

The Original Goodness of Human Nature

Introduction, Comment, and Section I (R 19–28)

1. Untitled introduction: Is humanity good or evil by nature?

The main text of Kant's *Religion* starts with an untitled introductory section. Despite its unassuming placement, this section, which appears to be little more than an additional preface, establishes a crucial context for the whole First Piece. The section (and with it also the main body of the book) begins with a paraphrased allusion to a verse from the Bible, followed by an interpretation highlighting several themes that appear regularly throughout the book.

19.01–16

FIRST PIECE

ON THE INHERENCE OF THE EVIL ALONGSIDE THE GOOD PRINCIPLE: OR ON THE RADICAL EVIL IN HUMAN NATURE

That the world is in bad /shape¹ is a lament as ancient as history, even as the still more ancient art of poetry—indeed, just as much so as the most ancient among all poetic expressions,² the religion of the priests. Nonetheless,³ all [of these traditions]⁴ have the world start from the good: from the Golden Age, from life in Paradise, or from a still more fortunate⁶ life in communion with heavenly⁵ beings.

¹ WP has “lies in baseness” for Kant’s *im Argen liege*; GH and GG have “lieth in evil,” putting the whole phrase in quotation marks, on the assumption that Kant is quoting 1 Jn. 5: 19b: “the whole world is under the control of the evil one.” But Luther’s translation reads *die ganze Welt liegt im Argen*, and Kant’s rewording expresses a German idiom. As occurs throughout *Religion*, Kant is alluding to LB but amending it (probably from memory), often in order to make a specific point. Here his point is that the biblical lament reflects common sentiment.

² WP has “poetic compositions” for Kant’s *Dichtungen* (cf. note 8.40 in this volume); GH and GG have “fictions,” but Kant has just referred to poetry (*Dichtkunst*); GG renders the latter loosely, as “poetic fiction,” thus preserving the relation between cognates, but with the wrong English word. GH appeared in the wake of Vaihinger’s 1925 book *The Philosophy of As If*; and, as di Giovanni 2012 confirms, the traditional reductionist interpretation of *Religion* still influenced GG heavily.

³ WP has “Yet” for Kant’s *gleichwohl*; GG and GH omit this word.

⁴ WP inserts “religions” here, while GH and GG allow Kant’s “all” to stand alone. I read Kant as making a statement not about all religions, but about all of the specific traditions referred to above: history, poetry, and priestcraft.

⁵ Following GG for Kant’s *himmlischen*; WP and GH have “celestial.”

But they soon make this *fortune*⁶ vanish like a dream, and *they*₁ now make the *lapse*⁷ into evil (the moral /aspect\), with which *the* physical /aspect\⁸ always *went in the same pair*⁹ hasten toward the [ever] *worse*¹⁰ in an accelerated fall—so that now (but this Now is as ancient (*alt*) as history) we live in the final age, the Last Day and the end of the world are knocking at the door,

Wasting no time, Kant reveals in the book's opening sentence that one of his main themes will be the "poetic" character of religious power structures, or "the religion of the priests." As we shall see in examining the Fourth Piece, when Kant calls such religion *poetic* he does not mean it is meaningless or promotes untruth any more than poetry itself or history does; rather he means that any truth found therein has a symbolic, hypothetical status, inasmuch as we human beings are the source of any meaning that resides in it (see *PSP* IV.3–4 and *PCR* V). The lament Kant borrows from 1 John 5: 19b says nothing about priestcraft; rather the context (1 John 5) focuses on the inward certainty that every genuine religious believer can obtain. Such certainty is not contrary to what might be called "poetic truth," for it is the final goal of symbolic or hypothetical reasoning.

All three of the disciplines Kant mentions in the first sentence (history, poetry, and priestcraft) share such a hypothetical character. Kant likewise notes the consonance between pagan, Greek and Hebrew myths about the origin of the world: they all hypothesize an original goodness at the foundation of the world. Yet, as Kant adds in the third sentence, they are equally univocal in depicting this original, dream-like state as giving way almost immediately to corruption and evil (both moral and physical) that is so serious that, for as long as we have recorded our own history, we human beings have pictured ourselves as being on the brink of final destruction.

In support of the common tendency to regard this "fall" as an ever-accelerating moral decline, Kant adds a footnote to that word, quoting three Latin verses from Horace's *Odes*:

19n.24–26

Aetas parentum peior avis tulit
Nos nequiores, mox daturos
Progeniem vitiosiore.¹¹

By citing an ancient text to support the notion that each generation tends to view itself as worse than the previous and better than the next, Kant shows that this tendency is nothing new. Not only is such pessimism "as ancient as history," but it also crosses cultural barriers, as Kant further emphasizes in the main text by referring to an alleged tendency among some Hindus to view the destroyer god Shiva as the most powerful component of their divine trinity.

⁶ WP, GG, and GH have "happier [or "more happy"] ... happiness" for Kant's *glücklichern ... Glück*; but see Glossary for the distinction between *Glück* ("fortune"), *glücklich* ("fortunate" or "happy"), and *Glückseligkeit* ("happiness").

⁷ WP and GG have "decline" for Kant's *Verfall*; GH has "Fall."

⁸ WP and GH have "moral evil, with which physical evil," for Kant's *das Moralische, mit welchem das Physische*; GG has "moral evil, with which the physical."

⁹ WP has "has ... gone hand in hand" for Kant's *zu gleichen Paaren ging*; GG and GH have "went hand in hand." But Kant's phrase is less metaphorical, not necessarily implying the kind of *cooperation* connoted by the English metaphor.

¹⁰ WP has "ever baser" for Kant's *zum Ärger*, rather confusingly putting the phrase *before* "hasten"; GH has "from bad to worse" and GG omits this phrase.

¹¹ Horace, *Odes*, 3.6. As a more readable alternative to Martin's archaic translation (quoted by GH) and to the literal translations of GH and WP, I offer this somewhat looser version, which more clearly expresses the gradual moral degradation that Kant has in mind: "Having been worse than their parents' generation / Our parents gave birth to us, still worse, / That we in turn might beget offspring more vicious than ourselves."

19.16–20

and in several regions of India¹² the World Judge and Destroyer *Rudra* (otherwise also called *Shiva* or *Siva*) is already being venerated as the god now reigning, after the World Preserver *Vishnu*, weary of the office he took over from the World Creator *Brahma*, has already resigned it centuries ago.

Putting aside questions regarding the historical accuracy of Kant's sources for such comments (cf. note 1.12), we can appreciate this passage as foreshadowing his later emphasis on the Trinity (see App. III); it also illustrates how *Religion* draws examples from a wide variety of religious traditions other than just Christianity.

In contrast to the various traditions that view the course of world history pessimistically, Kant now calls attention to a more optimistic tendency, characteristic of Enlightenment thinkers, who believed that education can reverse evil's influence.

19.21–23, 20.01–03

More recent, but far less proliferated,¹³ is the opposite, heroic opinion, which has probably¹⁴ found its place solely among philosophers and, in our times, above all among pedagogues: that the world advances incessantly {20} (though scarcely noticeably) in precisely the reverse direction, namely from the bad (*Schlechten*) to the better; [or] that at least the predisposition to this [[advance]] is to be found in human nature.

Here we find Kant's first use of the technical term "predisposition" (*Anlage*), though not yet employed in his own special sense. As we shall see, Kant is preparing to propose a theory that aims to account for both views: the enduring negative influence of evil throughout history as well as the presence of something in us that can counteract that influence. But, at this early stage in his argument, he has only hinted that both views have a measure of truth, the former being confirmed by human experience while the latter is grounded in a less tangible, "heroic opinion."¹⁵ Regarding the latter, he adds:

20.03–10

This opinion, however, they have assuredly not drawn from experience if the *moral aspect of* good or evil (rather than the civilizing [of people]) is at issue: for there the history of all times speaks far too mightily¹⁶ against them. Rather, it is presumably merely a well-meaning presupposition of the moralists, from *Seneca* up to¹⁷ *Rousseau*, [made] in order to impel us to cultivate indefatigably the germ (*Keimes*) for the good, which perhaps lies within us, if only one can count on such a natural foundation for the good in the human being.

In other words, human experience and history tend to support the pessimistic view of human nature as mired ever more deeply in evil, even though philosophers and moral educators may

¹² Following GH and GG for Kant's *Hindostan*; WP has "Hindustan." For interesting suggestions as to Kant's likely sources for his claim about Hinduism, see *BRR* 166–7, 249 and *GG* 457.

¹³ WP and GG have "widespread" for Kant's *ausgebreitet*; GH has "prevalent."

¹⁴ WP has "—I suppose—" (after "which") for Kant's *wohl*; GH has "indeed," while GG omits this word.

¹⁵ WP 19n and GG 457n both note that Kant refers to a similar "heroic faith in virtue" in *EAT* 332, published in June 1794. Although Kant quotes from Horace in both contexts, the *EAT* passage contrasts two types of religious imagery (heaven and hell), not two types of philosophical opinion, as here. Among the "pedagogues" who hold such a position, Kant is surely thinking of Johann Bernhard Basedow (1724–90), whose book (Basedow 1770) Kant had used as the textbook for his own lectures on education until 1780, when his friend Friedrich Samuel Bock (1716–85) published a better one (Bock 1780). For a thorough discussion of Kant's treatment of these optimistic and pessimistic options, see *BRR* 165–75; Bohatec traces Kant's position here back to the influence of Seneca, Rousseau, Hume, Ferguson, Shaftesbury, and Hutcheson.

¹⁶ WP and GG have "powerfully" for Kant's *mächtig*; GH has "loudly."

¹⁷ WP, GG, and GH have "to" for Kant's *bis zu*.

tend to prefer a more optimistic view of human nature as capable of cultivating ever-increasing goodness. Kant's argument will show that religion arises out of precisely this tension between experience and reason.

Kant concludes his second paragraph of this untitled section by introducing a hypothetical analogy, which suggests that nature might come to our aid in cultivating this inner seed or "germ for the good."¹⁸

20.10–17

Add to this, moreover, that since the human being must, after all, be assumed to be by nature (i.e., as he is usually born) sound in body, there is no reason for not assuming him to be by nature sound and good just as much¹⁹ in soul as well. Thus nature itself, [[so the moralists conclude,]] promotes the training²⁰ in us of this moral predisposition to the good. *Sanabilibus grotamus malis, nosque in rectum genitos natura, si sanari velimus, adiuvat*,²¹ says Seneca.

One who considers the infant mortality rate in Kant's day might question the accuracy of his analogy between the health of the body and the health of the soul at birth. But Kant's point seems to be more about the *moment* of birth (or even the moment just before it) than about the prospects of survival once the newborn inhabits the world outside the womb. Just as a prenatal infant's health comes naturally, through its connection to the mother rather than through its own effort, so the human soul, considered (hypothetically) *prior to all experience*, is naturally grounded in unchosen goodness. The analogy conveyed by Seneca's maxim provides hope to those who call upon that inner predisposition even *after* evil choices corrupt us: just as our body's diseases are best cured by letting the inborn processes of nature effect a healing from within, so our soul's tendency to be infected by evil desires is best counteracted by calling upon the natural power of the soul's inborn goodness.

At this point the hope offered to this second class of people stands as nothing more than a hypothetical opinion. To test its validity, Kant's third paragraph considers two other, "intermediate" alternatives.

20.18–22

But since, after all, it could certainly have happened that people (*man*) erred in both [these] supposed²² experiences, the question is whether a mean²³ is not at least possible, namely that the

¹⁸ Kant explicitly employs a "seed" or "germ" metaphor twelve times in *Religion*, alluding to it many other times. He appeals to the same metaphor in the final paragraph of *WIE* (41), where he applies it to the church (see note 7.84 in this volume). *BRR* 225–6 traces Kant's theory of the human predisposition (*Anlage*) as a "germ" of the good back to Iselin's *History of Mankind*, first published in 1764, which had gone through numerous editions, including one of 1791, published shortly before Kant wrote *Religion* (see Iselin 1791).

¹⁹ WP has "even" for Kant's *eben so wohl*; GH has "similarly" and GG "equally." For smoother English, I have also changed the position of "by nature."

²⁰ WP has "further the development" for Kant's *auszubilden ... beförderlich*; GH has "(is) inclined to lend ... aid to developing" and GG "would ... be promoting the cultivation."

²¹ As the translators all point out, this quotation from Seneca's treatise *On Anger* (*De ira* 2.13.1) appeared on the title page of Rousseau's *Émile*. Rousseau's original reads: *Sanabilibus grotamus malis, ipsaque nosque in rectum genitos natura, si emendari velimus, adiuvat*. Kant slightly changed it in two places, by omitting *ipsaque* ("itself," said of *natura*) and by replacing Seneca's *emendari* ("be improved") with *sanari* ("be cured"); and he emphasized Seneca's *in rectum genitos*. Kant's own version translates: "We suffer from curable ills; and, if we are willing to be cured, nature helps us, *who are born right*." WP 20n points out Kant's modifications of Seneca's original by giving a composite version.

²² WP's "alleged" (for Kant's *angeblichen*) can have negative connotations. GH and GG ignore this word's literal meaning and replace it with "reading of" and "ways of reading," respectively. But Kant is referring here to the supposition that we *actually experience* progress, either from bad to worse (indicating a fundamentally evil nature) or from bad to better (indicating a good nature).

²³ WP has "something immediate" for Kant's *ein Mittleres*; GH and GG have "a middle ground."

human being, in his genus, may be neither good nor evil, or as the case may be,²⁴ the one as well as the other, partly good, partly evil. —

These hybrid options—that the natural moral state of human beings is either neutral or mixed—are appealing because they seem to fit what we actually experience. Yet Kant goes on to dismiss both, making a distinction familiar to readers of his Critical writings (and mentioned in R 14), between the external *legality* of an action's results and the internal *morality* of an agent's intentions, as revealed by the maxims (or rules of action) adopted in the process of deciding how to act.

