

INTRODUCTION

The cloaks under which the embryo first formed itself into the human being must be cast off if he is now to step into the light of day (Religion, 121).

Religion within the Bounds of Bare Reason is among the most influential books in the history of the philosophy of religion; yet, as with many a great book, interpreters incessantly disagree over what the author intended to say. An introductory essay cannot possibly discuss, much less resolve, all the interpretive difficulties facing the twenty-first century reader of this classic text. Fortunately, this new translation corrects one major oversight of past translations by rendering the enigmatic title in a way that preserves the crucial *clothing* metaphor Kant presents in the Preface and employs throughout the book, whereby rational religion is a “bare” (*bloß*)¹ body that is inevitably clothed by some historical faith. Letting the power of this metaphor speak for itself, I divide this Introduction into four sections: after highlighting in §1 the key background issues relating to Kant’s experience of religion, I explain in §2 how *Religion* fits into the structure of Kant’s philosophical System; this prepares us for a detailed overview of the book’s contents in §3, so that we can assess in §4 its implications and ongoing relevance.

1. THE RELIGIOUS CONTEXT OF KANT’S LIFE: CONFLICTS OF HEART AND MIND

Kant’s conflicted religious upbringing is everywhere evident in the pages of *Religion*. The book is bound to be misunderstood by readers who forget that the author was raised as a devout Pietist at the hands of a loving mother whose moral instruction never ceased to inspire his appreciation and respect. Prussian Pietism arose as a movement within the Lutheran (state) Church, its leaders emphasizing private devotion to God, individual Bible study and moral integrity, while deemphasizing

¹ Twentieth-century English interpretations of *Die Religion innerhalb der Grenzen der bloßen Vernunft* were plagued by Greene and Hudson’s (1934) mistranslation of *bloßen* as “alone,” giving rise to a disastrous tendency to read Kantian religion as moral reductionism. Richardson, a much earlier (1798) but now virtually unknown translator, saw the metaphor in Kant’s title but used the rather too explicit “naked.” The Cambridge Edition (1998) uses “mere”; but this ignores the metaphor and has a pejorative connotation that is not present in Kant’s usage. For a full defense of the translation adopted here and its profound implications for a balanced interpretation of Kant’s *Religion*, see Stephen Palmquist, *Kant’s Critical Religion: Volume Two of Kant’s System of Perspectives* (Aldershot: Ashgate: 2000), Chapter VI.

church hierarchy, dogma and the theological presumptuousness that regards ritual as transmitting saving power. Kant's parents raised young Immanuel in a home some might call "fundamentalist" today, so that as a young adult Kant probably could not recall a time when he was not a "believer," so deeply ingrained was his youth in Christian Pietism.

Yet we also must not forget that the mother at whose beloved knee young Immanuel learned the ways of God was taken from this life when he was just thirteen years old and that, from ages eight to sixteen, he experienced such harsh treatment from his Pietist teachers that the very thought of those days filled the mature Kant with dread. Teachers at the *Collegium Fridericianum* would force students to *appear* devout by requiring them to pray aloud, memorize long passages of Scripture, recite creeds, etc. Kant explicitly disapproved of such pedagogical practices, yet ironically, they seem to have had their intended effect (to ingrain the truths of Scripture so deeply in his heart that they would flow through his blood and inform his whole being), for Kant's intimate knowledge of the Bible is evident throughout *Religion*. He often quotes or alludes to specific passages and weaves biblical themes into his rational arguments as if they were second nature; yet he seems at times to struggle against this tendency, even as he had refused to let his spirit be broken by the harsh schoolmaster of his youth.

The tendency of Kant's antireligious readers, including many Kantians who are attracted by other aspects of his System, to read into *Religion* a total disdain of anything religious must be avoided by anyone who wishes to understand this book in a hermeneutically responsible way. Tempting though it may be to think of Kant as renouncing all religion after those horrific days at the *Collegium* (or perhaps as never even embracing it), Kant went on to matriculate as a *theology* student at the University of Königsberg; and while serving as a private tutor for a pastor's family in his twenties, he probably filled in for his employer by occasionally preaching sermons. Moreover, Kant's writings nowhere provide clear evidence that he ever totally abandoned the essentials of his religious upbringing. Although his philosophical training and Critical reflections led him to remain silent about his personal beliefs, from his first publications to his last Kant seemed intent on discussing *theories* that have a direct impact on both theological belief and religious practice. One could argue that Kant's whole philosophy is *theocentric*, though perhaps only in the sense that the center of a storm is calm: all the energy generated by his central reflections on God ends up being channeled into *anthropology*, in an effort to help human beings understand who we are and how our lives should be lived in a religiously authentic way.

Kant's education and early publications took place during the heyday of the European Enlightenment and were deeply influenced by philosophers such as Christian Wolff (1679–1754), who declared himself a "supernaturalist" in matters of revelation even while arguing that reason provides us with a completely natural religion. This was a time when Reason reigned supreme in academia as well as in some quarters of the church. Optimistic rationalism—the assumption that reality is

inherently logical, so that reason can and should solve every problem—was taken for granted in one form or another as the proper philosophical standpoint for assessing any area of culture, including religion. Kant kept well informed on Enlightenment trends in theology and religious studies, as evidenced by his allusions in *Religion* to numerous contemporary scholars in these fields. Josef Bohatec points out that Kant borrowed much of his theological terminology from works of the Swiss theologian, Johann Friedrich Stapfer (1708–75), including his twelve-volume *Grounding for the True Religion*.² Kant was also conversant with the great biblical critic, Johann Salomo Semler (1725–91), a principal representative of German theological rationalism.

That Kant himself never fully adopted such Enlightenment assumptions may explain why he did not stand out as a remarkable student in his teachers' minds. His thinking was already advancing beyond theirs, already preparing for the paradigm shift he would later cause in the history of philosophy (see § 2), so to them he must have seemed like someone who did not quite "get" what philosophy is about. During his thirties, Kant appears to have loosened, or completely broken, the commitments he once had to organized religion; he began to lead a lively social life and (contrary to common caricatures of the lifelong bachelor) even had several love affairs.

As he entered midlife, already moderately well known for his early publications but several years before becoming Chair Professor at the University of Königsberg, Kant became intrigued by stories he had heard about the Swedish mystic Emanuel Swedenborg, but was disappointed by Swedenborg's refusal to reply to his questioning letters. His critique of Swedenborg's religious fanaticism, *Dreams of a Spirit-Seer Illustrated by Dreams of Metaphysics* (1766), exhibits more poignantly than any of his other writings the conflicted attitude that always characterized Kant's approach to things religious. While on the one hand calling Swedenborg a windbag fit for a mental hospital, he affirms Swedenborg's explanation of how spiritual visions occur as matching almost exactly with his own metaphysical theories. The crisis Kant experienced in the series of events that led him to write *Dreams* triggered a fifteen-year period of low productivity, when he published only a handful of small essays as he reflected deeply on how to solve, once and for all, the problem he now believed to be shared by metaphysicians and mystics alike. He recognized that the problem of metaphysics essentially corresponds to a *religious* and *theological* problem, so the solution to the former, as contained in his three *Critiques*, was bound to have implications for the latter. This helps explain why Kant taught classes on Rational Theology at least four times during the 1770s and 1780s and why the first book he wrote after publishing the third *Critique* (1790) was *Religion* (1793).

² See Pluhar's note 42 to the First Piece. Pluhar's notes provide numerous such references.

Kant could not have anticipated in 1766 that, over twenty years later, as he was diligently writing the books that constitute his System of transcendental (or Critical) philosophy, the religious tolerance that had characterized Prussia during his entire adult life would come to an abrupt end. King Frederick the Great (to whom Kant dedicated his *Critique of Pure Reason*) had been on the throne since Kant was sixteen; but shortly before the king died in 1786, Frederick's nephew, Friedrich Wilhelm II, was crowned. The new king, an orthodox Christian inclined toward mysticism, appointed an outspoken Freemason and Rosicrucian, Johann Christoph Wöllner, as his privy councilor for finance and de facto prime minister. In 1788 Wöllner issued an edict regulating the publication of ideas that might threaten the stability of the Lutheran church and other prominent religious groups, specifically mentioning the need to protect Christianity from the "Aufklärer" (i.e., Enlightenment scholars). By this time Kant was so firmly established at the top of Prussian philosophical academia that he dared to follow through with his long-planned application of his Critical System to religion, in spite of this edict. As a result, even the publication of *Religion* ended up being shrouded in conflict, this time of a political nature.³

Shortly after the 1794 publication of the second edition of *Religion*, Wöllner's edict became the basis for a royal order in October 1794, accusing Kant of "distorting and disparaging several principal and fundamental doctrines of Holy Scripture and of Christianity" and ordering him to cease and desist from all such activities. Kant had published the "First Piece"⁴ in a *Berlinische Monatsschrift* article in April 1792, thus alerting the authorities of the challenging nature of his position. When he submitted the Second Piece that June, it was rejected by the censor as too theological. Kant responded by compiling the four essays as a book and sending it for approval to the Dean of the Philosophy Faculty in Jena, who was able to approve it without submitting it for theological censorship. This slap in the face to Wöllner (the Second Piece already having been banned) had prompted a letter of reprimand, threatening harsh consequences should Kant attempt such an act of subversion again. So the royal order of 1794 was no laughing matter. Kant replied to Wöllner promising never to speak or write on religion again, "as long as the king shall live." Some criticize Kant for giving in to this pressure. One wonders, however, what more he could have done from a prison cell, where he would have ended up had he disobeyed, than he did from his professorial podium over the next few years.

³ See Manfred Kuehn, *Kant: A Biography* (Cambridge: Cambridge University Press, 2001), 362–366.

⁴ That Kant labels the chapters of *Religion* as *Erstes Stück*, *Zweites Stück*, etc. (literally "First Piece," "Second Piece," etc.) has been a matter of considerable confusion to English readers, especially since Greene and Hudson misleadingly used "Book One," "Book Two," etc. *Stück* appears to be a reminder that *Religion* consists of four essays ("pieces") originally written as a series of journal articles.

2. THE PLACE OF *RELIGION* IN KANT'S PHILOSOPHICAL SYSTEM

Kant was a late bloomer. He was 57 when he completed the first of his three *Critiques*, but he then completed the other two, along with several supporting works, in just nine years. His many years of deep reflection on Enlightenment pretensions about reason's capacities had paid off. In addition to *Dreams*, several of Kant's early writings provide important clues to his mature view of theology and religion. *The Only Possible Argument for the Demonstration of the Existence of God* (1763) argues that the traditional proofs for God's existence all fail but can be replaced by a new "possibility" proof. He then gives morality an independent status from religion and theology in "An Inquiry into the Distinctness of the Principles of Natural Theology and Morals" (1764). But the radical implications of his new "Copernican" insight, whereby the philosopher's task is to discern the necessary ("transcendental") conditions for the possibility of whatever type of experience is under consideration, were still largely unforeseen up to this point.

