provide such a theology would be to overstep the epistemological boundary-conditions set by the first *Critique*, where Kant argued that metaphysics can say nothing positive about ideas of reason that have no intuitive content – God, freedom, and immortality being the three key metaphysical examples. They therefore simply assume that, whatever Kant may be aiming to do in *Religion*, he *cannot* be telling us that a real God who exists outside of human beings actually *assists* us to become good. That is, there is no room in Kant’s philosophy for a *theology* of grace *as such*. To this extent, reductionist interpreters are certainly correct. However, what Kant actually presents in *Religion*, especially in the Second Piece and the main portion of the General Comment to the First Piece,<sup>16</sup> is what I call an *ethics of grace*.<sup>17</sup> Because his focus here (and throughout *Religion*) is (admittedly) on ethics, the reductionist interpretation has a prima facie plausibility. However, the *aim* of Kant’s ethical focus is not to pontificate on how unenlightened it is to believe in divine assistance (and all the baggage that typically comes with it – different for each empirical religious tradition); rather, his focus is to guide those who *do* hold to such an empirical belief (i.e., any religious believer who already believes God has provided some form of supernatural assistance) to *interpret* their belief in a way that enables them to maintain an ethically

good lifestyle. The danger of affirming a positive theology of grace is that it has a tendency to make people *morally lazy*: if God has saved me by providing what I lack, then why should I even *try* to do what I obviously cannot do under my own power? In the process of explaining how it is possible to believe in divine grace without thereby making oneself morally lazy, Kant offers some valuable theological reflections on the nature of conversion and grace. But his theory focuses on what Kant refers to in the first *Critique* as “a noumenon in the **negative** sense,” as distinguished from “the noumenon in a **positive** sense” (B307): that is, grace can be properly understood only when we interpret it against the backdrop of the limitations of human knowledge that come along with an affirmation of transcendental idealism.

According to Kant’s theoretical philosophy, the essence of human nature (and of each individual human being), like the true reality of God, lies hidden in the transcendent realm that is somehow mysteriously grounded in what he infamously calls “the thing in itself.”<sup>18</sup> The distinction between the thing in itself and things as they appear to us, together with the corresponding (though not identical) distinction between phenomena and noumena, forms the backbone of Kant’s transcendental idealism. Or, to be more precise, the former distinction defines the basic framework for transcendental idealism (i.e., the status of “objects” within the human mind), while the latter defines the corresponding framework for empirical realism (i.e., the status of determinate objects that we can come to know as, presumably, existing outside the mind [see note 18]). Kant’s position, carried over from the first *Critique* directly into *Religion*, is that, whereas the transcendent (noumenal) realm (grounded in the thing in itself) is theoretically unknowable for human beings, we must (primarily for moral reasons, but also for reasons relating to the integrity of scientific knowledge) *believe* that this transcendent realm is thoroughly intertwined with the empirical (phenomenal) realm (consisting of things as they appear to our minds). Again, this is the essence of transcendental idealism, and any reader of *Religion* who does not accept this as Kant’s presupposition is bound to find the text filled with irresolvable perplexities.

The perplexities become more tolerable once we realize that Kant’s theory of the role of grace in conversion defends two distinct (yet compatible) perspectives on one and the same subject-matter. The focal question of *Religion*’s Second Piece is: How is it *possible* for a radically evil human being to become good? From the transcendental perspective, Kant argues that the only way to conceive of this possibility is to presuppose that a *revolution* has taken place in the deepest part of our moral character, our *Gesinnung* –