20.22–29

We call a human being evil, however, not because he performs actions that are evil (unlawful), but because they are so made-up,²⁵ as to allow /one\ to infer evil maxims in him. Now, through experience one can indeed notice unlawful actions, [and] also (at least in oneself) that they are consciously unlawful; but one cannot observe the maxims, not even always in oneself, [and] hence the judgment that the agent is an evil human being cannot with assurance²⁶ be based on experience.

That we are unable to see into the depths of another person's intentions, and often cannot even accurately detect our *own* intentions, is a point Kant reiterates throughout *Religion* and elsewhere.²⁷ Here he mentions it as the first premise in his argument that a philosophical answer to the question of whether human nature is good or evil cannot be based on experience.

Kant's preferred alternative is to adopt a transcendental (synthetic a priori) approach to the question at hand.

20.30–34

In order to call a human being evil, therefore, one would have to be able to infer a priori from several consciously evil actions, [or] indeed from a single /one\, an evil maxim lying at their basis, and from it again a basis, itself in turn a maxim [and] lying in the subject universally,²⁸ of all particular morally evil maxims.

²⁴ WP has “perhaps” for Kant’s *allenfalls*; GH has “at all events” and GG “at any rate.” To avoid ambiguous overlaps of word usage, I translate *allenfalls* either as here (when options are implied) or as “in any case,” unless otherwise noted.

²⁵ WP and GG have “constituted” for Kant’s *so beschaffen*; GH has “of such a nature.” See Glossary for a defense of my use of “made-up” and “make-up” for the verb and noun forms of this word.

²⁶ Following WP for Kant’s *mit Sicherheit* (literally “with safety”); GH has “with certainty” and GG “reliably.” Cf. the “maxim of safety” that Kant discusses in R 188–189.

²⁷ See, e.g., *CPrR* 93–4, R 51, and *MM* 447. Neiman 1994: 131 argues that such ignorance functions as a necessary condition of human freedom. As Kant explains in some detail (*GMM* 407–9), even though people’s *actions* usually conform to duty, “if we look more closely at the intentions and aspirations in them we everywhere come upon the dear self, which is always turning up” (407)—so much so that the “cool observer” of human interactions may legitimately doubt “whether any true virtue is to be found in the world.” (In *CPR*, A315/B372, he goes so far as to call it a “fact that no human being will ever act in a manner adequate to what is contained in the pure idea of virtue.”) Kant’s response to such doubts is not to deny them but to appeal to the perspectival hope provided by transcendental idealism: virtue remains possible as an *ideal*, because “reason by itself and independently of all appearances commands what ought to happen” (*GMM* 408). Later in the same work (*GMM* 450–1) he appeals more explicitly to *CPR*’s distinction between two “standpoints” as the only way of escaping the “circle” created by our own ignorance: “Even as to himself, the human being cannot claim to cognize what he is in himself through the cognizance he has by inner sensation. For, since he does not as it were create himself and does not get his concept a priori but empirically, it is natural that he can obtain information even about himself only through inner sense and so only through the appearance of his nature and the way in which his consciousness is affected”; yet a rational being “must necessarily assume something else lying at their basis, namely his ego as it may be constituted in itself”—this noumenal standpoint being the source of hope.

²⁸ GH and GG read Kant’s *allgemein* as an adjective (“common,” modifying *Grund*), whereas WP (correctly) translates it as an adverb, adding a footnote that suggests that Kant seems to be alluding to the universality of the moral law. However, it is possible that Kant is simply contrasting *allgemein* (which can also mean “generally”) with *besondern* (“particular”). I reserve “generally” for Kant’s use of *überhaupt* (see Glossary).

The process of inference Kant describes here cannot be empirically grounded but must be an entirely rational, a priori process. If a person is aware of performing one or more acts that he or she regards as evil and if such a judgment is correct, then we can infer that the person must have adopted, as a rational basis for choosing, a maxim that *made* the action evil.²⁹ This explains for the first time in *Religion* why *all* such evil maxims must themselves be synthetic: basing a choice on an underlying maxim (like basing a judgment on intuition in *CPR*) is not a merely logical procedure. The result is what Kant later (for example, at *R* 31.23) calls “the supreme maxim” (*die oberste Maxime*), and its transcendental status, though only implicit at this point, will emerge as the argument proceeds.

Kant begins the fourth paragraph by clarifying an ambiguity in the use of the term (human) “nature” whenever we consider whether human beings are good or evil “by nature.”

20.35–37, 21.01–09

However, lest anyone straightaway³⁰ take exception to³¹ the expression³² *nature*, which, if it were intended to signify (as /it\ usually /is\ the opposite of the basis of actions [arising] from freedom, would stand³³ in direct contradiction to the predicates *morally* {21} good or evil, it should be noted that here [I mean] by the nature of the human being only the subjective basis of the overall use of his freedom (under objective moral laws), the basis which (*dieser Grund*)—wherever it may lie—precedes any deed that strikes the senses. This subjective basis itself, however, must always in turn be an act (*Actus*) of freedom (for otherwise the use or abuse of the human being’s volition in regard to the moral law could not be imputed to him, and the good or evil in him could not be called moral).

Attempts to determine whether human beings are naturally good or evil are not about nature *as opposed* to freedom,³⁴ but about nature as an *expression* of freedom, or (what amounts to the same thing) about *freedom* as an expression of our *nature*. This point often escapes interpreters of *Religion*, yet it is crucial to a proper understanding of the judicial standpoint Kant is adopting.³⁵ The special *Actus* that empowers human persons with the ability to make moral

²⁹ In Kant’s moral theory, the act itself should not be called good or evil, but only the maxim that motivates it (see, e.g., *GMM* 397). However, as we shall see in §2.4, Kant adopts a rather different standpoint on this issue here in *Religion*, focusing on actions because intentions are beyond our reach.

³⁰ WP and GG have “immediately” for Kant’s *sofort*; GH has “at once.”

³¹ WP has “be ... troubled by” for Kant’s *sich am ... stoße*; GH has “difficulty ... be encountered in” and GG “be ... scandalized by.” The most literal translation would be “bump oneself (*sich*) against.” Note that Kant uses the active voice, not the passive.

³² Following GH and GG for Kant’s *Ausdrucke*; WP has “term.” Cf. *R* 190.13, 31.22.

³³ Following GG for Kant’s *stehen*; WP has “be,” but normally uses “stand.” GH omits this word, translating the verbal phrase as “flatly contradict.”

³⁴ That is, Kant’s question about our good or evil nature is not directly parallel to the distinction between the phenomenal and noumenal (the realms of nature and freedom) as presented in the three *Critiques*—even though so many interpreters have claimed the contrary (see, e.g., McCarty 2009: ch. 3.10).

³⁵ On “judicial” see note 0.15. *BRR* 272–3 notes this synthetic feature of Kant’s special definition of “nature” in *Religion*, though without linking it to the standpoint of *CJ*; instead, he emphasizes its significance for the problem, actively debated in Kant’s day, of how to establish responsibility for human actions. Interpreters who are troubled by Kant’s references to a timeless deed are far too numerous to list, but they include the anonymous book reviewer to whom Kant refers in the second Preface (*R* 13–14); see SP-2013c for details. Even Bohatec (*BRR* 306–9) admits that he is unable to solve the apparent contradiction that the deed that secures human responsibility is “acquired” (*R* 29.03), yet “timeless”; he blames the problem on Kant’s (to Bohatec, regrettable) focus on problems set by theologians such as Schultz and Heilmann. Schultz, for example, had solved the problem by grounding evil in nature (*BRR* 311)—a position Kant adamantly rejects. I have argued in SP-2010b: §4 that proper attention to the perspectival nature of Kant’s argument renders his alternative position self-consistent.

judgments³⁶ is a free nature that is “the subjective basis” that “precedes any deed that strikes the senses” (i.e., every *observable* act of human volition). Regardless of whether human nature is good or evil, therefore, it must be grounded in the noumenal (i.e., in freedom) yet expressed in the phenomenal (i.e., in the causal nexus), otherwise we could not be held responsible for the empirical results of our choices.

Kant has not yet staked a claim as to whether the mystery of free volition makes our “nature” good or evil; however, he believes that its necessary presence at the root of all moral action enables us to establish *how* the question must be answered—namely, by examining a person’s supreme maxim.

21.09–23

Hence the basis of evil cannot lie in any object *determining* volition through inclination, not in any natural impulse, but *can lie*, only in a rule that volition itself—for the use of its freedom—makes for itself, i.e., in a maxim. Now, concerning this /maxim\ it must not be possible to go on asking what is the subjective basis, in a human being, for the adoption of this maxim rather than of the opposite /one\. For, if ultimately this basis itself were not a maxim any more but a bare natural impulse, then the use of freedom could be *traced back*³⁷ entirely to determination by natural causes—which, however, contradicts freedom (*ihr*). Thus, when we say: The human being is by nature good or; He is by nature evil, this means no more than this: He contains (*enthält*) a first basis (inscrutable to us) for the adoption of good maxims or the adoption of evil (unlawful) /ones\, and /this\ moreover (*und zwar*) universally, as a human being, [and] hence in such a way that through this /adoption\³⁸ he expresses at the same time the character of his genus.

The specific rules or maxims we give ourselves in the process of deciding how to act must ultimately arise (whether explicitly or implicitly) out of our volition; and that capacity for choosing maxims must itself be guided by a higher level maxim, otherwise our choices would derive (like animal choice: see note 1.36) from our natural impulses, thus rendering freedom otiose and destroying our potential to be moral. The answer to the question must therefore depend neither on an empirical assessment of a given person’s actions nor on the character of our genus, but on whether a person’s supreme maxim is good or evil.

Before we proceed to the next paragraph to see how Kant fleshes out this claim, let us examine the footnote added to the observation that the ultimate basis of all our maxims is “inscrutable to us.” This parenthetical comment ironically threatens to destroy the whole project Kant is undertaking, if he seriously believes that we cannot obtain reliable knowledge of what this ultimate basis is. The footnote only heightens the mystery.

³⁶ Kant elsewhere distinguishes human from animal volition: unlike humans, animals choose in a deterministic framework that makes them amoral (see *LM* 256–7; for helpful discussions, see McCarty 2009, especially §3.10, and *BRR* 121–2, which traces Kant’s position to Baumgarten’s influence). Illustrating what *FDR* calls the “translation thesis,” Bohatec (*BRR* 277–86) surmises that Kant’s insistence on grounding human responsibility in a noumenal *act* must have been motivated by a desire to cater to the assumptions of church dogmas such as original sin, under the influence of Heilmann, Achenwall, and Baumgarten; assuming such a motivation, Bohatec laments that Kant’s theory is little more than “an objectionable relapse into theological intuitions” (*ein bedenklichen Ruckfall in theologische Anschauungen*) (286). He demonstrates similarly significant resonances between Stapfer’s theology and Kant’s position on freedom, moral responsibility, and the propensity to evil (*BRR* 322–40), even as expressed in *CPrR*, though without decrying the source.

³⁷ Following GG for Kant’s *zurückgeführt*; GH has “trace.” WP’s “reduced” has misleading connotations. If *reduction* occurred, the contradiction Kant goes on to mention would no longer occur, since freedom would be recognized as otiose.

³⁸ WP has “through this nature” for Kant’s *durch dieselbe* (literally “through/by the same”), while GG has “by his maxims”; GH has “thereby,” thus retaining the ambiguity of the original. Kant’s reference here, however, could not be to “nature,” because that term occurs before the colon, nor to “maxims,” because they express nothing until we *adopt* them.

21n.29–37

*That the first subjective basis for the adoption of moral maxims is inscrutable can provisionally already be seen from this: Since this adoption is free, its basis (why, e.g., I have adopted an evil maxim rather than a good /one\ must be sought not in any incentive of nature, but always in turn in a maxim; and since this /maxim\ must likewise have its basis, while yet apart from the maxim³⁹ no *determining basis* of free volition is to be or can be adduced, one is referred back (*zurück gewiesen*) ever further, *ad infinitum*,⁴⁰ into the series of subjective determining bases, without ever being able to arrive at the first basis.⁴¹

Given that free choice requires every maxim to be based on another maxim, we must presuppose an *inscrutable* ultimate basis for our maxims, otherwise we would have to assume an infinite regress of maxims, giving rise to a never-ending story of moral motivation. The story would be much simpler if we could trace our maxims to a natural impulse; but this would take our actions out of the realm of free volition, and therefore out of the moral realm.

With the problem of the inscrutability of our own ultimate maxim still intact, Kant proceeds in the fifth and final paragraph of this opening section to emphasize that, if we are to avoid blaming (or praising) nature for each person's moral character, this ultimate basis of human morality must arise out of a person's free volition.

21.24–28, 22.01–09

Hence of one of these characters (*Charaktere*) (distinguishing the human being from other possible rational beings) we shall say: it is *innate* in him; and yet in doing so we shall always concede that nature does not bear the blame for these /characters\⁴² (if he⁴³ is evil), or that it does not get the credit (*Verdienst*) (if he⁴⁴ is good), but that the human being himself is author⁴⁵ of that /innate character that distinguishes the human being\ (*desselben*).⁴⁶ But, since {22} the first basis for the adoption of our maxims, which must itself always lie in turn in free volition, cannot be a fact that could be given in experience, the good or the evil in the human being (as the subjective first basis for the adoption of this or that maxim in regard to the moral law) is called innate merely in this sense,⁴⁷ that it is laid at the basis (in earliest youth, up to [the point of] birth, back),⁴⁸ prior

³⁹ WP and GG have “a maxim” for Kant’s *der Maxime*; GH has “maxims.”

⁴⁰ GH and GG have “endlessly” for Kant’s *ins Unendliche immer weiter*; following WP, I consistently translate *ins Unendliche* as *ad infinitum*.