Critique of Pure Reason (1781) first introduces God as one of three "ideas of reason" that constitute the proper subject-matter of metaphysics. Aside from some vague hints about God's mind, unlike ours, having the power of "intellectual intuition" (whereby objects come to exist merely by being present in the mind), Kant says very little about God until the Fourth Antinomy, where he argues that pure reason is incapable of determining whether the existence of the universe requires a *first cause*, because logic can be used to defend one side of the debate just as persuasively as the opposite side. This sets the context for Kant's discussion of proofs for the existence of God in the "Ideal of Reason," the longest chapter in the *Critique*. Kant there updates his earlier (1763) distinction between three types of proof and argues that all possible theoretical proofs fail to establish the existence of a divine being. The only nonempirical elements that can be admitted into a theoretical system are those with a transcendental status, serving as necessary conditions for the possibility of the experience relevant to the system; but God, freedom, and immortality are not necessary to scientific knowledge. From the standpoint of theoretical reason, we must therefore admit we are ignorant of God's existence. Kant thinks this result *protects* religion from philosophers who attempt to use pure reason to prove that God does *not* exist. The most Kant will allow here is a "regulative use" of the concept, whereby we may talk and act "as if" God exists, if we have sufficient practical reasons for doing so, even though we remain totally ignorant of the *fact* of God's existence in any scientifically significant sense.

Critique of Practical Reason (1788) defends what may seem to be a radically inconsistent position on God, but is better viewed as a natural outworking of the theoretical *Critique*. Having barred human beings from all scientific *knowledge* of God, Kant now claims our moral nature provides an overwhelmingly persuasive

reason for *believing* in God. He bases this claim on a peculiar feature of our moral situation: all rational beings who possess freedom of choice find themselves constrained to respect a “moral law,” whereby each person must give the interests of other rational/moral beings equal weight to one’s self-interest; likewise, all embodied rational beings should agree that a meaningful life, one that is truly worth living, must have an appropriate level of happiness. The problem is that following the moral law often requires us to sacrifice the very happiness we associate with life’s meaning. Kant thinks this puts us in a situation of practical absurdity, unless we “postulate” beliefs that can preserve the rationality of moral action. In short, he claims we must put forward as actual (i.e., believe, without knowing) some (unspecified) form of life after death and a moral God who will guarantee that “the highest good” (i.e., happiness in proportion to a person’s virtue) will become real in that future life.

While the first two *Critiques* answer Kant’s basic philosophical questions, “What can I know?” and “What ought I to do?” his *Critique of the Power of Judgment* (1790) is supposed to answer a third question, “What may I hope?” The problem is that the third *Critique’s* analysis of beauty, sublimity, and natural purposiveness—regarded as mysterious (but nonetheless real) symbolic forms of connectivity between the otherwise discrete realms of nature (knowledge) and freedom (morality)—does not address very explicitly the obviously religious tone of the third question. No doubt, this is why Kant added a lengthy Appendix setting out a “moral teleology” that presents the purposiveness of nature as pointing ultimately to a moral God whose creative purposes are fulfilled in humanity’s moral nature. Even with this Appendix, however, the third *Critique* provided an incomplete and only partially successful explanation of how human hope arises at the intersection of nature and freedom. The need for a more complete answer to the question of hope drove Kant to devote his next major work to a Critical philosophy of religion.

Religion relates to the other books in Kant’s System as part of his attempt to bridge the noumenal and phenomenal realms—and so also, the corresponding practical and theoretical standpoints. He accomplishes this by interpreting religious symbols as pointers to moral truths. This feature leads some (especially antireligious) readers to view the book as merely an extension of Kant’s ethics, intended to reduce all religion to morality. Although the First Piece on its own does constitute an important fine-tuning of Kant’s ethics, this accomplishes the very opposite of reduction: Kant’s argument is that evil corrupts our moral nature in such a way that morality must be *raised* to the status of religion in order for us to have any reasonable hope of fulfilling the moral demand. The detailed summary in § 3 highlights how *Religion* embodies the central conflict of Kant’s entire System, attempting to resolve it through a refined religious hope.

The censorship described in §1 did not prevent Kant from continuing to think and write about religion during his period of forced silence, for in 1798, soon after the king died in November 1797, Kant published his most mature account of the

relation between the theology faculty (the academic department that trained professional ministers to serve ordinary religious people) and the philosophy faculty. *The Conflict of the Faculties* enshrines the key feature of Kant’s lifelong experience of religion, *conflict*, as a principle for responsible dialogue between philosophers and other academics. His Preface tells the censorship story and claims *Religion* is grossly misunderstood if the reader fails to recognize its central goal of affirming Christianity as “the universal religion” of the human race. Part One of *Conflict* goes on to explain that the philosopher’s proper role in relation to the biblical theologian is to *maintain a healthy conflict*. The interests of the public (and for religion, this means ordinary religious believers) are best protected if scholars of each type adopt the assumptions and methods appropriate to their respective standpoints, in open dialogue with the other: philosophers rightly start with bare reason and seek to assess the meaningfulness of any alleged revelation, while biblical theologians rightly start from revelation and use reason to understand its content.

Kant explicitly identifies his intended readership in the Preface to *Religion* as consisting of these same two groups, whose natural inclination is to steer clear of each other’s influence—much as did the dogmatists and skeptics whose positions the first *Critique* attempted to bridge. First and foremost, he addresses Enlightenment philosophers, many of whom were dogmatic in their rejection of empirical religion as irrelevant to their project—not surprising at a time when churches were in danger of being hijacked by fanatical extremists like Wöllner. They typically assumed bare reason can account for *all* religious truth, so that real empirical religion can be discarded as wholly illusory. A central goal of *Religion* is to demonstrate the rational instability of that position, inasmuch as morality cannot reach its aim, in the historical fulfillment of human destiny, without the aid of empirical religion. The second, equally important target group consists of Christian pastors and theologians (as well as theology students), many of whom (including some of Kant’s friends) were skeptical about philosophy having any relevance to ordinary religious believers. This is why Kant never questions the *givenness* of religious experience but instead allows that the rituals and beliefs found within a specific religious tradition may have some genuine meaning not only to the believers but possibly (we can never know) even to God. To construct a persuasive transcendental argument, one must assume the basis for proof an experience one’s skeptical readers will grant as genuine (see note 5). Kant, so often misunderstood by interpreters such as Gordon Michalson, who blame him for their confusion when he refuses to take sides, therefore freely grants, as the basis for such arguments, the concepts and practices assumed by his religious readers. As he writes in the first Preface, he wants philosophical theologians and biblical theologians “to be at one.” The aim of *Religion* is to entice these groups into a creative conflict with the potential to transform both in the service of what Kant saw as the single greatest goal of humankind: the establishment of a community of right-minded and good-hearted persons.

3. CONFLICTS IN KANT'S *RELIGION*: SUMMARY OF THE TEXT

The four essays compiled in *Religion* constitute a system based on the same architectonic form Kant used as the pattern for his other systematic works. Following his table of four categories (quantity, quality, relation, modality), the “pieces” adopt, in turn, what I elsewhere call his transcendental, logical, empirical, and hypothetical perspectives. Subsections B–E, below, each highlights a specific conflict that arises when viewing religion from one of these perspectives. Subsection A examines the overarching conflict Kant introduces in the two Prefaces (the second added when Kant published the second edition in 1794, just before Wöllner's censorship silenced him). Subsection F groups together the four *parerga*, or byproducts of reason, that Kant discusses in the General Comment appended to each piece, thus enabling us to see them as a systematic whole that follows the same fourfold (categorical) pattern as the four essays they supplement.

A. PREFACES: THE CONFLICT BETWEEN BIBLICAL AND PHILOSOPHICAL THEOLOGIAN

The first Preface begins by positing the distinction that forms the basis for Kant's entire religious system. When viewing humans as free, rational beings, we need not appeal to concepts such as God or happiness to explain why we should follow the demands of our inner moral nature. From this rational (philosophical) standpoint, “morality in no way needs religion . . . ; rather, through the power of pure practical reason it is sufficient to itself.” Nevertheless, the maxims we adopt when following the moral law point beyond themselves to “necessary consequences” that constitute the “purpose” of our moral nature. To ask what results from our right action is to adopt an empirical (historical) standpoint, focused on “natural need” rather than on reason's formal requirements. From this standpoint, the idea of religion (i.e., belief in “a powerful moral legislator” who can guarantee the happiness of those who obey) “emerges from morality” as “a final purpose of all things.” Religion combines “the purposiveness arising from freedom with the purposiveness of nature,” the themes of the second and first *Critiques*. Morality is, therefore, independent from religion, yet leads “inescapably” to a religious goal that manifests itself in human history.

After an interlude claiming all treatises on religion ought to submit to public censorship, the first Preface distinguishes between theologians responsible for “the welfare of souls” (i.e., clerics) and those responsible for “the welfare of the sciences” (i.e., university scholars). Scholars, Kant claims (in a passage that must have infuriated the cleric Wöllner), should have priority over clerics in matters of censorship because the university has a built-in mechanism for self-censorship: it employs *two* types of theologians, each with a distinct responsibility to protect one of the two standpoints on religion. While the biblical theologian starts with

Scripture, using reason as a vehicle to clothe the explanation of religious truth, the philosophical theologian starts with “bare reason” and uses Scripture as a vehicle to enlighten its independent search for truth. Kant portrays these two as *neighbors* engaged in a creative conflict: each inhabits a separate “territory” and must respect the other's property rights, taking care not to encroach onto the other side of “the bounds of bare reason.” Just as biblical theologians use rational argument to defend the truths of Scripture, philosophical theologians may “borrow from” Scripture, though they cannot dictate what biblical theologians should tell clerics about the meaning of Scripture. Having clearly separated these two “sciences,” Kant tells us he will conduct the “experiment” of “consider[ing] them as united.” Kant therefore recommends *Religion* to biblical theologians as a text for teaching students how to make a right use of reason in interpreting the Bible.

The second Preface, published less than a year later, serves as a postscript to the first. Kant begins by refining his territorial metaphor to clarify a possible ambiguity in the book's title: the philosophical standpoint of “pure rational religion” and the historical standpoint of “revelation” relate to each other not as independent domains but “as concentric circles,” with the “*wider* sphere of faith enclos[ing] pure rational religion as a *narrower* one”; the philosopher's “bare a priori principles” of morality define the inner circle that marks the boundary between the two. The “experiment” mentioned near the end of the first Preface seeks to resolve the conflict between philosophical and biblical theology by developing a version of the former that encompasses the core ideas of the latter. Kant now adds that *Religion* conducts a “second experiment, namely to start from some supposed revelation and . . . to hold the revelation as a *historical system* up to moral concepts . . . and to see whether this system does not lead back to the same pure *rational system* of religion.” These two experiments have the same goal, to find “not merely compatibility but also unity” between the conflicting standpoints of the biblical and philosophical theologian.

The second Preface concludes with a response to a critic who claimed *Religion* is as incomprehensible as the rest of Kant's philosophical System and can safely be ignored by anyone who chooses not to ask Kant's questions. Both the tone and content of Kant's reply are noteworthy: “To understand this work in terms of its essential content, only common morality is needed, without venturing into the critique of practical reason, still less into that of theoretical reason”; while using some technical terms, such as “*phaenomenon*” and “*noumenon*,” is unavoidable as a concession to “the school” (i.e., so that *Religion* might serve as a useful textbook), “the matter itself is contained, even if in different words, in the most popular instruction for children or in sermons and is readily understandable.” Kant's nostalgic tone, hearkening back to his religious upbringing, suggests he saw *Religion* as resolving his own private conflicts over religion. Unfortunately, over two centuries have elapsed since he wrote those words, and few (if any) readers have found the book's content to be as simple as Kant portrays it to be. My goal here in §3 will therefore be to describe how the basic “matter” of each piece serves to resolve the conflicts

of religion by conducting two experiments: constructing a theory of rational religion that coincides with the core truths of biblical theology; and deriving religion's rational truths from biblical theology.