a technical term that is difficult if not impossible to translate into a single English word, though I have argued that “conviction” is the best alternative (see note 6). To be evil, for Kant, is to have freely chosen to adopt, as our deepest *conviction* regarding how we ought to make moral choices, the principle that our own interests (i.e., the fulfilment of any personal goal that we believe will make us *happy*) are more important than obeying the moral law; if the two happen to agree, our *actions* will turn out to be “legally” good, even though our *character* remains evil. Only if we experience a radical conversion in the depths of our heart, so that our conviction prioritizes the moral principle even if doing good will require us to sacrifice our own happiness, can we think of ourselves as genuinely “good.” The *problem* that (in concert with the presence of the propensity to evil) gives rise to the need for empirical religion is that *our convictions are largely inscrutable to us* (see, e.g., Rel 6:20, 51); we can to some extent guess what they are and *hope* that our deepest conviction is consistent with “the good principle,” but only God would be in a position to know for certain. The key to Kant’s ethics of grace, therefore, is to insist that *if* one believes that God has assisted in this revolution of the heart coming about, then this belief must be held as an expression of *hope*, not as a dogmatic claim to know the mind of God. In other words, Kant does not deny the possibility of divine assistance; he merely insists that *if* God assists us, the philosophical status of such assistance must be *transcendentally ideal* – that is, an assistance that occurs through the cooperation of our own mental presuppositions. That such assistance (i.e., *some* form of grace) must be *possible*, he further argues, is assured, since there is no other way to conceive of how evil human beings can become good, as the moral law requires.

The second major feature of Kant’s theory of conversion is that, because of our theoretical ignorance of our own (noumenal) character, we must look to our *actions* as (hopefully) accurate expressions of our deepest conviction; for the only option we have for gauging the morality of our character is to infer it from any *evidence* we can detect of reformation in our moral development over a lengthy period of time. Having started out under the influence of radical evil, our actions (taken as a whole) can never be perfect, so the most we can expect from this *empirical* perspective on our moral status is *gradual reform*. Kant insists that from *this* perspective, each person is individually responsible for his or her own moral reformation. This guards against any temptation we may have (due to the residual influence of the propensity to evil even on a converted person’s moral conviction) to believe that the goodness of our

*actions* is irrelevant to the goodness of our *character*. However, anyone who reads *Religion* against the backdrop of transcendental idealism must recognize that our own moral efforts represent only one side of the coin – and *not* the side that Kant is most interested in, as a philosopher. From the empirical perspective, a religious tradition is healthy only if it succeeds in motivating its members to continue striving to be good, despite the inevitable failures that we human beings will encounter along the way. Yet if an empirical religious tradition keeps moral religion (i.e., the religion of the first experiment) at its core, then the doctrine of grace provided by it (not by Kant himself!) can complement and even fulfill the purposes of rational religion.

What is often ignored by interpreters who focus only on the ethical/empirical side of Kant's ethics of grace is that his central argument in the Second Piece also has a transcendental side: given our radically evil starting point, the transcendental (grace-oriented) aspect of Kant's argument ensures that we must appeal to *some belief* that will empower us to do what would otherwise be (or at least seem) impossible: obey the moral law. Subsection A of Part I (of the Second Piece) thus introduces an "ideal of moral perfection" that he calls "the archetype of the moral conviction in all its integrity" (Rel 6:61), using language borrowed from the New Testament to describe it as a foreign influence that "has *come down* to us from heaven," as a "divine decree" – a "word" that bids us "*Become!*" (Rel 6:60). "Practical faith" in this idea is the second transcendental condition that makes empirical religion possible. Without it, we could not even conceive of how gradual reform could take place; but *with* it (as Kant goes on to argue in Subsection B) we must be careful to remember that our observation of gradual reform is our only available *evidence* that the noumenal revolution has actually occurred – that is, that God's grace has actually empowered us to become good. In other words, we can never actually *know* if God has assisted us; but we *can* know that our belief has motivated us to improve.