⁴¹ WP omits Kant’s concluding phrase *ohne auf den ersten Grund kommen zu können*. GH has “to reach” and GG “to come to” for *auf ... kommen*.

⁴² Following WP, who takes Kant’s *derselben* to be plural; GG has “for it,” following KV 23n, who assumes that Kant meant to write *desselben*. GH omits this problematic word. The referent of *derselben* could also be *der Unterscheidung* (“[the] distinguishing ...”)—an unlikely option, since it appears in parentheses. Kant would then be saying that nature is not responsible for what distinguishes human beings from other rational beings.

⁴³ WP and GG have “the character” for Kant’s *er*; GH has “it.” At this point in the text, however, Kant has never stated that *character* can be either good or evil; the last sentence of the previous paragraph explicitly refers to *the human being* as good or evil, not to character; so, even though the alternative translation of *er* as “it” (= the character) initially seems to be the best reading, it makes better sense for *ihm* (“him”)—i.e., *dem Menschen* (“the human being”)—to be the referent. Kant is still talking about *persons* being good or evil.

⁴⁴ WP has “the character” for Kant’s *er*, adding “for them” before the parentheses; GH and GG have “it.” Cf. note 1.43 immediately above.

⁴⁵ Following GG and GH for Kant’s *Urheber*; WP has “originator.”

⁴⁶ WP has “the originator of this character” for Kant’s *Urheber desselben*; GH and GG have “its author.” I read this as Kant’s first reference, after the semicolon, to the “one” character he has called “innate.” That is, Kant is saying that we are responsible for making *one* of the two “characters” innate.

⁴⁷ Following GH for Kant’s *in dem Sinne*; WP has “in the sense” and GG “*in the sense*.” Kant, rather oddly, emphasizes only the word that literally means “the.”

⁴⁸ WP has “back at the point of birth” for Kant’s *bis zur Geburt zurück*; GH and GG have “as far back as birth.” But the context and the most literal reading of this grammatical construction suggest that Kant is referring to the period *before birth* and *after* conception, not *soon after* birth.

to any use of freedom that is given in experience, and thus is presented⁴⁹ as present in the human being simultaneously with birth—[though] not exactly as having birth as its cause.

Since nature cannot be the source of this “first basis,” Kant argues, the only reasonable inference we can make is that one of the human being’s two possible moral “characters”—evil or good—is *innate*. The basis for being a good or an evil person cannot be derived from empirical facts; this basis cannot *result* from some experience we have had in the phenomenal world. Rather what makes a person good or bad must be presupposed to be “prior to” and must therefore “always lie ... in” (i.e., always be part of) every act of free volition, to the extent that it must already be present at our birth.⁵⁰ There is no evidence that birth *causes* either a good or an evil character; yet we can observe that, on the occasion of our very first free moral choice, this ultimate basis has already been put in place.

2. Comment: (A) Why moral neutrality is impossible

Having introduced an apparently irresolvable paradox before we even come to the first numbered subsection of his book, Kant explains in a supplementary “Comment” section why two easy ways of avoiding the “conflict” between the foregoing positions, pessimistic and optimistic, cannot succeed.

22.10–18

Comment

The conflict of the two hypotheses that were put forth above is based on a disjunctive proposition: *The human being is* (by nature) *either morally_s good or morally_s evil*.⁵¹ It readily occurs to anyone, however, to ask whether this disjunction is indeed correct, and whether someone could not assert /that\ the human being is by nature neither of the two, but someone else, /that\ he is both simultaneously, namely in several components.⁵² good, in others evil. Experience even seems to confirm this mean /position\ between the two extremes.

Kant admits here that, to the nonphilosopher—or to anyone who adopts what in *PSP* I called the “Perspective of Experience,” in contrast to “the Transcendental Perspective” that focuses always on the necessary conditions for the *possibility* of experience—it will seem obvious that human beings are neither good nor evil “by nature” but are a rather untidy, *piecemeal* (see note 1.52) mixture of the two. The challenge here, in other words, is that the question introduced above rests on a false dichotomy: being *all good* or *all evil* are not the only alternatives.

⁴⁹ WP and GH have “conceived” for Kant’s *vorgestellt*; GG has “represented.” Cf. note 0.47.

⁵⁰ See note 1.48 above. *BRR* 263–6 explores the influence of Stapfer, Heilmann, Schultz, and Crusius on Kant’s rationale for viewing evil as innate.

⁵¹ WP, GG and GH have “*morally good or morally evil*” for Kant’s *sittlich gut oder sittlich böse*. Here, as elsewhere, I add the subscript “_s” to indicate Kant’s use of *sittlich*; see Glossary.

⁵² WP has “points” for Kant’s *Stücken* (literally “pieces”); GH has “respects” and GG “parts.” But these options all fail to preserve the awkwardness of the German, which might have been intentional, given that Kant is using here the same word that he uses for the names of the four main divisions of *Religion*. (Recall that he wrote all four pieces, then sent them to the journal editor as a batch: see Introduction §1. So, even when this First Piece was published as an article, he would have had a motive for giving subtle hints regarding what might be coming up in *later* pieces.) If this usage was at least at the back of Kant’s mind here (and/or if passages such as this one—there are several such uses of *Stück* in the First Piece—played a role in his decision to use this rather odd word to designate the book’s four main divisions), then a subtext of this passage is to ask whether the “both good and evil” answer might pertain to one part of the book (e.g., the Second Piece), even though strict answers (either good or evil) pertain to the other parts (the First and Third Pieces). However, this possibility is remote, so I use “components”; further reasons are explained in the Glossary.

Kant responds forcefully to this (to his mind, unphilosophical) claim, arguing that moral matters should not rest on compromises.

22.19–28

However, the doctrine of *mores*⁵³ is greatly concerned overall to concede,⁵⁴ as long as this is possible, nothing morally intermediate,⁵⁵ neither in actions (*adiaphora*)⁵⁶ nor in human characters, because with such an ambiguity all maxims run the risk of losing their determinateness and firmness.⁵⁷ Those who are attached to this strict way of thinking are commonly called (by⁵⁸ a name that is intended *< soll >* to imply⁵⁹ a rebuke, but /that\ in fact constitutes *< ist >* praise); *rigorists*; and thus their antipodes may be called *latitudinarians*. Thus these /latter\ are latitudinarians either of neutrality and may be called *indifferentists*, or of coalition and may be called *syncretists*.

For Kant, *mores* (i.e., ethical/moral customs) are a product of reason; maxims should therefore be rational, and reason, being motivated in large part by the goal of seeking consistency in its comparisons, does not look kindly upon ambiguity. As if to illustrate his own, rigorously precise way of thinking, Kant diverts attention from the main topic just long enough to distinguish between his “rigorist,” black-and-white approach to moral teaching and the approach of those who prefer the “latitude” of focusing on the gray areas, either by refusing to take sides or by attempting to take both (or all) sides in a given dispute.⁶⁰

As often happens in Kant’s writings, the footnote he adds here is more informative than the sentence it supplements. Employing a mathematical method of reasoning that he had first put forward in 1763 (see note 1.69 below), Kant responds to the indifferentists.

22n.29–35

*If the good = a, then its contradictory opposite is the not-good. Now, this /not-good\ is either the consequence of a bare lack of a basis for the good, = 0, or the consequence of a positive basis for the reverse (*Widerspiels*) of the good (*desselben*), = -a; in the latter case the not-good can also be called the positive evil. (In regard to pleasure⁶¹ and pain, there is a similar mean,⁶² so that pleasure = a, pain = -a, and the state in which neither of the two is found, indifference, = 0.)

⁵³ WP has “doctrine of morals” for Kant’s *Sittenlehre*; GH and GG have “doctrine of ethics.” The term could also be translated “teaching of morals.”

⁵⁴ WP has “to admit ... no” for Kant’s *keine ... einzuräumen*; GH has “avoid admitting ... of anything” and GG “to preclude ... anything.”

⁵⁵ WP has the very awkward “no moral intermediate [some]things” for Kant’s *keine moralische Mitteldinge*. GH and GG have “anything morally intermediate,” giving the main verb a negative sense: “avoid” / “preclude.”

⁵⁶ Ancient Greek for “indifferent,” “undifferentiated.” The “indifferents” (*ta adiaphora*; singular *to adiaphoron*) constituted an important and much discussed classification of *things* in Stoic ethics. Kant, however, is applying the term to the moral character of *acts*. See also note 1.73 below.

⁵⁷ WP, GG, and GH have “stability” for Kant’s *Festigkeit*.

⁵⁸ WP, GG, and GH omit Kant’s *mit*, because they all move *Rigoristen* before the parentheses. I here preserve the dramatic effect of Kant’s word order, thus also highlighting the distinction between the two options.

⁵⁹ Following WP for Kant’s *in sich fassen* (literally “to grasp [or catch] within itself”); GH and GG have “to carry.”

⁶⁰ As WP 23n points out, BRR 176n (cf. 28n) traces the source of the technical terms Kant uses here to the Swiss theologian J. F. Stapfer. For a now classic discussion and defense of Kant’s “rigorist” position and its implications for Kant’s ethics in general, see Allison 1990: ch. 8. At MM 409 Kant clarifies that his rigorist position does not imply that all maxims have such a moral status. See also Wood 1999: 26–30.

⁶¹ Following GH and GG for Kant’s *Vergnügens*; WP’s “gratification” can have an inappropriately derogatory connotation in English. This German term refers to “pleasure” as enjoyment or fun, in contrast to Kant’s usual word, *Lust*, which typically refers to the satisfaction of an appetite.

⁶² WP has “something intermediate of that sort” for Kant’s *ein dergleichen Mittleres*; GH and GG have “a similar middle term.” Both options are too vague, as Kant is referring to something like a mathematical (e.g., statistical) “mean” between two “extremes,” not to just any term in between the other two.

Kant's special method treats qualities as if they could be mapped onto a number line, opposite qualities being represented by positive and negative terms, while the neutral point positioned between these two is represented by 0 (zero). He proceeds to apply this method of analysis to the issue being considered in the main text.

22n.35, 23n.04–12

Now, if in us the moral law were {23} not an incentive of volition, then moral-good (consonance of volition with the law) would be⁶³ = a, not-good would be = 0, but this [situation],⁶⁴ the bare⁶⁵ consequence of the lack of a moral incentive, would be = a \times 0. In fact (Nun), however, the moral law (es) is an incentive in us, = a; consequently the lack of conformity⁶⁶ of volition with this /law\ (=0) is possible only as the consequence of a real⁶⁷ opposite determination of volition, i.e., as the consequence of a resistance by it = -a, i.e., only through an evil volition; and hence between an evil and a good conviction (inward⁶⁸ principle of maxims)—by which the morality of the action must also be judged—there is no mean.

Kant presents here a subtle but powerful two-step argument. First, after defining what is “moral-good” as “consonance of volition with the law,” he argues that such consonance can come about only if the moral law relates to us as a positive “incentive,” because, without a genuinely positive basis for choosing good, the distinction between good and its opposite would collapse: both good and evil would amount to something essentially neutral (cf. note 1.64). Second, because moral goodness can and must be traced to something positive, its opposite must also be regarded not as a mere absence of conformity with the moral law, but as an active opposition to it. Therefore, when it comes to the choice that every person must make in order to be convinced of a fundamental principle designed to guide all of our maxims, the only viable options are to choose positive good or to choose positive evil; there is no room for evil as a mere absence (0) or privation of the good, as so many theologians had argued.⁶⁹

With its appeal to his previously defended mathematical methodology, this is Kant's strongest argument so far; yet he tucks it away in a footnote. Surprisingly, he also chooses to introduce here, for the first time after the Prefaces, what is arguably his most important

⁶³ Following GG for Kant's *Wäre ... sein*; WP and GH have “would.”

⁶⁴ GG and WP also insert an additional “would be” at this point. They and GH all take Kant's *dieses* to mean “the/this latter.” However, this renders the following phrase unintelligible. Kant's point here is that, if the moral law is not a genuinely positive incentive, then the distinction between good (a) and not good (taken as neutral, as 0) collapses: both options would turn out to be 0.

⁶⁵ WP has “mere” for Kant's *bloße*; GH has “merely” and GG “just.” On my use of “bare” here and throughout, see note 0.5, Introduction §4, and PCR VI.1. Although Kant is not using the term here in the same sense as in the title, as a reference to the inner circle of rational religion, the type of use is the same: morality would collapse if reason did not look beyond the bare form of the moral law to an incentive within us—namely respect.

⁶⁶ WP, GG, and GH have “agreement” for Kant's *Übereinstimmung*.

⁶⁷ WP's “really” is the literal translation of Kant's Latin word *realiter*; GH and GG have “real and,” though Kant does not have *und*. Kant is here employing the crucial distinction between real and logical opposition, first introduced in his 1763 essay on negative magnitudes (see note 1.69 below).

⁶⁸ WP and GH have “inner” for Kant's *innerem*; GG has “the inner.”

⁶⁹ Matušík 2008: 90 mistakenly takes Kant himself to regard evil as privative. On the relation between Kant's position and the theory of evil as a privation of good (often attributed to St. Augustine), see PCR 288–9. In LPT 1078 Kant himself appears to defend a privation theory similar to the one adopted by Leibniz; however, this could have been merely his summary of Baumgarten's textbook, which was the basis for that part of his lectures, because in his earlier (1763) essay on negative magnitudes (ICNM 182–4) he defends a position much closer to the one presented here, using the same logical apparatus to distinguish the various options. See Loudon 2000: 134–6 for further evidence that the theory of evil in *Religion* does not constitute a radical break from Kant's past, as many have claimed. AAE-Card 75f states six “controversial theses” that emerge from Kant's account of evil in the First Piece; she disagrees with Kant in defense of the “common sense” view. What Card's (and many others') criticisms neglect is that Kant is giving an account not of empirical evil, but of evil as a necessary condition for the universal human experience of religiosity.

technical term, *Gesinnung*, defining it as an “inward principle of maxims.” I translate this term as “conviction,” for reasons defended fully elsewhere.⁷⁰ The overall conclusion of the argument in this footnote is that this deep, inward conviction must be either good or evil; it cannot be neutral, because that would render morality ultimately meaningless.⁷¹

In R2 Kant adds a new sentence to the end of this note (without marking it with a dagger),⁷² reiterating a claim he has already made, though less pointedly, in the main text.