B. FIRST PIECE: EVIL AS THE TRANSCENDENTAL ROOT OF RATIONAL RELIGION

Do human beings start out good and become bad as time passes, or do they steadily evolve from bad to better? Kant's introduction to the First Piece argues that answering this eschatological question requires us to identify what can be ascribed to human beings *by nature*: we are either basically good or basically evil. Philosophers cannot resolve the "conflict" between these two "rigorist" positions by appealing to either intermediate option—that we are "neither good nor evil or . . . partly good, partly evil"—because their task is not to assess individual experiences, but to determine what we are "able to infer a priori from" any given choice(s). Even though experience "seems to confirm [an] intermediate position," philosophers must look to the purity of the maxim (or rule) that guides a person, and this cannot be a mixture but must be pure, because only purely good or purely evil incentives could serve as the basis for genuinely free choice. The conflict here is between the philosophical rigorism that judges "on the scales of pure reason (before a divine tribunal)" and the intermediate positions that judge "according to an empirical standard (by a human judge)." The philosopher's goal is to locate the subjectively universal basis of the good or evil maxims that give rise to human actions. Identifying this basis as our "nature" does not make it a natural impulse; rather, it is a deeply hidden "act of freedom" that is "inscrutable to us." Calling it "innate" means this fundamental choice to be good or evil is somehow already present at birth, not that birth is its cause. The only rational explanation for a person choosing to act in a way that does not conform to reason's internally-legislated moral law is to presuppose the person has adopted a "supreme" maxim that was determined by an opposing incentive consisting of "positive evil." This transcendental perspective relates to humans *as a species*; whether any individual could be immune from an evil nature predicated to the species—as some claim for religious figures such as Jesus—is a question Kant defers to "anthropological investigation."

Kant resolves this transcendental conflict by portraying human nature (i.e., the human species) as having an "original predisposition" to good that has been corrupted by a "propensity to evil." Sections I and II explain how each component, in turn, applies to three aspects of our nature: that we are animal (living), human (rational), and personal (accountable) beings. Animal self-love predisposes living beings to do good by causing them to preserve themselves, propagate the species, and form social groups for mutual protection. Human self-love predisposes rational beings to *compare* themselves with others, inclining each "to procure a worth for oneself in the opinion of others," and thereby drives us to create culture. While various

evils can be "grafted" onto the first two aspects, our "personal" nature ("the receptivity to respect for the moral law") is *entirely* good, for "absolutely nothing evil can be grafted" onto it. The mere awareness of such respect does not suffice to make us persons, because we would not actually *choose* to adopt the moral law into our maxims if there were not "present in our nature a predisposition" to do so. These three aspects of the predisposition constitute the first of two *transcendental boundary-conditions* for Kant's religious system, for "the conditions of their possibility" are grounded in reason; they "belong to the possibility of human nature," being necessary for the very possibility of "our power of choice." As the starting point of the first experiment, the predisposition to good does not make us actually good, but establishes only that we are *made to be good*.

If this predisposition were the only transcendental element of human nature, we would all inevitably become good, so our "habitual desire" to allow *evil* motives to determine our choices requires us to infer a second condition, a "propensity to evil," as "the subjective basis for the possibility of" choosing evil. As a "contingent" tendency to desire evil once we experience it, the propensity is, paradoxically, "natural" or "innate": we must conceive it as applying universally, to all human beings, even though we bring it upon ourselves through the choice whereby we first graft it onto the good predisposition. The "heart" (or "will") is Kant's term for the part of human nature that chooses whether to adopt this propensity as the supreme rule of choice (making an *evil heart*) or to preserve the moral law as the supreme rule (making a *good heart*). The evil propensity has three "levels": "frailty" is the tendency to let (animal) inclinations overpower one's rational choice to follow the moral law; "impurity" is the need to supplement the moral law with other incentives (e.g., based on the rational impulse to compare oneself with others) before making the right choice; and "perversity" is the habit of "revers[ing] the moral order in regard to the incentives" determining one's choice, so that personal happiness (self-love) comes *before* any consideration of the moral law. These three classes of propensity must underlie the actions of "even the best human being . . . if one is to prove . . . [it] as universal"—i.e., a transcendental element in a system of rational religion. A person whose actions comply with the moral law is "legally good"; but to be *morally good* one's choices must be motivated by *no incentive other than the moral law*. In order to determine free choice, the evil propensity cannot be *physical*, but must be *moral*; it must arise from "an intelligible deed, cognizable by bare reason, without any time condition"—a transcendental "deed" that is presupposed by all our *empirical* (physical) deeds.

After highlighting the tension between predisposition and propensity, Section III adds a new step to the argument: because evil clearly *is* a feature of human experience, "we may presuppose" the evil propensity in *all* human beings, as "a *radical, innate evil* in human nature"; otherwise we cannot explain how the good predisposition could fail to determine our moral character. In what may be the most infamous sentence in *Religion*, Kant says "we can spare ourselves the formal proof" that this evil propensity "must be rooted in the human being . . . in view of

the multitude of glaring examples that experience puts before our eyes.” He then cites examples of human evil, ranging from that found in primitive cultures to that of civilized states, making the Enlightenment’s dream of “the perfected moral improvement of all of humankind” look like mere “fanaticism.” These examples should not be read as an *excuse* for not proving the evil propensity, for Kant later says (see note 5 below) he has already given a “formal proof.” If human beings are good by nature (a necessity that does not make us *actually* good), then any evidence of actual evil requires us to infer a “bare” (timelessly “chosen”) evil propensity; these examples are the evidence that at least some human beings *are* evil; so the formal conclusion of what amounts to a transcendental argument is that the evil propensity must exist as a universal feature of human nature.⁵

The remainder of Section III clarifies various features of this proof. First, Kant rules out the two alternatives to the evil propensity: the basis of evil deeds can be found neither in “the natural inclinations” of our sensibility (because these provide the occasion for virtue), nor in a corrupted reason that lacks respect for the moral law (because without the moral law we would not have freedom to choose). Humans are neither mere animals (acting solely on natural inclination) nor devils (actively opposing the moral law). The “proper constitution” of the evil propensity “must be cognized a priori from the concept of evil,” not merely from examples, for a person’s moral character does not consist of deeds based solely on one incentive (the moral law) or the other (inclinations toward self-love), but depends on “which of the two [a person] makes the condition of the other.” An evil person merely “reverses the moral order of the incentives in admitting them into his maxims,” making self-love the supreme maxim instead of the moral law. If, for prudential reasons, one’s deeds remain consistent with the moral law, “the empirical character is good, but the intelligible character is always still evil.” This “*perver-*sity of the heart” gives us an “*innate* guilt” that we can perceive as soon as free choice manifests itself in our moral development.

Having completed the first stage of his rational system of religion, while alluding to many otherwise tangential biblical texts as confirmation of his first experiment’s success, Kant devotes Section IV primarily to the *second* experiment, explicitly focusing on the doctrine of original sin. To find the “origin” or “first cause” of a given effect (e.g., evil), he argues, we can look either for the “rational origin” of its *existence* or for the “temporal origin” of its *occurrence*. But a temporal origin of a free choice would be “a contradiction”; the search for a temporal

⁵ “The proper proof” that *all* people are evil, Kant claims in a footnote at the end of Section III, “is contained not in this section but in the previous one. This section contains only the confirmation of the judgment through experience.” This indicates he is following the standard form of a transcendental argument. For a detailed account of the structure of this argument and of why it is transcendental, see my article, “Kant’s Quasi-Transcendental Argument for a Necessary and Universal Evil Propensity in Human Nature,” *The Southern Journal of Philosophy* 46.2 (Summer 2008): 261–297.

origin therefore leads us to infer a universal evil propensity as the transcendental basis of our contingent evil deeds. Treating the biblical story of the Fall as an account of evil’s “*inheritance* from the first parents” is “the most inappropriate” interpretation, because it offers an empirical solution to a philosophical problem. However, if we interpret the story as a symbolic description of how in “[e]very evil action” a person has “fallen into it directly from the state of innocence,” this way of explaining evil’s origin “agrees quite well with” Kant’s rational account. Scripture, according to Kant, correctly portrays evil not as arising from a necessary trait of our nature, but as a free rational choice, in response to a strict “divine command,” whereby we “use subtle reasoning to downgrade [our] obedience to the command,” considering “self-love” and “sensible impulses” to be more important than the moral law—this being Kant’s philosophical definition of “sin.” Whereas the first sinner was innocent, we must presuppose “an innate propensity to transgression” in ourselves—a non-temporal (and so, unconscious) evil deed of preferring self-love—that grounds all our temporal evil deeds. Kant thinks the biblical depiction of Satan as “a seducing *spirit*” is an appropriate symbol of the “incomprehensibility” of this rational origin of human evil. Without straying into the biblical theologian’s domain, the second experiment here shows how the Bible leads directly to the core truths of rational religion concerning evil.

When published as a journal article, the First Piece appeared in five sections. In the book’s second edition, Kant moved Section V to the concluding General Comment, because it previews the Second Piece rather than elaborating further on the First Piece’s theme. However, this clashes with the General Comment’s purpose, to discuss a *parergon* corresponding to the perspective of the First Piece (see § 3.F, on p. xliii). A better way of dealing with this “misfit” section would have been to make it a transitional Appendix or even an introduction to the Second Piece. As such, I discuss it, in next subsection, as a transition to the second stage of Kant’s argument.

C. SECOND PIECE: GRACE AS REASON’S LOGICAL POWER TO COMBAT EVIL

Kant’s transition to the Second Piece (appearing in the First Piece’s General Comment) lays out the logic of how reason uses religious concepts to combat the dire consequences of evil, in hopes of “restoring . . . the good [predisposition] to its power.” While God’s cooperation may be needed for such restoration to be possible, Kant insists “the human being must . . . *accept* this aid” if any renewed goodness is to be imputed. That such a transformation “surpasses all our concepts” does not make it impossible, for as demonstrated in the First Piece, the corruption of the good predisposition by the evil propensity has occurred even though we cannot conceptualize it either. All we can do to “make ourselves receptive to a higher and to us inscrutable assistance” is focus on “a germ of the good” that remains in us—i.e., our continued ability to obey the moral law, despite our evil propensity. We

must restore the *purity* of the moral law as our sole incentive for making moral choices. Firmly resolving to do good may enable our *empirical* character to improve, but such change is inevitably gradual and can be motivated by self-love; to become “pleasing to God” (the goal of all religion) requires *virtue* in one’s “intelligible character,” and this can result only from “a change of heart.” Kant applies his key perspectival distinction to explain how this inward “revolution” is logically possible: God must assess the human heart and its supreme maxim from the intelligible standpoint, whereas we can assess our moral disposition⁶ only on the basis of our particular choices and the gradual reform we experience in time. God must view our “constant *progress*” as “a unity . . . tantamount to actually being a human being who is good (pleasing to him).” Many “impure religious ideas” arise because people tend to be morally lazy, inventing ways of obtaining God’s help without requiring a change of heart. Yet we need not know *how* God assists us in order to be empowered by the bare idea that such assistance is available to anyone *morally* prepared to receive it.