Kant argues that "the objective reality" of this archetype hinges on each person making himself or herself an *empirical example* of the transcendental goodness it conveys (Rel 6:62–66). For this reason, we cannot depend on the example of some great moral hero to save us vicariously; indeed, we could only recognize the goodness in a person such as Jesus or Buddha because this archetype serves as the necessary condition that shapes our understanding of goodness. Kant's position, then, is not that Jesus or Buddha are not *perfect*, or in some sense even *divine*,<sup>19</sup> but that perfection is a transcendental ideal that *everyone* must

emulate; insofar as Christians and Buddhists see their figureheads as accomplishing the goal, they merely confirm that the change of heart is *empirically real* – that is, that converted religious believers are justified in hoping that their good actions are expressions of a good conviction – not that no further moral effort is needed.

Subsection C then examines three “difficulties” that threaten the objective reality of the archetype, if they cannot be resolved. Without going through the details of each problem here, it will suffice to note that in each case Kant’s solution is perspectival, relying (at least implicitly) on the phenomenon-noumenon distinction and the crucial assumption of all Critical philosophy, that what is empirically real must be grounded in something transcendently ideal. Thus, Kant solves the problem of explaining how God can save a person who still does some evil deeds (for he fully accepts that even a good-hearted person sometimes slips back into evil) by affirming that God judges the (noumenal) conviction (which, when committed to the archetype, is perfect), not the always imperfect (phenomenal) deeds; he solves the problem of whether we can ever be certain that our heart has actually changed by observing that although the noumenal character of our conviction bars us from absolute certainty, the phenomenal character of our deeds does provide us with evidence, so our improvement provides *grounds* for hope; and he solves the problem of how pre-conversion evil can be forgiven by appealing to a special form of suffering that occurs *during* the change of heart, thus providing a *bridge* between the phenomenal and the noumenal.

Kant supplements his solution to the third difficulty with some very suggestive allusions to vicarious atonement. While he does not affirm this Christian doctrine literally, neither does he dogmatically deny it. Rather, he treats it as a profound *symbol* of what must occur in everyone’s experience, in order for conversion to occur: the “new human being” must *suffer on behalf of* the “old human being” (Rel 6:74; see also Rel 6:66), and in so doing atonement for pre-conversion sin becomes *possible* without violating God’s justice. (Kant makes no claim that such atonement actually *occurs*; confirming *that* is the task of biblical theologians.) Such “suffering” is a form of noumenal causality that makes no sense outside the framework of transcendental idealism: a person’s *moral principle* has changed, and this “change of mentality” (Rel 6:73–74) *causes* the converted person to make choices that the “old man” would have regarded as painful in the extreme; yet the “new man” welcomes them as opportunities for further and further gradual reform. If we take this as a second application of Kant’s second experiment (as Kant does in Part Two of the

Second Piece), his point is quite consistent with many traditional forms of biblical theology: from God's point of view what atones for our pre-conversion sin is our faith in Jesus' saving work (cf. our moral conviction); yet the genuineness of each Christian's faith can be confirmed only if he or she actually lives like a "little Christ" (cf. pursues a reformed lifestyle).

### *Kant's Copernican hypothesis as applied to religious service in a church*

The payoff of the foregoing account of Kant's theory of religion, confirming its (mostly hidden) dependence on his transcendental idealism, comes when we consider its implications for a distinction that appears in various forms throughout *Religion* but becomes the focal point only in the Fourth Piece, in the form of a distinction between two types of religious service. His first reference to this distinction comes in the General Comment to the First Piece, where he says: "All religions . . . can be divided into [the religion] of the *pursuit of favor* (of bare ceremonial worship) and the *moral* religion, i.e., the religion of the *good lifestyle*" (Rel 6:51). Kant's comments on this distinction, both here and in the various other contexts where he mentions it, leave little doubt that he regards the latter as the true core of all religion (and thus, the topic of the first experiment), while the former is the false approach to being religious – that is, it refers to any version of the second experiment that exposes an empirical religious tradition that *fails* to preserve moral religion as its core. Kant's chief task in *Religion* is to set out the principles for reforming the first type of religion so that it includes the second. However, to draw from this fact the inference that Kant sees ceremonial worship and all other manifestations of empirical religious practices as irrelevant to, if not subversive of, true religion is a grave error – one that has been made so often by commentators precisely because the backdrop of transcendental idealism, with its thoroughly perspectival implications, tends to be neglected.