23n.13–17

[[†]]A morally indifferent action (*adiaphoron morale*)⁷³ would be an action ensuing⁷⁴ merely from natural laws, which therefore stands in no reference whatever to the moral law, as law of freedom—because it is not a *factum*,⁷⁵ and in regard to it neither *command* nor *prohibition* nor yet *permission* (legal authority) has a place, or is needed.

Kant does not deny that human beings sometimes make morally neutral choices. Indeed *all* choices motivated by concerns determined by natural laws (i.e., concerns related to our physical nature, our inclinations, etc.) are in a sense irrelevant to morality, because a choice can be relevant to the morality of our *conviction* (our inward, subjective basis for making choices) only if it is *free*, and choices determined by laws external to our own reason are not free.⁷⁶

⁷⁰ GH and GG use “disposition,” the term now employed almost universally in scholarly discussions of Kant’s theory of *Gesinnung*, while WP uses “attitude” (defending his translation in detail at WP 13–14n). However, neither a disposition nor an attitude is normally regarded as a “principle.” A conviction, by contrast, normally does rest on a maxim or principle that one adopts because one has become convinced of its truth. For an initial defense of this new translation of *Gesinnung*, see SP-2009a: xxviii; I offer a thorough defense in SP-2015d, where I respond to the main objection to using this translation: Kant’s technical term, *Überzeugung*, is normally translated as “conviction.” The latter convention has persisted in Kant scholarship even though *all* other forms of the word *überzeug-* are translated as forms of “convince,” not as forms of “convict.” I solve both problems by translating *Überzeugung* as “convincement” (see note 3.83), reserving “conviction” for *Gesinnung* and using “convicting” (in the legal sense) for *überführung*. This new translation has the significant merit of portraying Kant’s theory of the (so-called) “disposition” as referring not merely to a noumenal component of human nature, as past interpreters have assumed, but to a form of human volition with a *phenomenal* grounding (see Glossary).

⁷¹ Recall that Kant had argued in *CPrR* that the moral law motivates us to make choices that are universalizable, to respect the humanity in all persons, and to foster an environment wherein all human beings can make similar choices for themselves—i.e., to conform our choices to these three formulations of the categorical imperative. Cf. note 0.37.

⁷² WP and GG present this added sentence as if it had been, like the next added footnote, a new footnote attached to the early part of the next sentence in the main text, rather than merely a sentence added to the end of the previous R1 note. What probably prompted this mistake is that *Ak.* formats the added sentence as a *second paragraph* marked with a dagger, even though in Kant’s R2 text it is merely an unmarked additional sentence. Only GH locates the note correctly, as a mere addition to the previous (R1) note.

⁷³ Cf. note 1.56 above; lest there be any doubt, Kant here adds the modifier *morale*, a Latin adjective (here in the neuter, to agree with the Greek *adiaphoron*) meaning “moral”: “what is morally undifferentiated.”

⁷⁴ WP and GH have “resulting” for Kant’s *erfolgende*; GG has “follows.”

⁷⁵ Latin for “deed,” in the sense of something that has actually been done and is considered an empirical event.

⁷⁶ A frequently discussed paradox plagues Kant’s theory at this point: if actions that are determined by our physical nature—and hence presumably motivated by sense-based inclinations, what Kant elsewhere calls actions—are determined by nature and therefore “morally indifferent,” then how can autonomous (freely chosen, *Wille*-motivated) actions ever be anything but morally good? And, in that case, how can we be held responsible for (so-called) evil actions, since we are not free to choose which inclinations determine our motives? *BRR* 258–62 claims that Kant inherited this paradox from Wolff; this could be why the problem was detected as early as 1786, in a book review of *GMM* (in *Tübingische gelehrte Anzeigen*, February 16, pp. 105–12) by J. F. Flatt, a devoted follower and translator of Storr. For an excellent recent discussion of this problem and its relation to the debate between realist and constructivist interpretations of Kant’s moral theory, see Giannini 2013. The best solution is to take seriously Kant’s claim that moral evil lies in rational choice rather than in inclinations. Evil, as we shall see below, is for Kant the *free choice* to allow oneself to be controlled by one’s natural inclinations, and this freedom to choose whether or not to be so controlled is an inalienable part of what Kant will call our predisposition to humanity (see my comments at R 26.01–11). That is, to be human is to be *forced* to choose, and thereby to be responsible for what we do *even if* we end up allowing ourselves to be determined by our natural inclinations.

Having established (in the footnote) the irrelevance of morally neutral choices to the question at hand, the main text now examines the first of the two previously mentioned options, the one Kant seems to prefer.

23.01–03, 24.01–05

{23} The answer to the suggested⁷⁷ question according to the rigoristic way of deciding[†] is based on this observation, [which is] important for ethics: the freedom of volition has the quite peculiar make-up⁷⁸ {24} that it cannot be determined to an action by any incentive *except insofar as the human being has taken up⁷⁹ the incentive* (sic) into his maxim (has made this (es) a universal rule for himself, according to which he wills to behave⁸⁰ himself). Only in this way can an incentive, whichever it may be, coexist⁸¹ with the absolute spontaneity of volition (i.e., with freedom).

This clarifies that an incentive *can* play a role in morality, but only if the moral agent specifically makes it part of the maxim governing the choice. If the chosen incentive is respect for the moral law, its coexistence with freedom is not problematic. But Kant appears to be saying that, if the incentive is a naturally determined motive that encourages a person to follow his or her inclinations, then, quite paradoxically, a person can *freely* choose a maxim that will in effect put an *end* to the freedom-based primacy that practical reason should have for the will.⁸²

Shortly after resuming the flow of his argument, Kant digresses again with another meaty footnote (added in R2) that in several respects is more informative than the main text. Responding to Schiller's recent criticism of his moral philosophy, Kant begins by portraying their positions as deeply complementary.

23n.18–23

[†]Professor (Herr Prof.) Schiller, in his masterfully composed essay on *gracefulness* (Anmuth) and *dignity* in ethics (Thalia, 1793, 3rd issue),⁸³ disapproves of this way of presenting⁸⁴ obligation,⁸⁵ as if it carried with it a Carthusian-like mental attunement (Gemüthsstimmung). However, since we are united on the most important principles, I also cannot deploy⁸⁶ any disunity on this one, if only we can make ourselves understandable to each other. —

⁷⁷ WP has “mentioned” for Kant’s *gedachten*; GH has “at issue” and GG “just posed.” Kant uses this term because he did not actually pose or state the question previously, but only suggested it indirectly.

⁷⁸ WP and GG have “characteristic” for Kant’s *Beschaffenheit*; GH has “nature.”

⁷⁹ WP has “admitted” for Kant’s *aufgenommen*; GH and GG have “incorporated.” This term is the basis for Allison’s influential claim that Kant is here proposing a so-called “Incorporation Thesis” (Allison 1990: 5–6, 40); but Kant gives no indication that he intended this to be regarded as a technical term.

⁸⁰ WP, GG, and GH have “conduct” for Kant’s *verhalten*.

⁸¹ Following GH and GG for Kant’s *kann ... zusammen bestehen*; WP has “is ... consistent,” acknowledging the literal meaning in a footnote.

⁸² Kant explicitly defends the primacy of practical reason in *CPrR* 119–21. Whether or not Kant really means that we can choose to be unfree is a matter of considerable debate among Kant scholars (see note 1.76 above). Whatever his “official” position may be, he appears to affirm this one here. What is less debatable is that, for Kant, we become *free* (i.e., autonomous) only by choosing to be “determined” by the moral law (as imposed on us by *Wille*). Paradoxically, then, genuine freedom entails choosing to limit one’s actions to what duty demands, while false freedom (i.e., evil) entails “freeing” oneself to do whatever one’s inclinations determine.

⁸³ This journal was edited by Schiller himself (see WP 24n). After appealing to Greek mythology to explain the relationship between gracefulness, beauty, and other themes, the essay Kant refers to here explicitly critiques Kant (Schiller 2005: 150), whose “idea of duty is presented with a severity that repels all graces and might tempt a weak intellect to seek moral perfection by taking the path of a somber and monkish asceticism.”

⁸⁴ WP has “way of conceiving” for Kant’s *Vorstellungsart*; GH has “manner of presentation” and GG “mode of representation.” WP usually has “presentation” for Kant’s *Vorstellung*.

⁸⁵ Following WP, GG, and GH for Kant’s *Verbindlichkeit*, which also means “liability”; see Glossary and note 7.98.

⁸⁶ WP has “establish” for Kant’s *statuiren*; GH and GG have “admit.” Kant’s term can mean to prescribe or decree, to suppose or allow, or to set up as an example, as when an artist *deploys* a model.

The Carthusian order (founded by St. Bruno in 1084) is a monastic community of hermits, still part of the Catholic church today, who spend most of their time in solitary, ascetic contemplation, in the hope of allowing what is good in human nature to regain its original prominence. Kant's "as if" indicates that he *rejects* Schiller's claim that his moral theory's focus on the absolute priority of *duty* over human inclinations requires such harsh asceticism.

Kant proceeds to explain Schiller's charge against him and why this charge raises no genuine objection for his moral philosophy. Although he is here discussing features of his own moral philosophy, Kant does so in a way that illustrates how his position leads naturally to the (quite distinct) position on religion that he will defend in *Religion*.

23n.23–30

I gladly admit that to the⁸⁷ *concept of duty*, precisely on account of this concept's (*seiner*) dignity, I cannot adjoin⁸⁷ *gracefulness*. For it contains unconditional constraint,⁸⁸ to which *gracefulness stands* in direct contradiction. The majesty of the law (like the /law\ on Sinai) instills awe (not dread, which repels; also not charm, which invites familiarity); /and\ this /awe\ (*welche*) arouses the *respect* of the subordinate toward his master; but in this case, where the master (*dieser*) resides in ourselves, it arouses a feeling of the sublimity of our own predetermination,⁸⁹ [and] the sublime (*was*) enraptures us more than anything beautiful. —

Schiller had criticized Kant's concept of duty for *excluding* gracefulness, arguing that graceful movement is an empirical reflection of a person's inner moral maxim. He claimed that gracefulness unites morality, as "a principle which resides beyond the world of sense" with beauty, as a principle that is "purely sensuous" (Schiller 1902: 200). Kant's response shows that he is not ignoring gracefulness but putting it in its proper place in relation to duty. Using language that strikingly foreshadows Rudolf Otto's account of the experience of the numinous,⁹⁰ Kant correlates the distinction between duty and gracefulness to his *CJ* distinction between the feelings of sublimity (as unconditional and awesome) and beauty (as conditional and charming). Whereas the giver of duty is a master who "resides in ourselves," the giver of gracefulness is (presumably) "outside the human being" (cf. *R* 6). Schiller had identified the latter with the charms of "nature," while Kant now compares the former to the awesome God who gave Moses the Ten Commandments on Mount Sinai. Thus Kant's theory of duty does not ignore gracefulness but relates to it as sublimity relates to beauty.⁹¹ This comparison suggests that, unlike the concerns examined in *CPrR*, where the goal was to understand how duty is generated from a conception of lawfulness (i.e., from the concept of *Wille*) that is binding on all human beings, here in *Religion* the contingencies related to our embodiment *are* relevant—apparently to both duty and gracefulness, since his purpose here is to persuade Schiller that their positions are *compatible*.

⁸⁷ WP has "with the ... associate" for Kant's *dem ... beigesellen*; GH has "associate ... with" and GG "associate ... with the." I reserve "associate" for *sich verbinden*.

⁸⁸ WP and GG have "necessitation" for Kant's *Nöthigung*; GH has "necessity." On the harsh connotations of *Nöthigung*, see note 12.38.

⁸⁹ WP and GG have "vocation" for Kant's *Bestimmung*; GH has "destiny." See Glossary for a defense of my translation of this term in such contexts.

⁹⁰ With frequent references to the law of Sinai, Otto 1923 adopts as technical terms some of the words Kant uses here (i.e., *Majestät* "majesty," *Ehrfurcht* "awe," and *Reiz* "charm"). For a brief analysis of Otto's position and an explanation of how it can be seen as a complement to Kant's philosophy of religion, see SP-2000b: §31, and Firestone 2009: 118–38. On the close relation between awe and respect, see *MM* 488–9, where Kant links both to God's *justice*.

⁹¹ In *CPR* Kant argued: "where you do perceive purposive unity, it must not matter at all whether you say, 'God has wisely willed it so' or 'Nature has wisely so ordered it'" (A699/B727). Here in *Religion*, in the wake of the "physicotheology" defended in *CJ*'s Appendix (§85), he seems to be suggesting that duty and gracefulness correspond to God/sublimity and Nature/beauty. For a detailed discussion of Kant's reply to Schiller, with special attention to *R* 23n, see Winegar 2013.

Kant's reason for inserting this footnote on gracefulness at this point is to prepare the reader for what he goes on to say later in the main sentence, after the footnote, about the proper role of sense-based incentives in moral choice. When it comes to the actual application of duty to our lives, to what we today might call the existential task of living a virtuous life, gracefulness does play a legitimate role, for virtue welcomes the company of “the graces.”

23n.30–35

But *virtue*, i.e., the firmly based conviction to fulfill one's duty strictly,⁹² is in its consequences also *beneficent*, more so than anything that nature or art may accomplish in the world; and humanity's glorious⁹³ image, drawn up in this figure of /virtue with beneficent consequences/,⁹⁴ quite readily permits being accompanied by the *graces* (*Grazien*), who, however, as long as duty alone is still at issue, keep at a reverential distance.