After an introduction praising Stoic virtue because it “presupposes an enemy” and thus discourages moral laziness, though it wrongly identifies the enemy as “the natural inclinations” rather than as a corruption of morally-legislating reason (i.e., as the temptation to act without regard to *any* principles rather than as a *rational choice* to adopt a perverted principle of self-love), the Second Piece examines “the struggle of the good with the evil principle for dominion over the human being.” Its two sections correspond (as do the main divisions of the Third and Fourth Pieces) to Kant’s two experiments. Section One adopts the philosopher’s rational standpoint to assess the “legal claim of the good principle,” while Section Two adopts Scripture’s historical standpoint to assess the “legal claim of the evil principle.”

⁶ Kant’s elusive term, *Gesinnung*, is normally translated as “disposition” in religious contexts. Pluhar’s use of “attitude” following his translations of the three *Critiques*, preserves self-consistency but is likely to be misleading in *Religion*. While I appreciate and agree with his misgivings about “disposition,” “attitude” is too vague and too empirical to convey the deep (and hidden!) *power* Kant attributes to *Gesinnung*, especially in the Second Piece. Here *Gesinnung* is not just any attitude, but a very special, *fundamental attitude* (a *Grundhaltung*) that directs the heart toward one goal (evil) or another (good); upon it rests nothing less than a person’s salvation! Ordinary attitudes change easily; yet a change in one’s *Gesinnung* is a rare and revolutionary event. I toyed with using “conviction,” a word that aptly refers to the belief-oriented commitment that is central to the way Kant thinks *Glaube* (faith or belief) influences our actions. Provided we understand it not as a reasoned (propositional) conclusion but as a heart-centered *resolve to act in a certain way*, “conviction” is my preferred translation. However, its similarity in meaning to *Überzeugung* (translated by Pluhar as “conviction”), a term Kant uses more narrowly and sometimes pejoratively, led me to give up this option for this essay. That Kant never uses *Haltung* (the standard German word for “attitude”) in *Religion* reflects that he does not think our *psychological attitudes* have the least relevance to religion. To alert readers of Pluhar’s translation to this potentially disastrous misreading of Kant’s religious theory, I use “disposition(s)” throughout this Introduction. Aside from this one term, I follow Pluhar’s translation in all quotations from *Religion*.

Subsection A of Section One presents the key element in the second (logical) stage of Kant’s rational system of religion as the “Personified Idea of the Good Principle.” Given that *perfect humanity* is “the purpose of creation,” the idea of a perfect human being “emanates from God’s essence” as an “only begotten Son,” a “word . . . without which nothing exists that has been made.” After quoting several biblical references to the *λόγος* (*logos*; i.e., ‘word’), Kant claims this “rational being,” as the “archetype of the moral disposition in all its purity . . . can give us power” to become good again. We can aptly depict this archetype within us by describing it (as in the biblical account of Jesus) as having “*come down* to us from heaven” to assume humanity, yet without being corrupted by the evil propensity, for our predisposition “by itself is not evil.” To serve as an archetype for us weak human beings, who are constantly “wrestling with obstacles” when trying to do good, our idea of such a perfect person must be of one “who would not only be willing to perform any human duty” and to teach others to do the same, but also “be willing to take upon himself all sufferings” even to the point of dying “for the sake of the world’s greatest good.” Only a person who has “practical faith in this Son of God”—i.e., confidence that he or she can “unshakably continue to adhere to humanity’s archetype and to imitate its example in faithful emulation”—“is entitled to regard himself as . . . an object not unworthy of divine pleasure.”

Subsection B argues that the “Objective Reality” of this idea of a perfect human person is self-evident, viewed from the non-temporal standpoint of practical reason, because duty demands that we always “be in conformity to it.” We need not understand how this can happen in order to know it must be possible; the pure rational idea can empower us without being embodied in any “example from experience.” One who demands *more* than the bare (non-temporal, intellectual) choice to adopt duty alone as a supreme maxim thereby “confesses . . . his moral *faithlessness*.” We cannot require a miracle, for example, as proof that God has made a person perfect, because we could identify such a miracle only by appealing to the idea of perfection. Examples of perfection in our temporal experience are possible, because “every human being should provide an example for this idea in himself.” Yet, paradoxically, as an inner idea, “no example in outer experience is adequate to it”; even through introspection, we cannot penetrate with certainty to the rational root of our deeds (i.e., their supreme governing maxim) but can only *hope*, based on our experiences in time. This conflict between reason’s limits and human hope is what gives rise to religion, so we should not be surprised to see Kant struggling with it throughout *Religion*.

The subtlety of Kant’s resolution of the religious conflict is no more evident than in his comments on the status of real religious/moral heroes, such as Jesus (who remains unnamed throughout the book). In the second half of Subsection B, Kant says even if “a truly divinely minded human being” were to appear in “outer experience” as “a human being pleasing to God,” no matter how great a “revolution in humankind” his life and teachings brought about, we would not need “to assume in him anything other than a naturally begotten human being,” because the

archetype *already* has a divine origin. We cannot rule out the theoretical possibility of a wholly non-natural origin; yet belief in such an origin is unlikely to empower us to experience a change of heart. Instead, it may *hinder* the required change: if it makes “any transgression absolutely impossible for him,” then this “divine human being could no longer be set up as an *example* for the natural human being.” This may inspire great “love and gratitude” toward the divine human, but not *imitation*. A fully human teacher, by contrast, could empower his followers to imitate him in a “completely valid” way by speaking of his moral disposition as divine, referring to its intelligible origin, for this is something all human beings possess. Faith in such a historical example empowers religious believers to appropriate the archetype’s perfection—if certain “difficulties” in comprehending such appropriation can be resolved.

Subsection C presents three difficulties threatening the reality of the archetype and solves each by relating it to the essential religious conflict. The first difficulty arises because at any given time our deeds are “deficient”; they cannot clearly reflect the “holy principle” duty requires us to adopt as supreme maxim. The moral law demands absolute purity, yet we start out with an evil propensity. How, then, can our corrupted disposition “count for the deed”? Kant’s solution distinguishes the timeless (divine) and temporal (historical) standpoints on moral judgment. By projecting our moral progress in time “*ad infinitum* toward commensurateness with that law,” the judge “who knows the heart” can create “in his pure intellectual intuition,” based on our temporal development, “a perfected whole also in terms of the deed (the way of life).” Even though we continue committing evil deeds after experiencing a change of heart, God can be pleased with a good-hearted person’s *potential* future perfection.

The second difficulty arises when a person “striving toward the good” wishes to be assured of eventually reaping the happiness that awaits one who persists on the path toward perfection. Kant’s chief concern throughout the Second Piece is with moral empowerment: *total* confidence that one is on the right path could breed laziness, while a total *lack* of confidence could lead to despair. The solution is to be content with *moderate* confidence: a person who can see “a basic improvement” in temporal deeds since “he adopted the principles of the good” can thereby infer the presence of a good heart, and this awareness can “increase one’s *strength*” to stay on the good path, not only now but in any future life. By contrast, one who cannot detect improvement in temporal deeds probably has not experienced a change of heart and so has no reason to hope God will be pleased. This approach empowers *both* good-hearted and (self-deceiving) evil-hearted persons, without requiring us to assume the reward (happiness) or punishment (misery) reaped from one’s way of life must be eternal. The good-hearted disposition thus serves as a spiritual “Comforter (Paraclete) when our lapses make us worried about its persistence.”

The third and “greatest difficulty” arises because the solution to the first difficulty does not explain how a just God can overlook *pre-conversion* evil. We

cannot “extract any surplus” of goodness from our good deeds, beyond what morality intrinsically requires, nor can the “original debt” incurred by pre-conversion evil deeds “be extirpated . . . by anyone else.” Unlike monetary debt, moral debt is not *transmissible*; it “is the *most personal of all* obligations, namely a debt of sins” that can justly be satisfied only by “an *infinite punishment*.” Kant rejects the two most obvious explanations of how a good-hearted person’s pre-conversion sin can be punished: punishment cannot happen *before* conversion, otherwise no conversion would be needed to please God; nor can it happen *after* conversion, for then “the human being is already living the new life and is morally a different human being.” Instead, punishment must be “commensurate with this change” so that the conversion experience *carries with it* (as on a cross) a satisfaction of divine justice. The converted person is *physically* the same, yet morally new; in adopting a good disposition, the new moral person bears the debt of the old person “as *proxy*,” agreeing to endure “a long series of life’s ills”—ills that would *seem* like punishment to the old person—for the sake of upholding the purity of the moral law. This process goes on “continually” in each religious person, even though the archetype (“the representative of human kind”—i.e., Jesus) suffers and dies “once and for all.” The “surplus” that solves this difficulty by being “imputed to us *by grace*” comes from this archetype; “empirical cognition” of our deeds gives us “no legal claim” to be regarded as good by God, except insofar as they give evidence of our “receptivity” to this inner archetype.

Having presented rational reinterpretations of the Christian doctrines of sanctification, eternal assurance and justification, Kant seeks in the remainder of Subsection C to complete this second stage of his first experiment by confirming its compatibility with the standpoint of biblical theology. He says the foregoing “deduction” (the type of proof used to justify the constitutive element of the logical perspective in the *Critiques*) of the archetype’s reality serves a practical purpose by answering “a speculative question” about how divine justice can be reconciled with “the hope for the human being’s absolution from his guilt” through grace. Realizing that hope in grace is rational “only on the presupposition of a complete change of heart” encourages believers to reflect on whether and how they are gradually casting off the old (evil) ways through an “awakening conscience.” Belief in a future, final judgment is morally empowering if we imagine an *inner* judge, because one “cannot bribe his own reason,” whereas viewing the judge as external encourages one to perform deeds of appeasement that do not require a disposition of moral improvement. Typical of the latter, wrong-headed view of religion is the notion that one can enjoy a “life of gratification,” then feign a conversion just before death—a temptation clerics should discourage by refusing to offer doctrinal “opium” for the conscience of one who is about to die but instead *stirring up* the person’s conscience.

Kant turns in Section Two to his second experiment, observing that Scripture symbolizes the same “intelligible moral relation” between the good and evil principles as a struggle between two (unnamed) persons—he is obviously referring to Satan and Jesus—outside the human being. The first humans were created good,

but God gave them freedom by letting Satan try “to acquire a dominion *over people’s minds*.” Despite the Fall, the legal claim of the good principle was safeguarded by the Jewish theocratic government; yet that external symbol of the purity of goodness was later replaced by a government “attuned to no incentives other than the goods of this world.” Jewish moral laws took on external forms, based on “coercion,” and were powerless to combat “the kingdom of darkness.” Suddenly there appeared “a person whose wisdom was purer still than that of the philosophers”; somehow he remains unaffected by humanity’s pact “with the evil principle” (a moral purity aptly symbolized by virgin birth). Jesus threatens Satan’s dominion, because faith in his good example can empower others to turn away from evil; Satan fights back, trying to discredit Jesus through physical persecution. Satan wins the physical battle (i.e., Jesus dies) but loses the moral battle: Jesus remains free by refusing to enslave himself to the evil principle’s rule over his mind. He thus “opens the gate of freedom to all who, like him, want to die unto everything that keeps them fettered to life on earth to the detriment of morality.” Jesus succeeds not in destroying Satan’s kingdom, but only in “the breaking of its power to hold, against their will, those who” have chosen to be slaves of evil. Like Jesus, his followers will experience the persecution involved in renouncing self-love. The obvious consistency between this story and the religious system defended in Section One shows that the “spirit and rational meaning” of the Bible’s teachings have “been practically valid and obligatory for all the world at all times”: Jesus reveals the highest wisdom, for he practices “the *holiest* teaching of reason.”