That *true* (philosophically well-grounded) religion must have a *moral* core does not render the nonmoral aspects of religion necessarily false or harmful; this would be like claiming that the transcendental ideality of space and time in Kant's theoretical philosophy proves that empirical objects are not *really* spatiotemporal. Just as the first *Critique* reaches the opposite conclusion for empirical objects, so also the goal of *Religion* is to pave the way for an *authentic* understanding of religious dogma and ritual, rather than to decry all empirical religion as worthless superstition and "delirium [*Schwärmerei*]." Indeed,

the institution in *Religion* that plays a role corresponding to that of science in the first *Critique* is probably the single most neglected technical term in Kant's entire *corpus*: the *church*.

*Religion's* Third Piece is best known as the locus for Kant's introduction of the "ethical community," as the only ultimate means for humanity to overcome the influence of radical evil. What is rarely recognized is that Kant portrays this community as (implicitly) transcendently ideal; it is, at best, an unreachable goal that we must do our best to approximate in our limited, humanly constructed communities. The bulk of Kant's argument in the Third Piece therefore focuses not on the ethical community as such, but on its empirical instantiation, in the *religious* form of a "church" – that is, a "people of God" that unites itself under a common belief in a transcendent lawgiver whose laws are *moral* (Rel 6:98–100) and thus internally manifested to "all well-meaning" people (Rel 6:152). The four basic (categorially organized) principles that Kant lists ("universality," "integrity [*Lauterkeit*]," "freedom," and "unchangeableness") are explicitly identified as principles of *church governance* (Rel 6:101–2); as such, they correspond directly to the principles of pure understanding that govern scientific knowledge in the first *Critique's* Analytic of Principles. Far from being vague and undetermined, as some commentators have alleged,<sup>20</sup> the "form and structure" of the church (Rel 6:94) is worked out in considerable detail in the second half of *Religion*, through a series of *applications* that show how the four basic principles serve to guide those who are charged with the organization of actual (empirical) religious communities. And with each application, Kant's distinction between the two approaches to being religious becomes more and more well-refined.

After §IV of Division One of the Third Piece introduces the four principles, the three concluding subsections elaborate on the church's universality, integrity, and freedom, respectively: §V clarifies how "pure religious faith" constitutes the "one true religion" even though it *always* starts out as one of the *many* types of "church faith," based on some revealed scripture (Rel 6:107); §VI clarifies how a true church can preserve its integrity only by focusing its scriptural interpretation primarily on finding a moral meaning in the text (Rel 6:110–11); and §VII clarifies how "sanctifying faith" must *start* from the principle of free volition (Rel 6:115–19), grounding the "kingdom of God" on the hope that "equality arises from true freedom" (Rel 6:122). Finally, Division Two concludes the Third Piece by elaborating on the principle of unchangeableness, appealing to Judaism and Christianity, considered not so much as historical religious

traditions but as ideal *types* of religion, to illustrate the difference between a religion that grounds obedience in *changeable* outward observances and one that grounds obedience in *unchangeable* inward conviction.