Kant here defines “virtue” as the firm conviction that one must strictly follow one's duty (i.e., obey the moral law). He responds to Schiller by explaining that, even though abstract moral philosophy must ensure that the graces remain “at a reverential distance” from duty, the *consequence* of a virtuous life will be to welcome just the sort of “splendid image” of a “beneficent” humanity that Schiller extols. If this can be done (e.g., through religious symbolism, as Kant will attempt here in *Religion*), then such consequences will serve to synthesize the realms of theory and practice even “more so than anything that nature or art may accomplish”; that is, religion can form the bridge uniting the Critical System even better than *CJ* did.

Continuing this footnote with an appeal to mythology, Kant explains why sensibility becomes an inevitable factor once we consider virtue (i.e., the real human *experience* of goodness) as opposed to duty (i.e., an abstract rational *conception* of goodness).

23n.35–42

But if one takes account of the graceful consequences that virtue, if it gained access (*Eingang fände*) everywhere, would disseminate⁹⁵ in the world, then the morally directed⁹⁶ reason also calls (through the imagination) sensibility into play. Only after subduing monsters does Hercules become *Musagetes*⁹⁷ a labor from which those good sisters recoil. These companions of Venus Urania are wanton sisters in the retinue (*Gefolge*) of Venus Dione as soon as they meddle in the business of determining duty and want to supply the incentives for this. —

Far from seeking to eclipse the role of sensibility in human life, Kant's moral theory puts incentives arising from our embodiment in their proper place: as a natural *outcome* of a virtuous life rather than as its motive force. Our “imagination” can then quite properly picture, as in Greek mythology, even the gods undergoing real struggles in the empirical world before they can reap the fruits of virtue. Hercules' successful struggles with monsters put the Muses under his control; yet these companions of Venus (goddess of love and beauty) lose their heavenly

⁹² Following GG for Kant's *genau zu erfüllen*; WP has “toward strictly fulfilling” and GH “strictly to fulfil.”

⁹³ Following GG for Kant's *herrliche*; WP has “splendid” and GH “august.”

⁹⁴ WP has “in the guise of virtue” for Kant's *in dieser ihrer Gestalt*; GH has “in this character” and GG “in the figure of virtue.” But *dieser ihrer* must refer not to mere virtue, but to virtue as *beneficent*, for Kant goes on to *contrast* it with “duty alone” (i.e., without beneficence). Only as beneficent does virtue appear (as Kant argued in *CJ*) as humanity's “figure,” “drawn up” by nature and art.

⁹⁵ WP, GG, and GH have “spread” for Kant's *verbreiten*, but I reserve “spread” for *ausbreiten* (when used without *sich*). The noun *Ausbreitung* is translated with a form of “proliferation.”

⁹⁶ Following GH for Kant's “-gerichtet”; WP and GG have “oriented,” but WP elsewhere always uses “directed.” See Glossary.

⁹⁷ That is, Apollo, leader of the Muses, the “good sisters” to whom Kant refers next.

("Urania") character and become "wanton" when they try to be mothers ("Dione") of virtue by serving as "incentives" for duty. Kant's picture of an ideal world where all persons are both virtuous and happy is therefore one that, far from being ascetic (as Schiller had claimed), *requires* sensibility—and with it, as we shall see, notions arising out of the religious imagination, such as grace. For, as we shall see in §1.4, Kant's theory of the threefold predisposition to good suggests that *humanity* necessarily entails not only striving to be moral (*personality*), but also struggling to be fully embodied (*animality*).

After a dash, this lengthy footnote concludes by examining a specific aspect of virtue's sensible nature: the role of human "temperament" in moral deliberations.

23n.42, 24n.23–33

Now, if one asks: Of what kind is the *aesthetic make-up*, the {24} *temperament*, as it were, of *virtue*, [is it] courageous [and] hence *cheerful*,⁹⁸ or *anxiety-bent*⁹⁹ and *depressed*?¹⁰⁰ then an answer is hardly needed. The latter, slavish mental attunement can never occur without a hidden *hatred* of the law, and the cheerful heart in *complying* with one's ⟨*seiner*⟩ duty (not the comfortableness in *recognizing* it)¹⁰¹ is a sign of the genuineness of a virtuous *conviction*. It is such a sign, even in *piety*, which consists not in the self-torment of a repentant sinner (which is very ambiguous and is usually only an *inward* reproach for having violated the rule of prudence), but in the firm resolve to do better in the future—, a resolve, that, *fired up*,¹⁰² by good progress, must bring about a cheerful mental attunement, without which one is never [quite] certain of also having *embosomed*¹⁰³ the good, i.e., of having *taken* it *up* into one's maxim.

These three sentences, rich in nuanced implications, deserve deep meditation. Kant here argues for a specific empirical outcome as a definitive sign of genuine virtue. The person with a virtuous conviction, as one who has nurtured a *love* for what is good, will be "cheerful"¹⁰⁴ when obeying the moral law, whereas a person with a "slavish"¹⁰⁵ temperament exhibits a "mental attunement" that lacks a genuine conviction to do good, actually *hating* the moral law and obeying the call of duty only regretfully, out of prudence.¹⁰⁶

Bearing in mind the crucial insight conveyed by Kant's footnote (i.e., that affirming Kantian morality frees us to embrace our animality rather than forcing us to deny it), we now resume

⁹⁸ WP omits Kant's emphasis on *fröhlich*.

⁹⁹ WP and GG have "weighed down by fear" for Kant's *ängstlich-gebeugt*; GH has "fear-ridden." Kant is using the term Kierkegaard later employed as the focal concept of his groundbreaking 1844 book *The Concept of Anxiety*. Cf. R 146n.33.

¹⁰⁰ WP, GG, and GH have "dejected" for Kant's *niedergeschlagen*.

¹⁰¹ WP has "acknowledging the law" for Kant's *Anerkennung desselben*, because he inserts "[according to the law]" just before this parenthetical phrase; GH has "the recognition thereof" and GG "the recognition of it."

¹⁰² WP has "energized" for Kant's *angefeuert*; GH and GG have "encouraged."

¹⁰³ WP has "grown fond of" for Kant's *lieb gewonnen*; GH has "attained a love for" and GG "gained a love for." Kant's metaphor here alludes back to Venus (goddess of love), mentioned just two sentences earlier.

¹⁰⁴ GH and GG have the more religious term, "joyous," for Kant's *fröhliche*. Schiller had accused Kantian morality of lacking the "joy" that he believed is present in the most highly developed forms of moral goodness (Schiller 2005: 149–50; cf. note 1.83 above).

¹⁰⁵ Schiller had accused "the immortal author of the 'Critique'" of trying to make inclination a slave to reason (Schiller 2005: 148, 151). Kant is here anticipating an even more influential criticism, later leveled against his moral theory by Schiller together with Goethe, who wrote in *Xenien*, their 1797 collection of critical poems: "Scruples of Conscience / I like to serve my friends, but unfortunately I do it by inclination / And so often I am bothered by the thought that I am not virtuous / Decision / There is no other way but this! You must seek to despise them / And do with *repugnance* what duty bids you" (Wood 1999: 28).

¹⁰⁶ For example, Kant thinks that those who decide not to steal because they realize that the chances of getting caught are too high are not morally praiseworthy, for their actions only accidentally comply with the moral law: they realized that it would be more prudent to obey. Nevertheless, Kant's theory of prudence is not entirely negative, as some have claimed; see SP-2014a and SP-2015e.

our examination of the main text at R 24, midway through the second paragraph of the “comment” added to the untitled introductory section of the First Piece.

24.05–15

However, in the judgment of reason the moral law is on its own an incentive, and whoever makes it his maxim is *morally* good. Now, if, in regard to an action referring to the law (*dasselbe*), the law nonetheless does not determine somebody's¹⁰⁷ volition, then an incentive opposed to the law (*ihm*) must have influence on his volition; and since, by dint of the [above] presupposition, this can occur only by the human being's taking up¹⁰⁸ this /incentive\ ([and] hence also the deviation from the moral law) into his maxim (in which case he is an evil human being), his conviction in regard to the moral law is never indifferent (never neither of the two, neither good nor evil).

For those readers who have attended carefully to Kant's footnotes, these two sentences convey nothing new but are more like a summary of what has been argued (mainly in those footnotes) up to now. The argument, in outline, is:

1. The moral law actively motivates every rational being to choose what is good.
2. Neutral determining influences (e.g., from nature or from our natural inclinations) cannot, on their own, counteract this positive incentive.
3. When a person performs an action that is contrary to the moral law, this indicates that the person must have actively allowed an *evil maxim* to determine the nature of his or her basic moral conviction.¹⁰⁹
4. Therefore, as far as his or her *conviction* is concerned, a person can never be morally neutral but will always exhibit *either* original goodness (as in premise 1) *or* evidence of having chosen evil (as in 2).

Kant assumes the validity of this conclusion throughout the remainder of the book, but not before explaining why one further alternative must also be mistaken.

3. Comment: (B) Could humans be partly good and partly evil?

Aspiring to completeness in his exposition, as always, Kant now briefly addresses the remaining option, that a person's basic moral status might occupy a middle ground.

24.16–22

But he also cannot be in several components, morally good, [and] simultaneously [[morally]] evil in others. For if he is good in one /component\, then he has taken up the moral law into his maxim; thus if in another component he were to be simultaneously evil, then, because the moral law of compliance with duty overall is only a single /law¹¹⁰ and is universal, the maxim connected to it would be universal, while simultaneously being only a particular maxim—which is contradictory.*

¹⁰⁷ Following GG, with WP's word order, for Kant's *jemandes ... einer*. WP has “someone's (action) ... the (power)” and GH “a person's (will*) ... an (action).”

¹⁰⁸ WP has “admitting” for Kant's *aufnimmt*; GH has “adopts” and GG “incorporate.” As explained in the Glossary, I translate forms of *aufnehmen* literally, to distinguish them from forms of *annehmen*.

¹⁰⁹ The occurrence of “conviction” (*Gesinnung*) in the passage from R 24.05–15, quoted above, marks Kant's first use of this technical term in the main text of *Religion*. See note 1.70 (above) for more about why I translate this term as “conviction.”

¹¹⁰ WP has “only one” for Kant's *nur ein einziges*; GH has “essentially single” and GG “a single one.”

One of Kant's fundamental assumptions is that the moral law, being a product of reason, cannot contradict itself. A good act is one based on a maxim that is universalizable under the moral law, and a good (or virtuous) person is one for whom such a moral maxim serves as a principle informing his or her *conviction*, as Kant has now officially named our inward moral nature; an evil act is one based on a maxim that is not universalizable and is thus contrary to the moral law, and an evil (or vicious) person is one whose conviction follows suit. The former adopts a universal law that governs all persons equally; the latter adopts a particular law that sets oneself apart as more deserving than others.¹¹¹ Just as a maxim cannot be both universal and particular, so the rational (and therefore timeless) basis of one's conviction must be *either good or evil* at any given time.

Kant attaches a footnote to this paragraph, again addressing an issue from ancient philosophy that seems at first to be peripheral but turns out to offer crucial insight into his intentions.

24n.34–38, 25n.26–34

*The ancient moral philosophers, who exhausted just about everything that can be said about virtue, also did not leave the above two questions untouched. The first they expressed thus: Whether virtue must be learned (Is¹¹² the human being therefore by nature indifferent toward it and vice)? The second was [this]: Whether there is more than one virtue (Does it hence not perhaps take place¹¹³ that the human being {25} is in several components virtuous and in others vicious)? Both [questions]¹¹⁴ they answered in the negative (*wurde von ihnen ... verneint*) with rigoristic determinateness, and rightly so; for they were considering virtue *in itself*, in the idea of reason (as the human being ought to be). But if one wants to make a moral judgment about this moral (*moralische*) being, the human being *in appearance*, i.e., as experience allows us to be aware of him, then both of the cited questions can be answered affirmatively; for then he is judged not on the scales of pure reason (before a divine tribunal), but according to an empirical standard (by a human judge). This (*Wovon*) will be dealt with further in what follows.¹¹⁵

Neglect of the important perspectival concession Kant makes here has resulted in many past misunderstandings of *Religion*. As a philosopher grounded in the a priori, Kant sees himself as a rigorist and takes comfort in the resonance he detects between his position and that of ancient philosophers on the twin issues of whether the moral nature of human beings could be neutral or whether it could be mixed. Nevertheless, he admits (somewhat shockingly, to interpreters who assume that Kant's point of view is exclusively anthropocentric)¹¹⁶ that this

¹¹¹ The literature assessing the legitimacy of this assumption is massive. Some have argued that immoral maxims are just as universalizable as moral ones, while others have sought to defend Kant's position. This wide-ranging debate, however, belongs to Kant's practical philosophy, not to his philosophy of religion (especially if we view the latter as part of the *judicial* wing of his system; see notes 0.15 and 5.114), so we can safely ignore it here.

¹¹² I have rearranged WP's word order to phrase this parenthetical insertion as a question, following GH. WP adds "and" at the beginning of both this and the next parenthetical phrase, in order to reduce the awkwardness of presenting each as a statement.

¹¹³ WP has "and hence [the alternative] does not perhaps have a place" for Kant's *mithin es nicht etwa statt finde*; GH and GG omit most of this difficult phrase, having only "so" and "and hence ... perhaps," respectively. GG adds in a footnote that Kant's *nicht*, in particular, "does not seem to make any difference." In this case, none of the three translators makes the parenthetical phrase into a question. However, Kant provides question marks for both parenthetical phrases, so he seems to have intended each one to clarify the *question* at hand. WP's addition of "[the alternative]" to translate Kant's *es* ("it") results in an incoherent reading, whereby Kant ends up saying the opposite of what he must have meant to say, in order for the two questions to correspond to the two options he mentions above.