D. THIRD PIECE: CHURCH AS THE EMPIRICAL REALIZATION OF TRUE RELIGION

Kant introduces the Third Piece by observing that, despite being liberated from evil’s dominion, a person with a good disposition “continues to remain exposed to the attacks of the evil principle” and must be “armed for struggle” in the empirical world. An “intrinsically contented nature” will not suffice to resist the corrupting influence of “hostile inclinations” (e.g., envy, greed, lust) when one is with others—even if they are also good-hearted. We must therefore establish “a union aiming quite expressly at the prevention of this evil and the furtherance of the good in the human being.” This third stage of Kant’s religious system requires “all who love the good” to create “an *ethical community*,” a moral “kingdom” assembled under the “flag of virtue.” Kant examines this idea’s “objective reality”—a term denoting the applicability of a concept to possible objects, typically from the empirical perspective—in two “Divisions,” adopting the philosophical and historical standpoints, respectively.

The first of Division One’s seven subsections distinguishes between “political” and “ethical” situations, wherein people are united either by “coercive laws” or

“laws free from coercion, i.e., bare *laws of virtue*.” The “state of nature” has “no *public* power-holding authority” (whether political or ethical); “each person legislates to himself.” External laws coerce people out of the state of nature into a *political* community but should not aim at “ethical purposes,” for this “would . . . not only bring about precisely the opposite of the ethical purposes, but would also undermine [the] political ones and render them insecure.” Political states should leave citizens free to decide whether to join an ethical community that takes them out of the ethical state of nature. Just as political states should be “connected by a public law of nations,” so also each group of people forming an ethical community constitutes “only a presentation or a schema” of “the ideal of the whole of all human beings” who unite themselves under laws of virtue.

Subsection II presents what amounts to a *religious argument* for the existence of God.⁷ Just as people in a state of nature must form political states to avoid “a situation of injustice and of war of everyone against everyone,” so also they must form ethical communities to defend against the natural state “of inner immorality,” the “*public* reciprocal aggression against the principles of virtue” that causes even good people to have a corrupting influence on each other. Forming such a community is a unique duty “of humankind toward itself,” because the “common purpose” of “every genus of rational beings is determined objectively” to be that of furthering its “highest good as a common good.” This “idea of a universal republic according to laws of virtue” is unlike all other duties inasmuch as it requires *cooperation*. Because “we cannot know whether, as such, it is also in our power” to reach such a lofty goal, yet its status means we must be able to fulfill it, we must presuppose “another idea, namely that of a higher moral being through whose universal arrangement the forces of the individuals . . . are united to yield a common effect.”

Kant clarifies in Subsection III that this presupposition amounts to viewing the ethical community as “a People of God.” To form any community, “all individuals must be subjected to a public legislation.” In political communities the people authorize the legislation; but ethical communities cannot use this method, otherwise the laws would be coerced, not free. As “one who knows the hearts,” God alone can penetrate everyone’s disposition, ensuring each person receives “whatever his deeds are worth.” God can fill this role only if all divine commands are ethical; were God’s legislation external and statutory, it would be coercive (as in a humanly governed theocracy). Thus, we can hope to fulfill our human duty to build an ethical community only by regarding “all *true duties* . . . *simultaneously* . . . as [God’s] commands.”

⁷ For a detailed account of the structure of this argument, 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 (2009), 3–22.

Subsection IV tends to be ignored by readers who view *Religion* as a handbook on morality. Yet Kant here defends the key element in the third stage of his religious system: the ethical community must take the form of a *church*. Because the “crooked wood” of human nature prevents us from building a “completely straight” society, we can conceive of the *form* of an ethical community, but actually *establishing* it on earth is “a work whose execution can be expected not from human beings but only from God himself.” Far from providing an excuse for laziness, this belief requires the good person to “proceed . . . as if everything depended on him,” for only then does one have good reason to “hope that a higher wisdom will bestow completion upon his well-intentioned effort.” The coming of God’s kingdom depends on cooperation between the “invisible church” (“a bare idea of the union of all righteous people under the divine direct but moral government of the world”) and the “visible church” (“the actual union of human beings to form a whole that harmonizes with that ideal”). A “true (visible) church” is one that exhibits the invisible church empirically, “as much as this can be done by human beings.” Kant recommends four (categorical) principles: (1) the *quantity* must be “numerical oneness” or “universality”; (2) the *quality* must be “purity” of moral motivation; (3) the *relation* must be “freedom,” both internal (between members) and external (between the church and “the political power”); and (4) the *modality* must be “*unchangeability*” in the part of its constitution featuring these principles.

Kant argues in Subsection V that a true church requires a *scripture* to compensate for “a peculiar weakness of human nature” that leaves us unable to establish a church on “bare rational faith.” Few realize that anyone with “steadfast diligence directed toward a morally good way of life” is “constantly in the service of God.” Because humans like “attestations of honor,” we tend to think a “morally indifferent” type of service, demonstrating “passive obedience,” will please God more; yet this is “absolutely impossible,” given that we “cannot act upon and have influence upon” God. Knowing “*how God wills* to be venerated (and obeyed)” is the key to all religion: if God commands through “purely moral laws,” then each person can know “through his own reason, the will of God” and “only *one* religion” (the moral) can exist; but if God commands through “merely statutory” laws, “not through our own bare reason,” then a revelation is needed, either “through tradition or scripture.” Special acquaintance with this revelation constitutes “a *historical* faith.” Those preferring the latter answer must recognize that the *former* approach “properly constitutes religion itself” and that “statutory religion can contain only the means to its furtherance and expansion.” For the only possible “universally valid” answer is the moral one; “statutory legislation (which presupposes a revelation) can be regarded only as contingent.”

This appeal to scripture is Kant’s attempt to resolve the Third Piece’s religious conflict (between ethical and political, or divine and human, organization) by conceding that religion does require *some* empirical (external) structure. The historical standpoint requires more than just “bare reason” to answer the question “how God wills to be venerated in a *church*.” Establishing an empirical church is “a duty,” yet

the members are “left entirely to themselves” to learn how best to do so. Claiming a divine origin for empirical religion is presumptuous, but no less conceited are those philosophers who rule out the possibility of “a special divine arrangement” even when the statutes *agree* with moral reason and appear to have arisen out of nowhere. Because human beings tend to believe “in statutory divine laws” that go beyond moral goodness, “church faith naturally precedes pure religious faith” even though “[m]orally it should happen in reverse.” Assuming this natural need “cannot be changed,” and pure religious faith must use statutory faith “as a vehicle,” scripture is a better vehicle than tradition, for “history proves that extirpating any faith based on scripture has been impossible,” whereas tradition-based faiths meet their demise “simultaneously with the breakdown of the state.” Any scripture that “contains . . . the purest moral doctrine of religion” may rightly “claim the authority equal to that of a revelation.” Although different empirical churches can manifest “one and the same true religion,” church-goers tend to confuse “religion” (universally valid inward moral dispositions) with “faith” (contingent, external statutory beliefs), deriding those who affirm other faiths as “unbelievers” and those who interpret the same faith differently as “heretics.”

Having given scripture an honored place in establishing the church, Kant offers hermeneutic guidelines in Subsection VI. Because human beings are embodied, we have a “natural need” for “some historical church faith” as an “experiential confirmation” of moral religion, “if the intention is to *introduce* a faith universally.” To satisfy the human need to particularize the universal, each church must interpret its scripture “to yield a meaning that harmonizes with the universal practical rules of a pure rational religion.” Even if they seem “forced,” when viewed from the standpoint of historical scholarship, meanings based on moral symbolism are more suitable for *church use* than literal ones that carry no moral content. Because “the supreme criterion” for regarding a text as “divine revelation” is that it must foster “moral improvement,” this must also be “the supreme principle of all scriptural interpretation.” Historical scholars best serve the church by attempting to authenticate the authority of an alleged revelation, using their knowledge of the original language and cultural context to establish “that the origin of Scripture contains nothing that would make the assumption of its being a direct divine revelation impossible.” To avoid fanaticism, churches should not accept interpretations based on “*inner feeling*” as authoritative, because we cannot “ascertain any cognition of laws” from personal feelings, but only from the universal feeling of respect for the moral law.

Subsection VII stipulates that the contingent, “experiential cognition” of historical faith can constitute a “true church,” provided it gradually focuses more and more on universal, “pure religious faith.” But another religious conflict arises over what actually saves the members of such a church. A “*saving* faith” must be based on a heart with pure moral dispositions; those who labor to please God through non-moral actions possess “a *slavish* and mercenary faith” that may coexist with an evil heart. What matters is how a person combines the “two conditions for . . .

hope of salvation”: the historical faith that God will forgive one’s guilt, and the moral faith that one can please God by following “a good way of life.” In the transition from historical faith to pure religious faith “a noteworthy antinomy of human reason” arises: (1) if living a good life is a condition for receiving God’s historically conditioned forgiveness, “pure moral faith will have to precede church faith”; yet (2) “if the human being is corrupt by nature,” then one cannot live a good life unless historically-conditioned “faith in a merit which is not his own” *precedes* “all endeavor toward good works.” As a theoretical question, this antinomy surpasses reason’s speculative power, so the only way to resolve it is through practical reason: we know our duty is to do what is good and hope for God’s assistance on that basis; “the theoretical concept” serves only to “*make* the absolution *comprehensible*,” whereas the practical is the “unconditional” basis for hoping to receive any forgiveness God might offer. Church faith views belief in divine forgiveness as a duty and “a good way of life” as a gift of grace, while pure religious faith views “the *good way of life*” as an “unconditional *duty*” and divine forgiveness as “a bare *matter of grace*.” Only by viewing these as two sides of the same coin can we avoid their respective dangers—“superstition” and “naturalistic unbelief” being the typical errors committed by the two types of readers (theologians and philosophers) whose religious conflict Kant is resolving here.

This long subsection concludes Division One by demonstrating that the antinomy of faith can be resolved only by realizing that “in the appearance of the God-man, the proper object of the saving faith is . . . the archetype that lies in our reason,” for they are “one and the same practical idea,” proceeding “in one case insofar as it presents the archetype as located in God” and “in another case insofar as it presents it as located in us, but in both cases insofar as it represents it as the standard for our way of life.” The antinomy remains unresolved only if we take either historical faith or pure religious faith on its own, without seeing each in its necessary relation to the other. Historical faiths tend to make this error by offering relatively easy, non-moral ways to please God; yet no one empirical path will appeal to everyone. Just as a person casts off childish ways when growing up, religion will “be detached gradually from all empirical determining bases,” enabling the kingdom of God to become a reality on earth as churches find their basis in “pure rational religion.” This “practical regulative principle” of transformation to a world wherein all humanity lives “as a community” in “eternal peace” contains, “as in a germ that develops and that will later bear seed in turn, the whole that some day is to illumine and rule the world.”