In conclusion, I shall clarify and confirm the potentially contentious claim made earlier in this chapter, that Kant grants that empirical religion may be a *necessary* means of propagating morality universally. Demonstrating this claim is one of the main tasks of the Fourth Piece. Thus, the untitled introductory section to Part One acknowledges two ways of defining religion: (1) “according to its first origin and its intrinsic possibility,” as either “natural” or “revealed” (cf. the transcendental perspective); and (2) according to what “makes it capable of *external communication*,” as either “natural” or “scholarly” (cf. the empirical perspective) (Rel 6:155). Here, as previously hinted (cf. Rel 6:84–85), he argues that even though a divine revelation is not necessary as a *rational* condition of salvation for human beings, it may be *historically* necessary in order to *awaken* people from their ignorant adherence to a nonmoral religion. He then goes on to present a summary of Jesus’ basic message in the Gospels, arguing that it can be regarded as “a complete religion,” given its deep resonance with the religion of bare reason (Rel 6:162). He states that such “a pure *rational faith*” can *also* be regarded “as a *revelation faith*” (Rel 6:163–64), even though it is of course impossible (given the limits of knowledge established in the first *Critique*) to know for certain if it was literally revealed from a divine source.

In the Fourth Piece Kant also distinguishes between direct and indirect service of God in a way that parallels the first *Critique*’s refusal to reject God, freedom, and immortality, even though the limits it establishes for human knowledge cast into question the very *possibility* of ever locating objects that would correspond to these ideas. At most, from the standpoint of theoretical reason, an affirmation of God (like freedom and immortality) can only be hypothetical. While the second *Critique*’s confirmation of these ideas as objectively real retains a *theoretically* hypothetical status, the same ideas are now also regarded as *practically* “constitutive” (see CPrR 5:135). Likewise, in *Religion*’s Fourth Piece we discover that real, empirical religion can serve the highly relevant practical function of *motivating* people to propagate moral convictions to others.<sup>21</sup> Kant cites the Christian rituals of prayer, baptism, churchgoing, and communion as specific examples of how the perspectival methodology of Critical philosophy can revolutionize the way we think about religion. Countless readers have mistaken Kant’s comments on these topics (especially in

the General Comment to the Fourth Piece) for an out-and-out rejection of all ritual. Yet once we read the text against the backdrop of Kant's distinction between transcendental idealism and empirical realism, a whole new (and more affirmative) picture arises. The (in itself transcendently ideal) *spirit* of prayer is what determines whether any verbal expression of (empirical) prayer is efficacious and thus worth preserving.<sup>22</sup> Churchgoing does not please God in itself, but is interpreted as a helpful means of spreading the moral fruits of love to the community. Baptism is no longer a tool for the church magically to control who is and is not admitted into heaven, but becomes a meaningful confirmation of its *duty* to instruct newcomers in morality. And the communion service is not a mechanical transmission of grace to any recipient, but a profound symbol of the unity and equality of all community members. Once we recognize that Kant's aim is not to destroy religion but to revitalize it by applying the insights gained from his transcendental idealism, even his approach to ritual can be acknowledged as constituting a *Copernican revolution in religion*.

---

<sup>1</sup>. For a defense of this translation of the title of Kant's *Die Religion innerhalb der Grenzen der bloßen Vernunft* (1793–94), see my “Does Kant Reduce Religion to Morality?,” *Kant-Studien*, 83, 2 (1992): 129–48; revised and reprinted as chapter 4 of my book, *Kant's Critical Religion: Volume Two of Kant's System of Perspectives* (Aldershot: Ashgate, 2000); hereafter abbreviated KCR. Werner Pluhar subsequently adopted this title for his translation of Kant's book (Indianapolis: Hackett, 2009). All quotations from Kant's *Religion* are my own, taken from my forthcoming *A Comprehensive Commentary on Kant's "Religion within the Bounds of Bare Reason"* (West Sussex: Wiley, 2014), which includes a new translation that relies heavily upon Pluhar's.

<sup>2</sup>. The most prolific interpreter whose work focuses on these interpretive tensions is Gordon E. Michalson, Jr. See, for example, his book *Fallen Freedom: Kant on Radical Evil and Moral Regeneration* (Cambridge: Cambridge University Press, 1990), 8–9, 28,

---

where he refers to various “wobbles” that plague Kant’s exposition. On the influence that Kant’s transcendental idealism had on the whole subsequent development of German Idealism, with special reference to theology, see Gary Dorrien, *Kantian Reason and Hegelian Spirit: The Idealistic Logic of Modern Theology* (West Sussex: Wiley, 2012).