¹¹⁴ Following GH. WP adds "[issues]" while GG leaves Kant's *Beides* ("Both") to stand on its own. Kant is clearly referring to the two *questions* he just posed.

¹¹⁵ Following GH for Kant's *in der Folge noch*; WP has "also ... later" and GG "More ... in what follows."

¹¹⁶ In PCR I.1–3 I have argued that Kant's philosophy, properly understood, is both anthropocentric (as far as its focus on *knowledge* is concerned) and theocentric (as far as its equally important focus on our *necessary ignorance* is concerned). The statement Kant makes near the end of the footnote quoted above is one of many textual justifications for the accuracy of this claim.

(rigorist) position—the position he himself defends in *CPR*—adopts the perspective of a “divine” judge of the human heart. If, by contrast, we adopt the perspective of a merely human judge of human actions, we must employ some different, “empirical standard.” This hint of what is to come in *Religion* is intriguing, to say the least. It reveals the gross inadequacy of any approach to interpreting *Religion* that sees it as a mere appendix to Kant’s ethics. In fact Kant is offering a fresh (judicial) account of how we human beings are to understand *and judge* their own moral nature, given the limitations implicit in our status as embodied beings.

After all these preliminaries, Kant is ready to draw to a conclusion the reflections that prepare the way for the first numbered section of the First Piece. He begins the final paragraph of the Comment by making an important clarification about our responsibility for our moral nature—a clarification that includes, somewhat belatedly, a formal definition of “conviction” (*Gesinnung*).

25.01–09

{25} To have the one or the other conviction as an innate make-up by nature also does not mean here that it has in no way been acquired,¹¹⁷ by the human being who harbors it, i.e., that he is not the author; rather it means only that it has not been acquired over time (*in der Zeit*) (that he *has always been, from his youth onward*,¹¹⁸ one or the other). Conviction, i.e., the first subjective basis for the adoption of maxims, can only be a single /basis,¹¹⁹ and it applies universally to the entire use of freedom. But conviction (*Sie*) itself must also have been adopted through free volition, for otherwise it could not be imputed.

Officially, then, “conviction” refers to “the first subjective basis for the adoption of maxims”; that is, it is the deepest layer of a moral agent’s decision-making process, determining what *kind* of maxims we will adopt when making free choices. The paradox Kant here acknowledges by way of clarification is that, even though the state of a person’s conviction (i.e., whether it is “by nature” a good or evil “characteristic”) can be considered “innate,” in the sense that it has “*always been*” present (phenomenally), it nevertheless must have somehow been “adopted” (noumenally), otherwise the person would not be responsible for his or her moral choices. Since it was there from the beginning, the person must have “procured” it in some *nontemporal* manner. This claim surely rates as one of the most difficult to understand in Kant’s entire philosophical corpus, for he offers only a few hints (e.g., at R 31) as to precisely what he has in mind. In §2.2 I shall explain and defend Kant’s position in more detail.

The middle part of this closing paragraph conveys what may be an initial clue to Kant’s opaqueness on the matter of how we procure our original conviction.

25.09–17

Now, the subjective basis or the cause of this adoption¹²⁰ cannot again be cognized (although inquiring about that (*darnach*) is unavoidable,¹²¹ because otherwise one would in turn have to adduce a maxim into which this conviction had been taken up, [and] this /maxim\ must likewise have its basis in turn). Because, therefore, we cannot derive this conviction, or rather its supreme basis, from

¹¹⁷ Following GH for Kant’s *erworben*; WP has “procured” and GG “earned.”

¹¹⁸ For Kant’s *auf*, GH has “up” and GG “on,” while WP takes this as included in the meaning of *immerdar* (“always”). I have rearranged WP’s word order in this parenthetical clause.

¹¹⁹ Following GG for Kant’s *eine einzige*; WP and GH have “one.”

¹²⁰ Following WP, GG, and GH for Kant’s *Annehmung*; this word can also mean “acceptance.” Cf. note 0.87 and Glossary.

¹²¹ All three translators move the closing parenthesis from the end of this sentence, where Kant put it, to this point. *GH* 20n merely states: “our alteration seems necessitated by the meaning.” *GG* 74n offers a more detailed explanation: “The clause starting with ‘for otherwise’ provides no explanation why we should not be asking about the cause, but it makes sense as an explanation of why no further cause can be known.” *WP* 26n agrees: “This is clearly an error”—and offers a similarly brief explanation. But I disagree. Kant’s *darnach* is ambiguous: assuming that it refers to “the subjective basis or cause of this adoption,” these translators render *darnach* as “into/about it”; but, if *darnach* means “into/about that,” then it can refer to the *fact* that the basis/cause “cannot again be cognized.” In that case, the closing

any [particular] first act of volition in time (*Zeit-actus der Willkür*), we call it a make-up of volition belonging to it by nature (even though in fact it has [its] basis (*gegründet ist*) in freedom).

The first sentence, provided we trust the way Kant wrote it (see note 1.121 above), states that we cannot cognize what causes us to adopt our innate moral conviction, but that we must nevertheless explain *why* it is unknowable, because this (Critical) inquiry alone will prevent us from assuming a never-ending series of maxims grounded in ever-deeper causes (cf. *R* 21n). Unlike the will (*Wille*), volition (*Willkür*) acts temporally. Because we are unable to find “any [particular] first act” of the latter (any temporal free choice) that causes us to have the innate moral conviction we find in ourselves, we say that our conviction belongs to human volition “by nature.” What this really means, however, is that its *ultimate* “basis” must lie “in freedom,” the nontemporal “fact” of human nature that Kant calls *Wille*.

Having highlighted our necessary ignorance of what causes us to adopt a particular initial moral conviction, Kant draws the Comment to a close by hinting that the answer to the question at hand, whether our conviction is good or evil “by nature,” will focus *not* on specific individuals but on the *entire race*.

25.17–25

However, that by the human being of whom we say [that] he is by nature good or evil we do not understand¹²² the individual /one\ (since then one /human being\ could be assumed to be *(als)* good, and another *to be* evil, by nature) but are authorized to understand the entire genus—this can be proved only later on, if in anthropological probing¹²³ it turns out that the bases that entitle us to attribute one of the two characters to a human being as innate are so made-up that /there\ is no basis for exempting any human being from it, and that the character *(er)* therefore holds for the genus.

This appeal to proof through “anthropological probing” is rather curious for Kant, who normally relies solely on a priori proofs.¹²⁴ If intended seriously (as I believe it was), this deferral

parenthesis can stay where Kant put it, and his meaning makes sense: even though we cannot cognize how we acquire our innate conviction, we must nevertheless inquire into that fact (i.e., into our ignorance of the cause), otherwise we must postulate an infinite regress of unknown causes. If the translators’ amendment is correct, then Kant would be explaining why the cause of the innate conviction is unknown by claiming that, if it were known, then another maxim together with its basis would have to be discerned. But that on its own offers no reason for stopping one’s inquiries at this level! Kant uses *fragen ... darnach* only one other time in the First Piece, stating that we “cannot inquire into the temporal origin of this deed, but must inquire merely into its rational origin” (*R* 41). In both passages he means that, because we are ignorant of the *temporal* cause of our adoption of a given conviction, we must (in keeping with the goal of Critical philosophizing) inquire into its unknown status as a *rational* cause; only this justifies the cessation of our inquiries at this level. Kant is not justifying why the innate conviction must be regarded as unknown; rather, he is assuming its unknown status and using that to justify why we must inquire into this conviction as rational rather than as temporal, as the next sentence in the current passage clearly states. (See also note 2.46.)

¹²² Following GG for Kant’s *verstehen*; WP has “mean” and GH “be understood.”

¹²³ WP has “investigation” for Kant’s *Nachforschung*; GH and GG have “research.”

¹²⁴ Fenves 2003: ch. 4 interprets this appeal as a direct (though unconscious) contradiction of Kant’s claim to be following a strictly a priori method in *Religion*. Wood 1999: 286, by contrast, sees this as evidence that Kant *intended* his subsequent proof of the radical evil in human nature to be empirical. See SP-2008b for an argument against both positions and MRB-Wood for Wood’s most recent reply to those who see Kant’s argument for evil as having an a priori grounding. Without accounting for the substantial textual basis that leads most readers to *expect* an a priori argument for evil to be present in *Religion*, Wood ironically labels my argument in SP-2008b (and all others that take at face value Kant’s claim to have provided such a proof) as “highly inventive” (54). Yet his allegation smacks of projection, for surely Wood’s own theory—that Kant grounds the very origin of evil in the empirical fact of our “social unsociability”—is not so much as hinted by Kant; indeed, the notion that an empirical solution could be given to a rational/transcendental problem directly contradicts the perspectival thrust of Kant’s whole Critical project. Claims such as Wood’s suffer from a gross failure to appreciate the perspectival difference between the topic of the First Piece

to a posteriori science appears to be motivated by a desire to make it virtually impossible to prove *beyond doubt* that there have been no exceptions whatsoever to the conclusions Kant will reach about human nature in the main sections of the First Piece.

4. Section I: Human nature's original predisposition is good

Because Kant initially published the First Piece as a separate essay, before publishing the rest of the book, its structure differs in several respects from that of the other pieces. Each piece has an untitled introductory section and a concluding General Comment. But, whereas the other three each have two main divisions and a General Comment that deals *only* with one of the four parerga,¹²⁵ the First Piece has no divisions except its four numbered sections, and most of its General Comment section (the portion labeled Section V in *RI*) actually foreshadows the content of the Second Piece. The first numbered section of the First Piece offers Kant's initial (although in a way still preliminary) answer to the question he has been defining and refining up until now. We shall therefore examine that section here.

Kant starts Section I by defining the teleological “elements” that determine every human being's “original predisposition” in terms of three “classes”—a distinction that will influence the structure of many subsequent arguments.

26.01–11 {26}

I. ON THE ORIGINAL PREDISPOSITION IN HUMAN NATURE TO THE GOOD

We can properly apply¹²⁶ this /original predisposition/,¹²⁷ in reference to its purpose, to three classes (*Klassen*), as elements of the determination of the human being:

1. the predisposition to the *animality* of the human being as a *living* /being/;
2. to the *humanity* of him (*desselben*) as a living and at the same time *rational* /being/;
3. to his *personality* as [that of] a being /who is/ rational and at the same time *capable of imputation* [[of actions to him]].*

Suddenly and with very little warning,¹²⁸ Kant now takes a definite stand on how the question of human nature should be answered. We know only from the section's title that he intends to argue that our predisposition is good. It is proper or “fitly” (*füglich*), he declares, to divide the

(i.e., the source of evil *in reason*) and that of the Third Piece (i.e., the *solution* to evil in community building). Rather than hinting at a (nonexistent) empirical origin of evil, Kant is here leaving a “space” for a person to be born *without* the usual innate conviction that, for the rest of the race (as we shall see in Ch. 2), is evil. This explanation dovetails nicely with Kant's expressed goal of performing *two* “experiments” (see Introduction §4): only the second experiment needs an appeal to anthropological research, not the first (see §4.4).

¹²⁵ As I explained earlier, I comment on the text of each of the four parerga in an appendix at the end of the corresponding part of this commentary.

¹²⁶ For Kant's *auf ... bringen*, GH's “divide ... into” makes sense but is too loose, GG's “bring ... under” is too literal and awkward English, while WP's “reduce ... to” is counterintuitive. The context dictates that Kant is dividing *one* thing into *three* classes, whereas reduction moves in the opposite direction. By contrast, “apply” (the primary meaning of *aufbringen*) fits the context well.

¹²⁷ WP, GG, and GH have “this predisposition,” but Kant's *sie* (“it”) refers specifically to the predisposition named in the title, the *original* one.

¹²⁸ Prior to the title of Section I, Kant used *Anlage* (“predisposition”) only twice: in the first paragraph of the First Piece, briefly referring to those who believe that a good predisposition can help reverse the trend from bad to worse in human development; and in the footnote to that paragraph, where he suggests that nature might come to our aid in supporting this “moral predisposition.”

predisposition into three progressively advanced “classes” of human nature,¹²⁹ each of which he will discuss in a separate subsection. What is apparent to anyone sensitive to the architectonic underpinnings of Kant’s method of system building is that these three classes correspond directly to the standpoints of the three *Critiques*: the fact that we are *living* makes us animals, so the first *Critique*’s theoretical standpoint begins the Critical System by establishing the limits placed on reason by our sensibility (i.e., by *nature*, the phenomenal source of our inclinations); the fact that we can be held *accountable* for our actions makes us persons, so at the other extreme the second *Critique*’s practical standpoint establishes our potential to transcend these mortal limits through the ideas of God and immortality (i.e., through *freedom*, the noumenal source of the duties that bind our will, *Wille*); and, mediating between these extremes, the fact that we are *rational* makes us human, so the ultimate task of the Critical System is to adopt the third *Critique*’s judicial standpoint (shared by *Religion*) in order to establish how nature and freedom can coexist in a being who seeks, paradoxically, to embody goodness through free volition (*Willkür*).