With pure rational faith firmly established as the proper goal of all historical faiths, Kant turns to his second experiment in Division Two, offering a “Historical Presentation” of a gradual transition to a world dominated by “the Good Principle.” Kant here focuses not on religion as an individual matter, but on the development of *church faith*, starting from its first recognition that pure religious faith lies at its core, and on how the modifications of faith reveal more and more of this core uni-

fying principle. As the Jewish faith set out to be a theocracy, with external statutes given by God, any moral elements within it were accidental: its laws were coercive, relating to specific actions rather than moral principles; far from having a universal outlook, the community *excluded* all other nations from their faith. The true church’s history, therefore, begins with Christianity’s turn to the internal, moral core of all religion, whereby “no statutes at all” are needed and the faith is “valid for the world.” The revolutionary message of “the teacher of the Gospel” (i.e., Jesus) replaced “that slavish faith” with “moral faith.” His reported miracles were not required to authenticate the content of his teaching as pure religious faith; yet they served the purpose of motivating the early adherents to record the story in a “testament,” thus giving rise to a new historical faith. The Roman “scholarly public” scrutinized the Christian faith, but only “after more than a generation,” so we have enough information to assess whether the early Christians experienced genuine moral improvement. Once “Christianity became itself a scholarly public,” the records are clear: “mystical fanaticism,” “blind superstition,” oppressive orthodoxy, and controlling priesthood incited all manner of evils in the name of Christian (historical) faith. These elements, though corrupt, served the purpose of “winning over to the new faith, through its own prejudices, the nation accustomed to the old historical faith,” so that Christianity could become “a universal world religion.”

In light of this overview, Kant claims “the present” is the best period in church history: if we allow “the germ of the true religious faith . . . to develop further and further unhindered,” it will gradually become “the church that unites all human beings forever and constitutes the visible presentation (the schema) of an invisible kingdom of God on earth.” This process has two guiding principles: (1) a modest awareness that revelation is possible, that religion needs *some* historical faith as a vehicle, should lead us to adopt the existing Scripture “as the foundation for church instruction,” without forcing it on people “as required for salvation”; (2) the exhibition of religion’s moral core in “sacred history” should be used to combat passive faith by emphasizing that “true religion” consists “not in the knowledge or confession of what God allegedly does or has done for our salvation, but in what we must do to become worthy of this.” We should interpret Scripture’s account of the goal of this process, in terms of “a visible kingdom of God on earth” arriving after an apocalypse, as a symbol of hope and courage, empowering us to press on toward an ultimate happiness. As Jesus taught, moral people cannot count on happiness now but are more likely to suffer. Our present condition, with good and evil people living together, has moral benefits; but Scripture’s depiction of the end of history, when “the form itself of a church is dissolved,” with good and evil people being ultimately separated, is also appropriate for that stage. Historical faith may never actually cease, we may always need to “look ahead” for the realization of this “beautiful ideal,” yet we must believe “it can cease” in order to highlight “the intrinsic stability of the pure moral faith,” as a kingdom that is “within” each person.

E. FOURTH PIECE: SERVICE AS A HYPOTHETICAL WAY OF PLEASING GOD

The Fourth Piece conveys the final element in Kant's religious system, a principle informing good-hearted people, from the hypothetical perspective, how God desires to be served in and by the church. The introduction reminds us that "the arrangement of this church is incumbent on human beings" and that its goal, the kingdom of God, has begun to appear "if so much as the principles of its constitution start to become *public*." Although "God himself must be the originator of his kingdom . . . , we do not know" how this will happen; yet "we find within ourselves the moral vocation to be citizens and subjects in . . . that kingdom." Once God provides the constitution, "through reason or through Scripture," we must found "the organization" and "manage the public affairs of the church." This organization concerns "only the visible church," because the invisible church has *no officials*: "each member of the community receives his orders directly from the supreme legislator." A true church will be "constantly approaching the pure rational faith," so its officials must "direct their doctrines and regulation always to that ultimate purpose." Officials who instead declare "the historical and statutory part of church faith as alone bringing salvation" confuse the means with the end and promote "pseudoservice," "the persuasion that one is serving someone by actions that in fact undo the latter's intention." Kant attempts to resolve this final religious conflict, between service and pseudoservice, in two parts.

Part One's introduction defines religion as "the cognition of all our duties as divine commands." A *revealed* religion requires me to know "a divine command" before identifying something "as my duty," whereas *natural religion* requires me to know my duty before identifying something as a divine command. While a *naturalist* denies all revelation in contrast to a pure *supranaturalist* who requires it, a *pure rationalist* requires only natural religion but allows for both the possibility and "the necessity of a revelation as a divine means for the introduction of the true religion." The dispute between the latter two is about which claim is necessary and which is contingent. As regards religion's *spread*, in "*natural religion* . . . everyone can be convinced through his reason," but in revealed religion "one can convince others only by means of scholarship." An objectively natural religion may need to be expressed subjectively, as a revealed religion, in order to be "universally communicable." Yet "once the religion thus introduced is there and has been publicly promulgated, everyone can henceforth convince himself of this religion's truth by himself and his own reason." At this point, the pure rationalist thinks any "supranatural revelation" can be "entirely forgotten," whereas the supranaturalist must preserve it, lest religion "disappear from the world." Treating revelation as "a pure concept of reason," Kant surveys the content of the New Testament to "test" Christianity (as "a revealed religion") in Part One's two sections, "first, as a natural religion, and then, second, as a scholarly religion"—a division that corresponds again to his two experiments.

Natural religion, Kant reminds us in Section One, makes universality "the great requirement of the true church"; "the bare, invisible church" thus needs "a ministry" (though not "officials") in order "to spread and be maintained." The required "agreement" is unlikely to become universal "unless the natural laws, cognizable by bare reason, are supplemented by certain statutory regulations" that impose "a special duty" to make "their permanent union into a universal visible church." If a teacher appeared who propounded such a pure religion and who "then added certain statutes . . . as means for bringing about a church that was to be founded upon those principles," without turning them into "new burdensome regulations," that church could be called a "true universal church." The New Testament clearly shows us such a person, "the *founder*," not of the pure *religion* that "is inscribed in the heart of every human being . . . , yet of the first true *church*." Citing Matthew's Gospel extensively, Kant explains why Jesus' teachings are "indubitable documents of a religion as such": Jesus demands that, "not the observance of . . . statutory church duties but only the pure moral disposition of the heart shall be able to make a human being pleasing to God"; he regards those who try "to evade their true moral duty and to indemnify themselves for this by fulfilling the church duty" as misinterpreting the law; instead, he requires pure dispositions "to be proved in *deeds*"; and "he collates all duties (1) in a *universal rule*," the "love God" command, "and (2) in a *particular rule*," "Love everyone as yourself." Virtue is striving to follow such "precepts of *holiness*"; thus Jesus denies any worth to the good deeds done by those who instead passively await "a heavenly gift." While promising happiness as "a reward in a future world" to those who do good "for its own sake," he concedes only minimal benefit to those who do good out of self-interest, "the god of this world." Only the former are "the proper chosen for his kingdom." Jesus introduces "a complete religion" in an "intuitive" form, providing "an archetype to be emulated," yet without appealing to scholarship.

In Section Two, Kant focuses on his second experiment, viewing Christianity not "as a pure *rational faith*" but "as a *revelation faith*": if dogmas not grounded in reason are deemed necessary, then the only way to avoid "a continual miracle of revelation" is to regard the revealed text "as a sacred property entrusted to the care of the *scholars*." Even though it "is built upon facts, rather than upon bare concepts of reason," the Christian faith affirms two types of service, historical (scholarly) and moral (rational)—neither being "self-subsistent." Service that is scholarly cannot be "intrinsically *free*," but tends to require blind obedience. Starting Christian doctrine "from *unconditional faith* in revealed propositions" forces the unscholarly to be slaves to the divine command. To avoid this, scholars must draw "the supreme commanding principle" from natural religion, with "the doctrine of revelation" being merely a means enabling the unscholarly to grasp the concepts. This prepares the way for "true *service*." Pseudoservice arises when "the moral order is entirely reversed and what is only means is commanded unconditionally," because those unable to interpret Scripture must then accept as their duty whatever the scholars and church officials say has been revealed. The officials may *claim* to serve the

church, yet the true outcome is “*domination* of its members.” Christianity’s great advantage over Judaism is that its central message “issued *from the mouth of the first teacher* not as a statutory but as a moral religion.”

Kant introduces Part Two by calling pseudoservice a practical form of “religious delusion,” defined as taking statutes intended for use by a visible church “as essential to the service of God in general.” He examines this theme in two sets of paired sections: the subjective basis of religious delusion and the moral principle opposing it, followed by the governance of “priestery” and the proper governing role of conscience opposing it. Though anthropomorphism “is scarcely avoidable” in presenting God’s nature to ourselves, Kant warns in §1 that a problem arises when we “make a God for ourselves”: we tend to think that what “we do solely in order to please the divinity” can exempt us from acting according to our inner moral disposition. Because such deeds are otherwise useless, people tend to think they “indicate more intensely the unbounded (though not moral) submission to his [i.e., God’s] will.” They “attest dedication to God,” so people attribute to religious *devotion* (i.e., “the mind’s attunement to a receptivity for dispositions dedicated to God”) the worth of moral dispositions themselves. But to think devotion is good in itself, apart from a moral disposition, is a “religious delusion.”

Kant starts §2 by positing a basic principle: “Apart from a good way of life, anything further which the human being supposes that he can do to become pleasing to God is a mere religious delusion.” This does not deny there may be something “only God can do to turn us into human beings pleasing to him”; yet to *require* faith in a revelation of this mystery, as a condition for being good, is worse than slavery, for it contradicts conscience. Reason bids those with “a truthful disposition devoted to duty” to believe their deficiency “will be compensated for by the supreme wisdom *in some way*”; by contrast, “he who absolutely claims to know this manner of redemption of human beings from evil, or, should he not, gives up all hope for it” *lacks* genuine faith, hoping to gain God’s favor without attempting to become good. One who “tenders everything to God, except for his moral disposition,” thinks devoting “his heart to God” means having “a heartfelt wish that those offerings may be received in payment for that disposition.” Those who believe non-moral service is “on its own pleasing to God” are all equal in “worth (or lack of worth),” regardless of how refined may be their appeal to sensuality over “the sole intellectual principle of genuine veneration of God.” For true service, “everything hinges on the acceptance or abandonment of the sole principle, to please God . . . only through the moral disposition.” Some might call this a “dizzying delusion of virtue,” yet a virtuous disposition is “something *actual* that by itself is pleasing to God and harmonizes with the world’s greatest good.” The delusions of fanaticism and superstition, discussed further in the sections on *parerga* as “effects of grace” and works of grace, abandon all principles, causing “the moral death of reason.” The principle that “forestalls all religious delusion” is that all church faith must “bring about the religion of the good way of life as the proper goal.”