<sup>3</sup>. Goethe, one of the first contemporaries to respond to Kant’s *Religion*, famously claimed that in this book Kant had “slobbered on his philosophical cloak” by treating the doctrine of original sin so seriously. For a detailed response to such caricatures, see my article, “Kant’s ‘Appropriation’ of Lampe’s God,” *Harvard Theological Review*, 85, 1 (Jan. 1992): 85–108; revised and republished as chapter 4 in KCR; see especially KCR 129n.

<sup>4</sup>. For a lively discussion of one such “affirmative” interpretation of Kant, see the symposium published in *Faith and Philosophy*, 29, 2 (Apr. 2012), focusing on the book by Nathan Jacobs and Chris L. Firestone, *In Defense of Kant’s “Religion”* (Indianapolis: Indiana University Press, 2008); hereafter abbreviated IDKR. The problem that surfaced in that discussion is that affirmative interpreters themselves seem to disagree quite strongly on which aspects of Kant’s theory are coherent, *and* on whether affirming the coherence of the theory commits one to affirming it as a viable approach to *religion*. The authors make clear in their essays that they do not *accept* Kantian religion, but seek only to affirm that the theory makes sense. Michalson’s contribution to the symposium, by contrast, affirms that, just because Kant’s argument is filled with “wobbles,” this does not make it unworthy of acceptance! Rather, he portrays such tensions as interesting traces of two incommensurable historical movements that were influencing Kant as he wrote *Religion*, with his attempts to unite them being as plausible as any union could be.

---

<sup>5</sup>. For details of this review and the inadequacies of Kant's response to it, see my article (coauthored with Steven Otterman), "The Implied Standpoint of Kant's *Religion*: An Assessment of Kant's Reply to (and an English Translation of) an Early Book Review of *Religion within the Bounds of Bare Reason*," *Kantian Review*, 18, 1 (Jan. 2013): 73–97. Kant claims to "dispose of" the reviewer's criticism in a few sentences (Rel 6:13), not admitting that the reviewer actually raises five substantive criticisms of *Religion*, some of which Kant himself responds to in new footnotes that he added to the second edition, while others he simply ignored – to his own peril, since these same perplexities continue to be raised by interpreters today.

<sup>6</sup>. Throughout this chapter (as in my *Commentary*), I translate Kant's technical term, *Gesinnung*, as "conviction." The traditional translation of Kant's *Gesinnung* is "disposition," but Pluhar's translations all use "attitude." In a forthcoming article, "What Is a Kantian *Gesinnung*? On the Priority of Volition over Metaphysics and Psychology in *Religion within the Bounds of Bare Reason*" (under review), I argue that (at least in its religious contexts) "conviction" fits the *volitional* emphasis that Kant gives this word far more appropriately than does the more metaphysical "disposition" or the more psychological "attitude."

<sup>7</sup>. Following the publication of the first edition of the first *Critique*, Kant was accused of defending a position close to that of Berkeley. In response, he added the new "Refutation of Idealism" section to the second edition (B274–79). As that section clarifies, Kant's special form of "idealism" rejects Berkeley's idealism (as well as Descartes's), inasmuch as Kant does *not* allow (much less require) us to consider *empirical* objects as ideal.