Kant makes a similar distinction at *CJ* 210 (§5), between the three types of “satisfaction” (Pluhar’s translation uses “liking”), which relate “to *inclination*, or to *favor*, or to *respect*.” These are aroused by “what GRATIFIES us” (the agreeable), “what we just LIKE” (the beautiful), and “what we ESTEEM” (the good), respectively. He correlates these to a triad strikingly similar to the one he employs here in *Religion*: animals, humans, and rational beings in general. Of the three, “only the satisfaction involved in taste for the beautiful is disinterested and *free*,” because the first and third types are grounded in the faculty of desire, while the second is grounded in the faculty of pleasure and displeasure. That is, the first and third types focus on “interest” in the existence of a thing and thus constitute *the* basic conflict that is the main topic of the second *Critique*—namely the conflict between every human being’s desire to fulfill our inclinations and our desire to act out of respect for the moral law. What this *CJ* passage states more clearly than anything Kant writes in *Religion* is that the *middle* predisposition (i.e., our humanity) is what makes us “free” to engage in this struggle, because (as Kant is about to state: see *R* 27.04–12) rationality is essentially our ability to *compare*—in this case, to compare inclinations with respect. In this way the volition that forms such a central feature of Kant’s argument, especially in the First Piece, arises directly out of the position human beings occupy, at the crossroads between two forms of necessitation.¹³⁰ This also explains why each

¹²⁹ Kant’s use of “predisposition” sometimes in the singular and sometimes in the plural is the source of considerable confusion, some commentators assuming that he proposes three (or more) entirely distinct predispositions. Admittedly, in relation to education, Kant says: “Many germs lie within humanity, and now it is our business to develop the natural predispositions proportionally and to unfold humanity from its germs” (*LP* 445; cf. *IUH* 18). One way to understand Kant’s reference to a *single* predisposition in *Religion* is to compare it to a triangle: the triangle is *one entity*; yet, if we look at its sides one by one, they cannot be recognized as aspects of a triangle until the third side is added. Likewise, animality and humanity, on their own, do not yet constitute *the* (or even *a*) predisposition to good; rather, they can be seen as good only when viewed as classifications of our *overall* predisposition, with the third class firmly in place. Only when Kant’s theory is viewed in this way can we appreciate how the inevitable corruption of the first two classes can nevertheless be put to one side, so that goodness can be seen *in* our animality (our inclinations to bodily self-love) and *in* our humanity (our inclinations to social self-love). For alternative ways of interpreting Kant’s use of the plural in the case of humanity, see note 1.147 below.

¹³⁰ This correlation between *CJ*’s basic threefold distinction and *Religion*’s theory of the three classes of predisposition provides further evidence that *Religion*’s standpoint corresponds to that of *CJ* more closely than to that of *CPrR*. That *humanity* aligns most closely with *CJ* and *free volition* is aptly illustrated not only by the choice between good and evil (the focus of most scholarly discussions of *Willkür*, because, as we shall see in Chapter 2, the propensity to evil arises directly out of this choice), but also by the choice that human beings have (as we shall see in Part III) between various empirical religions: we choose a particular faith because its symbols and rituals satisfy our aesthetic taste, promoting the kind of spirituality that will be most likely to empower us to live a moral life. The point here is that the focus in *Religion* is not on defining the moral endpoint of this story (i.e., on “personality”); that was fully explicated in *GMM* and *CPrR*. Rather, the focus is now on the *aesthetic* path that best leads us human beings to the goal of realizing our humanity.

class of our predisposition counts as a predisposition to *good*, even though (as we shall see) the first two are easily corrupted: only when the three classes are *taken together* does freedom arise.

Having posited this architectonic distinction without defense or explanation, Kant launches into a series of three arguments, demonstrating how each classification predisposes us to be good. Fortunately Kant first adds a footnote, once again filling a gap in his exposition and clarifying the third (and most important) class.

26n.21–32

*One cannot treat this /predisposition\ as already contained in the concept of the preceding /one\, but must necessarily regard it as a special predisposition. For from [[the fact]] that a being has reason it does not at all follow—at least as far as we can comprehend¹³¹—that this /reason\ contains a capacity to determine volition unconditionally, through the bare presentation of the qualification of its maxims for universal legislation, and thus to be practical on its own. The most rational of all beings of the world (*Weltwesen*) might yet always need certain incentives, coming to him from objects of inclination, in order to determine his volition, but might apply to this the most rational deliberation as regards [finding] the greatest sum of incentives as well as the means for achieving¹³² the purpose determined by them (*dadurch*), without suspecting even¹³³ the possibility of such a thing as the moral, absolutely commanding law, which proclaims itself as itself, namely, as the highest incentive.¹³⁴

From the human perspective (“as far as we can comprehend”), it seems quite possible that a being might be rational—able to formulate universal rules of thought to determine his or her choices—without being moral: in other words, a being can make such choices on the basis of an internally given incentive, which presents itself as determining our choices “unconditionally” and “absolutely.” This is why Kant refers to freedom and the self-legislation of the moral law as the one *fact* of practical reason (see, e.g., *CPrR* 31). We know that moral self-legislation is possible because we *experience* it, not because it is analytically implied by the very possibility of thinking rationally.¹³⁵ This will have an important consequence for Kant’s consideration of wickedness or *devilish* thinking, as we shall see in Chapter II.

The footnote concludes by marveling at the unlikelihood of us human beings ever *discovering* or *creating* for ourselves a moral incentive, if it were not presented to us—as if it were an inward gift.

26n.32–37

If this law were not given within us, no reason would ever enable us to excogitate it as *such*,¹³⁶ or to talk volition into it; and yet this law is the only /thing\ that makes us conscious of the independence of our volition from determination by any other incentives (conscious of our freedom) and thereby at the same time of the imputability (*Zurechnungsfähigkeit*) of all actions.

¹³¹ WP, GG, and GH have “see” for Kant’s *einsehen*.

¹³² WP, GG, and GH have “attaining” for Kant’s *erreichen*.

¹³³ WP has “so much as” for Kant’s *auch nur* and places it after “suspecting”; GH has “ever” and GG “thereby even.”

¹³⁴ WP has “that it is itself an incentive, and moreover the highest one” for Kant’s *sich als selbst und zwar höchste Triebfeder*; GH has “that it is itself an incentive, and, indeed, the highest” and GG “to be itself an incentive, and, indeed, the highest incentive.” The German here is so awkward that KV suggested moving *als* after *zwar*, thus doing away with “itself as itself.”

¹³⁵ Kant’s arguments in *GMM* sometimes adopt the latter (analytic) strategy. This need not be regarded as contradicting the synthetic arguments of *CPrR*, for these two approaches are meant to be complementary. See *PSP* III.3–4 for an account of how these works relate to *Religion* in the overall structure of Kant’s system.

¹³⁶ WP has “as a law” for Kant’s *als ein solches*; GH has “into existence” and GG “on our part,” paraphrases apparently based on these words. Kant’s meaning seems to be “as such a *given*,” since *es* (“it”) already refers to the law. Preserving the ambiguity of Kant’s grammar therefore seems advisable.

Our awareness that this inner law imposes itself onto our volition is what makes us free (i.e., able to *reject* other incentives) and thereby accountable for what we do. Although in the First Piece Kant has not yet mentioned the need for a higher being as the ultimate source of this law (cf. R 4–5), by referring to it as “given” this footnote seems to be preparing the way for this crucial feature of what distinguishes religion from bare morality.

In the first of the three numbered subsections that argue for the goodness of the human predisposition, Kant interprets our animal nature—what we might nowadays call our “embodiment”¹³⁷—in terms of the physical need of members of the human species to look after their own well-being (i.e., the need for “self-love”).

26.12–18

1. The predisposition to ANIMALITY in the human being can be brought under the general title of a physical and merely *mechanical* self-love, i.e., a kind /of self-love\ for which reason is not required. This /predisposition\ (*Sie*) is threefold: *first*, to the human being’s *(seiner)* preservation of himself; *second*, to the propagation of his species *(Art)* through the impulse *(Trieb)* to sex, and to the preservation of what is *generated*¹³⁸ therewith through *intermingling*,¹³⁹ *third*, to community with other human beings, i.e., the impulse to society. —

Kant’s use of “mechanical” should not be taken too literally, as if animal bodies engaging in self-loving behaviors *are* machines; rather, he is alluding to the *machine-like* operation of physical processes that are determined by the law of causality, as defined in *CPR*. We, like all other animals, do not need to think rationally in order to engage in behavior aimed at preserving our own life or the lives of other members of our species. Our rational capacity is as unnecessary for feeding, fleeing, or fighting as it is for reproducing and nurturing offspring or for forming social bonds with other members of our species. The argument here is merely implicit, but is fully consistent with modern biological science: like all animals, we instinctively behave in self-loving ways, and this predisposition is *good* for the human species. Why? Because (as subsequently demonstrated by Darwin) the behavior caused by it, such as the incest taboo (see note 1.139 above), enhances the likelihood of our *survival*.

Instead of making the above argument explicit, Kant concludes this brief subsection by arguing that, although the same forms of behavior can sometimes be used in evil ways, such uses are *derivative*, not grounded in the predisposition as such.

26.18–20, 27.01–03

On it can be grafted all sorts of vices (which, however, do not *sprout*¹⁴⁰ on their own *(von selbst)* from that predisposition as a root). They may be named vices of the *coarseness*¹⁴¹ {27} of nature, and in their utmost *(höchsten)* deviation from the natural purpose they are called *bestial vices*, of *gluttony*, of *lust*, and of *savage lawlessness* (in relation to other human beings).

¹³⁷ One of the most common criticisms of Kant’s philosophy is that it presents us human beings as if we were disembodied moral agents (see, e.g., Lakoff and Johnson 1999: ch. 20). Recently more and more scholars are recognizing, however, that nothing could be further removed from the truth (see, e.g., Shell 1996, Svare 2006, and Bunch 2010). Throughout this commentary we shall see that in *Religion* Kant regards our embodiment as *the* crucial factor about human nature that requires us to be religious: at its best, religion is, for Kant, *embodied* morality.

¹³⁸ WP has “produced” for Kant’s *erzeugt*; GH and GG have “begotten.”

¹³⁹ WP has “copulation” for Kant’s *Vermischung*; GG has “breeding,” while GH omits this word, rendering the phrase as “of offspring so begotten.” But Kant’s use of this specific term seems to be an allusion to the intermingling of the *sexes* (i.e., to heterosexuality as opposed to homosexuality) and/or to what we, in the post-Freud age, call the incest taboo: the instinct to *intermingle* rather than having sex with close relatives or same-sex partners serves to preserve offspring better.

¹⁴⁰ WP has “spring ... from” for Kant’s *entspringen*; GH has “spring from” and GG “issue from.”

¹⁴¹ WP has “crudeness” for Kant’s *Rohigkeit*; GH has “beastly” and GG “savagery.” I use “coarse” for *roh* (literally “raw”), to highlight its contrast with what is well developed or cultured. At *CJ* 303 Kant similarly contrasts *grob* (“coarse”)

Our impulse to preserve the species through healthy forms of self-love is natural, and therefore essentially good. Yet we pervert it whenever we give it unrestrained reign—for example, by eating or drinking too much; our predisposition to procreate is likewise essentially good but goes awry when we engage in forms of sexual activity that contradict its purpose;¹⁴² and, when we fail to temper our social relations by conforming them to the rule of law, we spoil them too. Although Kant's argument remains implicit, it is quite clear: the fact that these three aspects of our animal nature are *perverted* when we use them improperly enables us to infer that the plant¹⁴³ itself is originally *good*, enabling us to stay alive.

Kant's argument in defense of the second claim, that the rational (or “human”) aspect of our predisposition is good, is more explicit.

27.04–12

2. The predispositions to HUMANITY can be brought under (*auf*) the general title of a no doubt (*zwar*) physical but yet *comparing* self-love (for which reason is required): namely, to judge oneself as *fortunate* or *unfortunate* only by comparison with others. From this /self-love\ stems the inclination to *procure a worth for oneself in the opinion of others*, originally, to be sure (*und zwar*), merely that of *equality*: to permit no one superiority over oneself, *linked* with¹⁴⁴ a constant worry that others might strive for this, from which arises gradually an unjust desire to *acquire it* over others.¹⁴⁵—

What separates rational from merely animal beings is the ability to *compare* oneself with other members of one's species.¹⁴⁶ Just as the (physical) predisposition to *animality* is a form of

with *schön* (“fine” or “beautiful”), describing the former as someone having “no feeling for beautiful nature,” but whose enjoyment is limited to “mere sensory sensations at table or from the bottle.” This distinction between the undeveloped or *uncultured* and the cultured is closely related to the animality–humanity distinction, as Kant explains in *CBHH* 115: “the departure of the human being from the paradise which reason represents to him as the first abode of his species was nothing other than the transition from the crudity [*Rohigkeit*] of a merely animal creature into humanity, from ... instinct to the guidance of reason—in a word, from the guardianship of nature into the condition of freedom.” Only by keeping in mind that animality is also a predisposition to *good*, though the one that is most easily corrupted, can we properly understand Kant's use of *roh*; such persons are not *more evil* than other persons but merely *less educated*.

¹⁴² Kant develops his theory of human sexuality in *MM* 277–80. While his views appear sexist to most readers, I have argued (in an unpublished paper that I hope to make the basis of a book tentatively entitled *Egalitarian Sexism*) that in many cases, when properly understood, they are not sexist in any objectionable sense. On the convention of linking sexual perversion to nature's purposes, see *SP*-2003: §§10, 15.

¹⁴³ Kant's organic understanding of human nature is aptly illustrated by his use of the grafting metaphor in this passage and of related metaphors throughout the book, for example “seed,” “tree,” “root,” etc. The predisposition is like a plant whose ability to bear good fruit is compromised by the foreign influences that, as we shall see in Chapters 2 and 3, corrupt it at its root.

¹⁴⁴ WP has “combined with” for Kant's *mit ... verbunden*; GH has “bound with,” and GG “bound up with.”

¹⁴⁵ WP has “to gain superiority” for Kant's *sie sich ... zu erwerben*; GH has “to win it for oneself” and GG “to acquire superiority for oneself.” WP 29n points out the similarity between Kant's view here and that expressed in Rousseau's *Émile*. The position advanced in this subsection is also a further development of *IUH*, where Kant introduces his notion of the “unsocial sociability” that characterizes human community building. However, despite Wood's insistent wish to the contrary (cf. note 1.124 above), the proper place to discuss this position, given the architectonic structure of *Religion*, is not here but at the beginning of the Third Piece, where Kant more explicitly introduces the need for community as the ultimate solution to the problem of moral evil.