In §3, Kant observes that the worship “of powerful invisible beings” can be servile idolatry, legalistic “temple service,” or “church service,” where “the moral education of human beings” is at least partly in view. These differ in “the *manner*, but not in the *principle*, of having faith,” as long as church-goers define “their service of God” in terms of “faith in certain statutory theses” and “contingent observances,” rather than a good disposition. They all wish “to guide to their advantage the invisible power holding sway over the destiny of human beings”; while those who think of God as “an intelligent being” may include moral actions in these observances, they also include non-moral actions “that cannot become familiar to us through bare reason.” Since one of these must be “the supreme condition” of pleasing God, the only rational approach is to assume “actions that have no moral worth in themselves” are “pleasing to God only insofar as they serve as means for the furtherance of what is directly good in actions.” One who thinks non-moral actions can gain “God’s direct pleasure in him” has the delusion of “fetishism,” not essentially different from *magic*, for “the basest human being can perform [them] just as well as the best.” One who performs such observances not to influence God directly, but “to make himself merely *receptive* to the attainment of the object of his good, moral wishes,” does count on some divine assistance to compensate “for his natural incapacity,” yet “not as something *effected* by the *human being* . . . , but as something received, which he can hope for but cannot produce.” What matters most “when one wants to link two good things” is “the order in which one links them!” Recognizing that moral duties burden the conscience far less than non-moral religious observances (“because everyone sees on his own the necessity of complying with them, and hence nothing is thereby thrust upon him”) constitutes “true *enlightenment*,” whereby “the service of God . . . becomes for the first time a free and hence moral service.”

In the second half of §3, Kant defines *priestery* as “the constitution of a church insofar as a *fetish service* governs in it.” If the church’s “essential feature” reverses the proper order of direct and indirect service of God, then no matter how few “imposed observances” are deemed necessary, “the multitude is governed and is robbed, through the obedience to a church (rather than to religion), of its moral freedom.” Whatever political form “the hierarchy” uses, “its constitution . . . is and remains always despotic.” When a clergy becomes “the only authorized guardian and interpreter of the will of the invisible legislator,” both reason and scholarship (the conflicting standpoints highlighted in the Prefaces) tend to be ignored. Such priestery presents the message of faith in a simple form that is easy to communicate even to the foolish; but for “scholars and philosophers,” “accepting such a faith, which is subjected to so many controversies . . . , as the supreme condition of a universal and only saving faith is the most paradoxical thing that can be conceived.” By contrast, “the law of morality” is equally accessible to everyone, carrying “*unconditional* obligation in everyone’s consciousness” and “leads even on its own to faith in God.” Hence, prudence directs us to make this pure religious

faith “the supreme condition under which alone we can hope to come to partake of whatever salvation a historical faith might promise us”; however, “the person of moral faith” should be “open to the historical faith,” provided it invigorates “his pure religious disposition.” Religious teaching should reflect this by presenting “the *pure doctrine of virtue*” as the supreme condition of a moral disposition, and “the *doctrine of godliness*” as the means to acquire it, for the two “obviously stand in necessary connection with each other.” Here “everything hinges on the supreme concept to which one subordinates one’s duties”: if godliness comes first, the God worshipped “is an *idol*” and worship is a “vain endeavor” to become good merely “from wishing”; if virtue comes first, godliness can strengthen one’s “courage to stand on one’s own feet.”

The antidote to the false governance of priestery, Kant argues in §4, is to allow “conscience” to “serve as guiding thread” in all moral decisions. Whereas practical reasoning tells us the *content* of right and wrong, conscience is “the moral power of judgment passing judgment on itself,” telling us whether we have “actually undertaken” moral reasoning. It is self-authenticating “consciousness that is by itself a duty.” A basic “postulate of conscience” is that, for any action “I want to undertake I must . . . be *certain* that it is not wrong.” For example, an inquisitor who condemns a heretic to death for “unbelief” exhibits “lack of conscience”: even if obeying a command that seemed to come from God, he “could never have been entirely certain that he was not perhaps thereby doing wrong.” Yet moral reason tells us it is wrong to “take a human being’s life because of his religious faith.” Similarly, a cleric who compels “the people to confess as an *article of faith*” a revealed doctrine, even though the doctrine might be erroneous, is “proceeding against his conscience”—a “damnable” form of inward “untruthfulness.” People “under a slavish yoke of faith” adopt the principle “that it is advisable to have faith in too much rather than in too little,” since doing extra “can at least do no harm”; yet this “maxim of safety” is a delusion that turns “insincerity in religious confessions into a principle.” Only a hypocrite feigns certainty of something “even before God,” despite being “conscious that it is not of the character required to affirm it with unconditional confidence.” Settling this religious conflict with a clear conscience demands a “genuine maxim of safety”: I cannot know with certainty the truth *or falsity* of any aspect of historical faith that “does not contradict the pure moral principles,” yet I can “count on” any promised assistance, provided I do not “make myself unworthy of it through the deficiency of my moral disposition in a good way of life.” Surely, “even the boldest teacher of the faith would tremble” if asked, “in the presence of him who knows the heart,” to “affirm the truth of [revealed] propositions” with such absolutely certainty as to say that if any of them ends up being false, “*then let me be damned!*” Such presumptuousness stands in sharp contrast to the person of good will who in all sincerity exclaims: “I believe, dear Lord; help my unbelief!”

F. FOUR GENERAL COMMENTS: THEOLOGICAL SUPPLEMENTS TO RATIONAL RELIGION

Kant’s changes in the second edition mainly highlighted his second experiment, introducing it in the second Preface and appending to the First Piece’s General Comment an overview of the four parerga of rational religion: “(1) On Effects of Grace, (2) On Miracles, (3) On Mysteries, (4) On Means of Grace.” These theological ideas are not constitutive elements of rational religion, yet they arise in direct response to reason’s inability to solve the moral problems raised by our evil nature; they “abut” or *press against* reason’s boundary, offering to clothe the bare body of rational religion with useful concepts—especially for Christians, for whom the second experiment is conducted. These ideas can be accepted with a “reflective” faith, as compensations for reason’s moral and theoretical incapacity, yet if we presumptuously adopt them as articles of a “*dogmatic* faith, which proclaims itself to be a *knowledge*,” they will result in fanaticism, superstition, deluded claims to be illumined, and attempts to control God.

The first parergon is the belief that moral goodness “will be not our deed, but that of another being, and therefore that we can *procure* the effect of grace by just *doing nothing*.” As Kant explains at the end of the first General Comment, the belief that divine grace can magically change our inner disposition provides no theoretical knowledge of such an effect, “because our use of the concept of cause and effect cannot be expanded beyond objects of experience and hence beyond nature.” Moreover, this idea “is entirely self-contradictory” for practical reason, because if we do *nothing* to receive such a free gift of goodness, then it cannot make us genuinely good. Rational religion’s proper response to this theological idea, therefore, is to “grant the effect of grace as something incomprehensible,” without admitting it as a core element of religion that can determine how we form moral maxims. We can thus take advantage of the empowering idea of grace without falling into the delusions of fanaticism discussed in the Fourth Piece.

The second parergon is a belief in miracles. Cultic religion depends on miracles for its authority by promising to make people pleasing to God without requiring them to become good. Miracles “cloak” moral religion when it first appears, thus attracting converts. Yet, by insisting that only a good disposition can please God, moral faith renders “dispensable the very faith in miracles as such.” To grant “that miracles had occurred *of old*” while disallowing “*new* miracles” is not a total denial that miracles still happen, for we cannot have “objective insight” into their existence; instead, it is “only a maxim of reason” encouraging people not to *depend* on miracles. In *theistic* miracles, God acts through the natural and moral laws that already govern the world, so calling such events miracles adds nothing to our understanding. But if an *apparent* divine command tells us to break the moral law, we may safely assume it did not come from God, otherwise reason would be “as though paralyzed.” *Demonic* miracles are even less rational: we cannot assume a

command is *not* from a devil simply because it appears to be consistent with morality, for “the evil spirit often disguises himself as an angel of light.” Professionals such as clerics and doctors should not encourage people to appeal to miracles but should focus on “the moral improvement of the human being”—a goal that is impeded by the “senseless notion” of a “theoretical faith in miracles” that would allow us to “assail heaven.”

The third parergon is belief in *holy mysteries*; they remain theoretically useless, since we cannot “establish a priori and objectively” whether they exist, yet they can be “sufficiently cognized inwardly for practical use” if we appeal to our moral predisposition, especially our capacity for *freedom*, to interpret them. That we have a duty to work toward the highest good, yet cannot achieve it on our own, opens up an “abyss of a mystery of what God is doing” to assist us. Complementing our theoretical understanding of God’s nature, practical reason portrays God as “*holy legislator*,” “*benign governor*,” and “*just judge*.” This is not a mystery, provided “it expresses solely God’s moral conduct toward humankind” as contained in “the concept of a people as a community” (not “what God is in himself”). Similar concepts of God appeared in many world religions but remained mysterious until the Christian faith “first put forth publicly to the world” the doctrine of the Trinity, viewing God not as a human sovereign, mixing moral qualities in one personality, but as “one and the same being” with three distinct personalities: God’s laws are not arbitrary but directed toward holiness; God’s goodness is not unconditional but considers people’s “moral constitution” before compensating “for their inability to fulfill this condition on their own”; and God’s justice is neither unduly benign nor demanding of holiness, since humans are capable only of virtue. Moral reason helps clarify three other mysteries: (1) “the *calling*” of free beings “to citizenship in the divine state” is morally clear, yet we cannot understand how free beings are *created* (i.e., made subject to causal laws); (2) God’s *satisfaction* of the moral demand on behalf of imperfect humans is morally necessary, yet reason has no insight into how it happens; and (3) the *election* of only some people to receive the grace to adopt a “disposition pleasing to God” may be wise, yet seems unjust to human reason. As with freedom, “we still would not *understand*” such mysteries theoretically, even if God revealed their causes; yet we know just enough to inform our conduct. Therefore, demanding additional revelation is immodest. Biblical scholars and philosophers should agree on how to interpret these mysteries, because a “literalist faith more readily corrupts rather than improves the true religious disposition.”

The fourth parergon is a belief that special religious deeds control how God distributes *grace* (i.e., divine assistance in becoming good, beyond what we can do through our free choices, or “nature”). Such support for our “weak disposition to fulfill all our duty” is a transcendent idea that reason can neither prove nor disprove, whereas we know the moral laws that define “the good way of life.” To remain rational and avoid “passive idleness,” we may assume “that whatever nature is not capable of will be brought about by grace.” But a person’s only means of

pleasing God is “an earnest endeavor to improve as far as possible his moral constitution”; only this makes a person “receptive to the perfecting” influence of grace, “which is not in his power.” Treating *non-moral* deeds as a “means of grace” is self-deception, true service of God being an invisible “service of hearts.” “Yet for the human being the invisible needs to be represented by something visible (sensible); even though such “intuitive” ways of serving God are typically misinterpreted, reason can discern their “spirit and true significance” by dividing them “into four observances of duty,” useful “for arousing and sustaining our attention to the true service of God.” Thus, (1) private prayer *establishes* goodness “firmly *within ourselves*,” arousing “in our mind the disposition concerning it”; (2) churchgoing *spreads* the good “through public assembly” so people can hear and communicate “religious doctrines and wishes (and with them dispositions of this sort)”; (3) baptism *propagates* “this good to posterity” by giving an occasion to instruct “newly joining members into the community of faith”; and (4) communion *preserves* “*this community* through a repeated public formality which makes continuous the union of these members into an ethical body.” Nevertheless, such formalities become “a *fetish faith*” if one thinks they will persuade God to “satisf[y] all our wishes”—as if God would accept “*the manner* (the formality) . . . for the deed itself.”