<sup>8</sup>. That my claims in this paragraph are not wholly uncontroversial is demonstrated by the aforementioned symposium on IDKR (see note 4). IDKR 114–19 identifies and locates Kant's second experiment *solely* in

---

the fourth main part of *Religion*. My article, “Cross-Examination of *In Defense of Kant’s Religion*,” *Faith and Philosophy*, 29, 2 (Apr. 2012): 170–80, argues against such an interpretation, defending the position I develop more fully in KCR (especially ch. 7–8). Unfortunately, the authors’ responses (published later in the same issue) merely sidestep the core problems with their position, choosing instead to attack several alleged “fallacies” that they read into my article. They never address the substance of my criticism, which centers on the fact that IDKR fails to consider the most viable alternative to its own position: the two experiments do not consist (as they claim) merely of the main philosophical arguments of *Religion* (in the first three parts of the book), and an *illustration* (in the fourth part), respectively; rather, as a metaphor for Kant’s core philosophical strategy of viewing transcendental idealism and empirical realism as two perspectives on one and the same subject-matter, the two experiments permeate the entire book. The debate is minor because it hinges on whether or not KCR actually defended such a position: although my symposium article points directly to the lengthy account of the two experiments in KCR, the authors’ *only* substantive response was that their (2008) book was justified in neglecting the alternative presented in my (2000) book because my account of the two experiments *had not yet been published* prior to the appearance of IDKR.

<sup>9</sup>. I follow Pluhar in translating *Stück* (the word Kant uses to mark the heading of each major division in *Religion*) literally as “Piece” (as when it refers to an essay or journal article), rather than loosely as “Book” (as in Greene and Hudson’s 1934 translation) or “Part” (as in di Giovanni’s 1998 translation for the *Cambridge Edition*).

<sup>10</sup>. For a summary of some of these claims regarding the incoherence of Kant’s argument in the First Piece, see my article, “Kant’s Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Assistance,” *The Journal of Religion*, 90, 4 (Oct. 2010): 530–53. Much of my exposition in this section is a summary of this article.

<sup>11</sup>. Kant’s official moral theory relates primarily to choices and only secondarily (if at all) to phenomenal actions. However, the careful reader of *Religion* quickly detects that here Kant repeatedly reverses this previous tendency, viewing choices as beyond the reach of human observation and therefore treating evil as something that we can primarily (and in some contexts, *only*) detect by observing peoples’ actions.

---

<sup>12</sup>. I argue this point in detail in KCR, ch. 7, pt. 2 – a point Firestone and Jacobs totally ignore, both in IDKR and in their follow-up attempt to defend their claim that the second experiment operates *only* in the Fourth Piece (cf. note 8, above).

<sup>13</sup>. Firestone and Jacobs call this theory of the relation between Kant’s two experiments the “*Religion-as-Translation thesis*” (IDKR 69–81). It is the view that the first experiment just *is* Kant’s moral philosophy, while the second experiment is Kant’s attempt to translate the key theories of Christian theology into the terms of his moral philosophy (IDKR 104). They take John E. Hare’s book, *The Moral Gap: Kantian Ethics, Human Limits, and God’s Assistance* (Oxford: Oxford University Press, 1996), as a key example of this approach. As such, it is not surprising that Hare ends up regarding some of Kant’s main arguments as a “failure” (*The Moral Gap*, 60–62).

<sup>14</sup>. For the fullest account of the distinction between traditional and affirmative interpretations of Kant, see Chris L. Firestone and Stephen R. Palmquist, “Editors’ Introduction,” in *Kant and the New Philosophy of Religion* (Indianapolis: Indiana University Press, 2006), 1–39. For some important qualifications on the use of the term “affirmative,” see my article, “To Tell the Truth on Kant and Christianity: Will the Real Affirmative Interpreter Please Stand Up?,” *Faith and Philosophy*, 29, 3 (July 2012): 340–46. See also KCR, ch. 6, and note 1, above.

<sup>15</sup>. The tradition of interpreting Kant’s Critical philosophy as destroying the possibility of metaphysics (and so also, theology) can be traced back to Moses Mendelssohn’s early response to the first *Critique*. But the most frequently quoted source is Heinrich Heine, who referred to Kant as “the arch-destroyer in the realm of thought,” who put forward “destructive, world-annihilating thoughts” (*Religion and Philosophy in Germany* [1834], trans. John Snodgrass [Boston: Beacon, 1959], 109). I provide a detailed account of and response to this tradition in KCR, ch. 1.

<sup>16</sup>. In the first edition of *Religion*, the first Piece had no General Comment; the material renamed under this heading in the second edition was called section “V” in the first edition. The second edition added a lengthy paragraph introducing four “parerga” and relating them to the General Comments of each Piece. The portion of text I am referring to here is part that is *common* to both editions. Technically, the first parergon refers only to the newly added portion of the first General Comment; the foregoing portion (the former §5) constitutes Kant’s *solution* to the parergon.

---

<sup>17</sup>. For a fuller version of the argument presented in this section, see my article, “Kant’s Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Assistance,” *The Journal of Religion*, 90, 4 (Oct. 2010): 530–53.

<sup>18</sup>. Kant’s theory of the thing in itself is notorious for its interpretive difficulties, not the least of which is that he portrays it as the ultimate source of everything real, yet also insists it is by definition unknowable. I propose solutions to several of the most intractable of these difficulties in my first two published articles: “Faith as Kant’s Key to the Justification of Transcendental Reflection,” *Heythrop Journal*, 25, 4 (Oct. 1984): 442–55; and “The Radical Unknowability of Kant’s ‘Thing in Itself,’” *Cogito*, 3, 2 (Mar. 1985): 101–15. Another nest of problems concerns how “thing in itself” relates to the various terms Kant employs for the object of knowledge at various stages in its determination. For solutions to these problems, see my “Six Perspectives on the Object in Kant’s Theory of Knowledge,” *Dialectica*, 40, 2 (1986): 121–51, where I argue that “thing in itself” and “noumenon” are *not* mere synonyms, as many interpreters take them to be. Revised versions of these three articles became chapter 5, appendix 5, and chapter 6, respectively, of my book, *Kant’s System of Perspectives: An Architectonic Interpretation of the Critical Philosophy* (Lanham, Md.: University Press of America, 1993).

<sup>19</sup>. Contrary to the common assumption that Kant dogmatically denies Jesus’ divinity, I argue in “Could Kant’s Jesus Be God?,” *International Philosophical Quarterly*, 52, 4 (Dec. 2012): 421–37, that his official position is agnostic, yet includes an explanation of how one must *interpret* Jesus’ divinity, in order for such a belief to be consistent with the religion of bare reason.

<sup>20</sup>. See, for example, Philip J. Rossi, *The Social Authority of Reason: Kant’s Critique, Radical Evil, and the Destiny of Humankind* (Albany: State University of New York Press, 2005), who claims that Kant fails to specify “the concrete means” for establishing the social authority of reason (e.g., 9, 60). Surprisingly, however, Rossi uses the term “church” only occasionally, preferring the non-religious term “ethical commonwealth.” For an analysis of one of the key arguments in the Third Piece that requires the latter to be conceived *only* in religious terms, see my article, “Kant’s Religious Argument for the Existence of God – The Ultimate Dependence of Human Destiny on Divine Assistance,” *Faith and Philosophy*, 26, 1 (Jan. 2009): 3–22.

---

<sup>21</sup>. Along these lines, I attempt in chapter 9 of KCR to develop a biblical theology that puts worship at the core of historical faith *even though* from the philosopher's transcendental perspective such a prioritization seems upside-down. Kant makes clear in his 1798 work, *Conflict of the Faculties*, however, that philosophers and theologians *ought* to work with opposing principles and that a healthy conflict between these two types of scholars is therefore not only inevitable but potentially fruitful for both camps. Theologians, with their special access to (alleged) divine revelation, can contribute *insight* and *power*, while philosophers must focus their efforts on the guiding light of reason.

<sup>22</sup>. For a more detailed account of Kant's philosophy of prayer along these lines, see my article, "Kant's Critical Hermeneutic of Prayer," *The Journal of Religion*, 77, 4 (Oct. 1997): 584–604; revised and reprinted as appendix 8 of KCR.