After examining each of the four means of grace,⁸ Kant concludes the fourth General Comment, the Fourth Piece, and the whole book by correlating the “three kinds of delusory faith” (miracles, mysteries and fetishes) to the “three divine moral properties—holiness, grace, and justice,” rooting them in the human desire to win favor by mixing these three qualities instead of keeping them distinct. The false hope that “deedless wishes may serve also to compensate for the transgression of [God’s moral] commands” leads believers to focus on “*piety* (a passive veneration of divine law) rather than . . . *virtue* . . . , even though only virtue, *combined with piety*, can constitute the idea that one means by the word *godliness* (true religious disposition).”

4. IMPLICATIONS AND ENDURING RELEVANCE OF KANT’S *RELIGION*

The hope Kant expressed in the last sentence of the Preface to the second edition of his *Critique of Pure Reason* applies equally to *Religion*: “if a theory is internally stable, then any action and reaction that initially portend great danger will in time serve only to smooth away the theory’s unevenness; and in a short time they will even provide the theory with the requisite elegance, if those who deal with it are

⁸ In Kant’s detailed account of the four means of grace (*Religion*, 194–200), his aim is twofold: to explain how each observance can strengthen a person’s moral disposition, yet to warn against misusing it as a deluded way of trying to control God. For an interpretation of Kant’s account of prayer along these lines, see “Kant’s Critical Hermeneutic of Prayer,” Appendix VIII in *Kant’s Critical Religion*.

men of impartiality, insight, and true popularity.” Unfortunately, the immediate reaction to *Religion* was far from welcoming: not only did the censor respond by obtaining a royal order preventing Kant from making public statements on religion, but many early readers simply failed to grasp its key insights. Goethe, for example, wrote to Herder in a letter of June 7, 1793: “Kant required a long lifetime to purify his philosophical mantle of many impurities and prejudices. And now he has wantonly tainted it with the shameful stain of radical evil, in order that Christians might be attracted to kiss its hem.”⁹ Friedrich Schleiermacher (1768–1834) likewise found little insight in *Religion*: while accepting many of the fundamental tenets of the Critical philosophy, he rejected the basic components of Kant’s approach to religion, including the moral argument for God’s existence and morality as the core of religion, opting instead for *feeling* as the core. All in all, *Religion* had relatively little immediate impact on scholarly discussions of religion and theology in Kant’s day.

Schleiermacher, typically regarded as the father of modern liberal Protestantism, has dominated theology over the past two centuries. The tendency to portray him as theologizing in the spirit of Kant, however, ignores the fact that he oversteps many of the bounds Kant established in *Religion*, much as did Kierkegaard after him. The liberal theologian, Albrecht Ritschl (1822–1899), more explicitly aimed to present a Kantian version of liberal Christianity grounded in the distinction between “facts” and “values.” Karl Barth (1886–1968), in his *Protestant Theology in the Nineteenth Century*, traces Kant’s influence on many theologians, including his own neo-orthodoxy. The popularity of the “quest for the historical Jesus” in the nineteenth century was partly due to Kant’s call for “anthropological investigation” (*Religion*, 25). Similarly, his emphasis on symbols as the key to religious meaning has obvious affinities with the ideas of Paul Tillich (1886–1965),¹⁰ whose theology was largely a synthesis of Kant and Kierkegaard, though he was reluctant to affirm Kant’s influence. Among Catholic theologians who were deeply influenced by Kant’s religious thought, Joseph Maréchal (1878–1944) and Karl Rahner (1904–1984) stand out—though the list could go on and on. Most recently, as Kant’s philosophy has been interpreted in more and more consistent and holistic ways, a trend has arisen among various more conservative theologians and philosophers to interpret Kant as surprisingly amenable to the concerns of ordinary Christians. Whether this can be taken as far as the recent “affirmative” interpreters have suggested remains to be seen.¹¹ What is clear from the overview in §3 is that no

⁹ Quoted in Emil Fackenheim, “Kant and Radical Evil,” *University of Toronto Quarterly* 23 (1954): 340.

¹⁰ For an excellent account of the Kantian roots of Tillich’s theology as well as other twentieth-century developments, see Adina Davidovich, *Religion as a Province of Meaning: The Kantian Foundations of Modern Theology* (Minneapolis, MN: Fortress Press, 1993).

¹¹ See e.g., Chris L. Firestone and Stephen R. Palmquist (eds.), *Kant and the New Philosophy of Religion* (Bloomington: Indiana University Press, 2006), especially the Editors’ Introduction.

reading of *Religion* can be adequate without taking account of both Kant’s readiness to accommodate real religious beliefs and practices, and his severe criticism of any approach to religion that views these as anything but temporary manifestations of a deeper, moral reality.

The orientation of Kant’s position is neither clearly liberal nor clearly conservative. To highlight his deep awareness of the inherent conflict in the human experience of religion, I have elsewhere characterized his position as a “conserving liberalism” (for the philosophical reader) and a “liberating conservatism” (for the religious reader). Kant’s unwillingness to succumb to either of the two traditional theological extremes is, as suggested in §1, rooted in his childhood experience of religion. Kant refers to the highly personal source at several points in *Religion*, perhaps most notably in a footnote on p.190, where he describes in the first person:

. . . our manner of upbringing, above all on the point of religion, or, better, of doctrines of faith, where faithfulness of memory in answering questions concerning them, without regard for the faithfulness of the confession . . . , is accepted as already sufficient to produce a person of faith, who does not even understand what he affirms as being holy, and one will no longer be surprised by the lack of sincerity which produces nothing but inward hypocrites.

Here Kant lets slip through the cracks of his philosophical armor that *Religion* can be read as a religious autobiography, a detailed account of how the Critical philosophy helped him resolve (or cope with) the religious conflicts of his youth. Likewise, in a note on p.184, Kant critiques what he sees as the (nonessential) “temperament” of Christians

. . . . in *piety* (by which is meant the principle of a passive conduct regarding a godliness to be expected, from above, through a power); for they never posit a confidence in themselves but look around in constant anxiety for a supernatural assistance, and even in this self-contempt (which is not humility) suppose themselves to possess a means of gaining favor, of which the outer expression (in pietism or affected piety) proclaims a *servile* cast of mind.

This is Kant’s mature observation on the shortcoming of his childhood teachers. True piety, by contrast, “consists not in the self-torment of a repentant sinner . . . but in the firm resolve to do better,” and *empowers* a person by producing “a cheerful mental attunement” (160). The whole book, as we saw in §3, attempts to rescue religious faith from a fate worse than oblivion: the delusions and false practices prompted by our natural propensity for self-deception.

The reductionist interpretation of *Religion* that tended to dominate English Kant scholarship during much of the twentieth century, whereby Kant is read as virtually identifying religion with morality so that everything in the former is

already included in the latter, fails to take into account not only the depth of Kant's own religious disposition, but also the way he employed the clothing metaphor (as implied in the title) throughout the book. Such moral reductionism (see note 1) rightly stresses that one who sees religion in *reason's* light has no pure or *philosophical* need for any non-moral cloak. Yet we do need such clothing once we take into account the tendency toward self-deception and moral weakness that besets our nature as a direct result of our evil propensity. The reductionist reading assumes Kant wrote *Religion* only for philosophers. Yet, as we have seen, the book is filled with evidence that he wrote for two kinds of readers: those who, like Kant, might be called "recovering fundamentalists," having called into question many of the religious dogmas of their youth, but do not wish to uproot the grain of truth it planted in them; and for those who, like the antireligious *Aufklärer* among Kant's fellow philosophers, are pessimistic about the possibility of empirical religion ever being of practical use. Kantians who shun all religion are bound to be challenged by Kant's concessions to religious believers. They may choose to adopt religious convictions only privately (as Kant did, preferring in his mature years not to darken the door of his local church), but a balanced reading of *Religion* will force them to admit that their philosophical hero supports neither a dogmatic atheism nor an agnostic denial of all legitimacy to more openly religious ways of life.

Kant remains mostly silent on his personal religious beliefs because *Religion* is, from start to finish, a work of *philosophical* theology, not biblical theology, and he does not think philosophy is capable of specifying a *particular* historical vehicle as being necessary to fill the gaps left by reason. While philosophy can expose the "naked body" of reason, the bare *rational form* that can be used to test the validity of any historical religion, it cannot determine in advance which clothing will fit that body most appropriately at any given time. The First Piece (especially Section IV) thus assesses the rational stability of the Christian doctrine of original sin: its typical historical-hereditary interpretation is "inappropriate"; but if understood as referring to the *rational* origin of all evil, it is a perfectly acceptable account of the practical (moral) problem that cries out for a religious solution (i.e., the problem of the evil propensity in human nature). Likewise, the Second Piece assesses the Christian doctrine of grace through Jesus' atoning sacrifice as morally harmful if it is presented in a way that removes the believer's duty to effect his or her *own* self-improvement, but "completely valid" (66), "holding also as a precept to be followed" (64), if understood as depicting an internal rational space (the archetype of a perfect human being) that each of us should strive to imitate in our moral conduct. As we have seen, the book's second half offers equally "Critical" assessments (combining the theologian's interest in what goes beyond reason with the philosopher's interest in preserving reason as the core) of the doctrines of the church and its service to God, respectively.

Despite this book's depth of insight and wide-ranging influence on two centuries of theology, its central aim has only recently begun to be widely recognized: Kant is seeking, at one and the same time, to provide followers of literalistic,

fundamentalist religion with a viable alternative that allows them to maintain their faith with a high degree of philosophical sophistication *and* to persuade philosophers not to decry all religion as an irrational or futile endeavor but to affirm its ultimate significance for the development of the human race. Whether one thinks this twofold aim destroys or upholds the essential features of an empirical religion such as Christianity will surely depend on how one conceives of that religion. Forming a clear conception of genuine and illusory approaches to religion could not be more crucial than it is today, in a post-9/11 world where nobody can ignore the vastly different ways that followers of one and the same historical faith view religion. Kant's comparatively shallow comments on traditions such as Judaism, Islam, and Eastern religions may not carry the same depth of insight as his perceptive analysis of Christianity; yet followers of any faith can draw from this text important lessons regarding the dangers of religious illusion that apply equally to their own tradition. In the postmodern world, Kant's talk of universal "human destiny" may seem far-fetched; yet his own experience of religious conflict beckons us to regard the tensions and paradoxes of this text not as signs of weakness or insecurity, but as an authentic awareness of the conflicts of human existence that any modern reader who is intellectually, morally, and spiritually honest is bound to experience. Liebmann's well-worn aphorism, "You can philosophize with Kant, or you can philosophize against Kant, but you cannot philosophize without Kant," surely applies nowhere more poignantly than to his philosophy of religion, making Kant's *Religion* essential reading for any twenty-first century reader who wants to understand how best to clothe the bare body of religion's essential nature for the betterment of human beings.¹²

¹² I would like to thank Chris Firestone, Guy Lown, Richard Palmquist, Werner Pluhar, and Philip Rudisill for providing helpful feedback on earlier drafts of this essay.