¹⁴⁶ That Kant assumed that only human beings are able to exercise the low-level rational activity required to compare oneself with others should not be held against him, given the state of science in his day. Nowadays, scientific research would require us to treat this second class as a feature of our predisposition that we share with at least some other mammals, since the behaviors Kant discusses here have also been observed in the so-called animal kingdom. In *CJ* 355–6 Kant also associates humanity with our ability to communicate, adding that this requires “freedom ... with some constraint.”

self-love, so, too, the (mental) predispositions¹⁴⁷ to *humanity* are manifested as a tendency to prefer (or love) one's self over others. Everyone who sees another person who appears to be fortunate naturally wishes to be at least as fortunate as this other person. This, in turn, inclines us to be well behaved, so that others may view us as persons of worth. Provided that it goes only this far, the competition that naturally arises out of our tendency to compare ourselves with each other is evidence that our original predisposition is good.

Kant has already observed that, unfortunately, just as we easily pervert our animal nature by using physical self-love as an excuse for vice, we also easily pervert our human nature by using rational self-love to seek not equality, but an inappropriate level of superiority. He thus continues:

27.12–21

On this, namely on *jealousy* and *rivalry*, can be grafted the greatest vices of secret and overt hostilities against all whom we regard as alien to us—[vices] that actually do not, after all, sprout on their own from nature as their root; rather, in view of (*bei*) the worrisome¹⁴⁸ endeavor of others to gain a (to us hateful)¹⁴⁹ superiority over us, they are inclinations to procure, for security's sake, superiority (*diese*) over others as a preventive measure [for] ourselves, even though (*da ... doch*) nature wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) only as an incentive to culture.

Kant now makes his overall mode of argument in this section fully explicit. The fact that our predisposition, as rational beings, to compare ourselves with others is closely related to problems that plague the whole human race, to “the greatest vices” that stem from “*jealousy* and *rivalry*,” is indisputable. However, these evils are “grafted” onto the predisposition; they do not “sprout on their own from nature as their root.” If the “competitiveness” rooted in our human predisposition gives rise to evil only when we allow our inclinations to dominate the ability in question (i.e., our rationality), then that underlying predisposition must be good in and of itself. This predisposition to compare ceases to be a healthy form of self-love and gives rise to evil when, in hopes of preventing others from gaining superiority over us, we try to *appear* superior to others, even when our achievements do not merit such an assessment.

Kant concludes this second subsection by giving a new twist to the claim that our predisposition to make rational comparisons is essentially good.

27.21–26

The vices that are grafted on this inclination may therefore also be named vices of *culture*, and in the highest degree of their wickedness¹⁵⁰ (because they are then merely the idea of a maximum of evil that surpasses humanity), e.g., in *envy*, in *ingratitude*, in *gloating*,¹⁵¹ etc., they are called *diabolical vices*.

¹⁴⁷ Kant never explains why he uses the *singular* (“predisposition”) for animality and personality, but the *plural* (“predispositions”) for humanity. While this might be simply a reflection of the logical distinction between a set and its three classes (see note 1.129 above), another possibility is that he is thinking of males and females as having fundamentally distinct forms of human predisposition. Kant is notorious for holding views many nowadays take to be sexist (see note 1.142 above); however, his previous reference to sex under the heading of *animality* makes this unlikely. More plausible is that the plural usage foreshadows the theological claim that some human beings are predisposed to good and others to evil—a possibility he considers in the third General Comment (see App. III) under the heading “The mystery of the *calling*” (R 142–3), but neither rejects nor condones.

¹⁴⁸ WP has “feared” for Kant’s *besorgen*; GH and GG have “anxious.”

¹⁴⁹ WP and GH have “hated” for Kant’s *uns verhaßten*; GG has “hateful.” They all ignore Kant’s *uns*, yet this makes explicit that Kant is referring to *our* hatred. I add parentheses to clarify that the whole clause modifies “superiority.”

¹⁵⁰ GH and GG translate Kant’s *Bösartigkeit* more literally, as “malignancy,” an option that fits in nicely with Kant’s metaphor of human nature as a diseased plant. Another good option would be “maliciousness.” But this causes ambiguity at R 37, as WP points out in his persuasive defense of his translation at WP 41n.

¹⁵¹ WP has “malicious glee” for Kant’s *Schadenfreude*; GH has “spitefulness,” while GG translates the term literally, as “joy in others’ misfortunes.” BRR 227–31 traces Kant’s appeal to these “vices of culture” back to the influence of Rousseau, Shaftesbury, and Iselin.

The competitive instinct that arises out of our predisposition's second aspect has the essentially good purpose of motivating groups of humans banding together to develop distinct *cultures*. Cultures appear wherever rational beings (those capable of comparing) agree to temper the desire each individual has for superiority through a "reciprocal love" that enables them to work together for a common good. Obviously, culture is not necessarily good; but Kant explains this fact as a result of us human beings grafting vices onto our original predisposition. The closing sentence hints at a notion that Kant develops further in Section III: some types of evil reach such a "maximum" intensity that they go beyond what we normally regard as within the realm of "human" comparisons; this results in "diabolical vices," whereby a rational being wishes *positive evil* for another. Again, the fact that we humans do sometimes experience envy, gloating, and so on, is secondary to (or a derivative from) the more basic fact that the rational ability that makes such evil possible (our ability to make comparisons in the first place) is fundamentally good.

The longest and obviously most important subsection arguing that our predisposition is good is the third, since it focuses on our moral nature as persons.

27.27–34

3. The predisposition to PERSONALITY is the receptivity to respect for the moral law, *as an incentive, sufficient by itself, of volition*. This receptivity to bare respect for the moral law within us would be the moral feeling, which by itself does not yet amount to a purpose of the natural predisposition, but amounts to such a purpose, only insofar as it is an incentive of volition. Now, since this becomes possible solely through free volition's taking up the moral feeling (es) into its maxim, the make-up of such a volition is a good character.

Here, at last, Kant provides a more or less complete argument. Not all rational beings are persons, but only those who make respect for the moral law the sole and sufficient force motivating their volition. Apparently only those rational beings who have a good character deserve to be called *persons* in the fullest sense; nevertheless, *all* human beings are persons, at least to the extent of having a *predisposition* to make this crucial choice. That is, everyone has an innate "receptivity" to feel respect for the moral law that presents human volition with a free choice; but only those who say *yes* to the offer of freedom enter fully into the realm of personhood.¹⁵²

Given Kant's emphasis in the first two subsections on perversions of the good predisposition, his next remark may come as quite a surprise.

27.34–37, 28.01–07

Such /a character\,¹⁵³ as generally, every character of free volition, is something that can only be acquired, but for the possibility of which there must nevertheless¹⁵⁴ be present in our nature a predisposition on which absolutely nothing evil can be grafted. The {28} idea of the moral law alone, with the respect inseparable from it, cannot properly be called a *predisposition to personality*; it is personality itself (the idea of humanity considered entirely intellectually). But that we take up this respect into our maxims as an incentive¹⁵⁵—the subjective basis for this seems to be an addition to personality and thus seems to deserve the name of a predisposition on behalf of it.

¹⁵² For a comprehensive set of 67 essays dealing with various aspects of Kant's theory of human personhood, see SP-2010a. In particular, Kawamura 2010 highlights the special character of Kantian personhood as more of a "task" to be undertaken than a given character trait that all human beings necessarily possess.

¹⁵³ Kant merely has *welcher*, as the German here continues one long sentence.

¹⁵⁴ WP has "nonetheless" for Kant's *dennoch*; GH and GG omit this word.

¹⁵⁵ WP omits Kant's *zur Triebfeder*; GH has "as a motivating force" and GG "this incentive" (omitting Kant's *Achtung*).

Unlike the physical and rational aspects of our predisposition, this third aspect *cannot* be perverted, because respect for the moral law just “is personality itself.” Those who allow such respect, rather than the desires stemming from human inclinations, to serve as their “incentive” for maxim formation must have a “subjective basis” for doing so. That determining basis of volition must therefore be an “absolutely” good predisposition that exists in everyone.¹⁵⁶ A human being has “personality” only to the extent that he or she *acts* on this predisposition, because there is no personality apart from respect for the moral law; but everyone must have the *predisposition* in order for this free choice of a good character (i.e., “the idea of humanity”: the second predisposition, considered in abstraction from animality, the first) to be possible at all. Given this requirement, the predisposition to personality can be regarded as an “addition” to our nature as human *animals*, encouraging us to respect the moral law; and this predisposition is an unmitigated good, even though not all human beings necessarily *become* good—given that some do not actually take up this “subjective basis” into their maxims in response to such respect.

Kant concludes Section I of the First Piece with a paragraph that presents a brief transcendental reflection¹⁵⁷ on the status of the above three applications of the predisposition to good; perhaps inconsistently, he now refers to them as three separate predispositions.

28.08–14

If we consider the three mentioned predispositions according to the conditions of their possibility, we find that the *first* is rooted in no reason; the *second* indeed [in] practical reason, but only [as] subservient to other incentives; but the *third* alone [in] reason practical on its own, i.e., legislative unconditionally. All these predispositions in the human being are not only (negatively) *good* (they do not conflict with the moral law) but are also predispositions *to the good* (they *promote*¹⁵⁸ compliance with that [law] \ *desselben*)).

This hierarchy in the transcendental relation between practical reason and the three classes of predisposition explains why only the first two can be perverted. Each type positively predisposes us to be morally good, because each applies to our *relations* with our fellow human beings. However, our *animal* nature is easily perverted as soon as it is employed, because it has no rational conditions constraining us to exercise self-love according to principles; our *human* nature is constrained by concerns arising out of the pragmatic side of practical reason (i.e., prudence), but at this level we tend to give priority to the inclinations arising out of our natural self-love, so our powers of comparison are also easily perverted right from the start. Only the third predisposition gives explicit priority to morally practical reason, thereby pointing us firmly and solely in the direction of goodness. Yet, as we shall see in Chapter 2, the mere presence of this predisposition does not prevent us from being overcome by evil at the very outset of our moral experience; rather, as long as we have not actively chosen to base all of our moral choices on this third predisposition, its mere *presence* is, ironically, what makes us *evil* at the onset of our moral development.

¹⁵⁶ To readers familiar with Kant’s moral writings, this claim will not come as a surprise, for he famously argues at the beginning of *GMM* that only a good will can be called absolutely good.

¹⁵⁷ “Transcendental reflection” is Kant’s technical term for thinking about the conditions that make something possible. I have argued in *PSP IV* that each system in his Critical System begins with a stage that adopts this transcendental perspective in order to determine the limits or boundary conditions of the overall subject under consideration. The paragraph of *Religion* now under consideration suggests that Kant was doing essentially the same thing here, so that the First Piece can also be regarded as stage one in the first of his two “experiments” (see Introduction §4)—i.e., in the argument defending his system of rational religion.

¹⁵⁸ WP has the somewhat awkward “further” for Kant’s *befördern*; GH’s “enjoin” is rather loose, while GG’s “demand” is potentially quite misleading.

Kant continues by explaining that, even though the first two can be misused, all three predispositions to good are “original” and thus always present, in the transcendental sense that we must presuppose them in order to conceive of how human nature is possible.

28.14–17

They are *original*; for they belong to the possibility of human nature. The human being can indeed use the first two contrapurposively, but cannot exterminate¹⁵⁹ either of them.

This is the climax of Kant’s first main argument. Its thrust cannot be appreciated without recognizing the perspective he is assuming. To call our predisposition(s) good is to make not an *empirical* claim about our experience as human beings, but a *transcendental* claim about the origin (i.e., the very possibility) of human nature. Human beings just *are* animals conditioned through practical reason by an unconditioned, self-imposed inner law. This fact about our nature predisposes us to be good before we ever perform our first moral act; but it does not make us *empirically* good, especially because two of these predispositions are so easily used for purposes other than what is good. Although its first implication is to make us *aware* of, and thus responsible for, our shortcomings, the predisposition to good can never be completely wiped out, no matter how badly we may pervert it.

The concluding paragraph of Section I, and so also of the first step in Kant’s exposition of the first stage of his system of religion within the bounds of bare reason, now ends with a (belated) definition of “predispositions.”

28.17–24

By the predispositions of a being we understand the constituent components, as well as the forms of their association, that are required in order to be such a being. They are *original* if they belong to the possibility of such a being necessarily, but *contingent* if the being would intrinsically also be possible without them. It should be noted, still, that here no other predispositions are at issue than those that refer immediately to our capacity for desire (*Begehrungsvermögen*) and the use of our volition.

Kant’s first way of answering the main question of the First Piece—whether human beings are good or evil by nature—has been to examine the necessary conditions for being human, insofar as these relate “to our capacity for desire,” the rational faculty that governed Kant’s considerations in *CPrR*.¹⁶⁰ As creatures of desire who are “condemned to be free” (as Sartre 1966: 186 and 567 later put it) in the way we use our volition, we are animals who *must* choose a rational principle to govern our desires. Our original predisposition to good is not a “contingent” possibility; it is not one that we could do without and still be human persons; rather, it is necessary for the very possibility of our nature that we desire to remain alive, to compare ourselves with others, and to respect the moral law. This predisposition therefore functions as the first aspect of what (in the three *Critiques*) Kant would have called the *transcendental boundary* defining his topic. The next chapter will deal with a very different aspect of this boundary: we will examine how Kant answers the same question as it applies to our *actual choice* of moral incentives.

¹⁵⁹ WP and GH have “extirpate” for Kant’s *vertilgen*; GG has “eradicate.” My translation highlights that here Kant is yet again thinking metaphorically.

¹⁶⁰ Kant concludes the Introduction to the third *Critique* (*CJ* 197) by noting that the proper domain of practical reason is the capacity for “desire,” that of theoretical reason is the “cognitive” capacity, and that of judging (or “judicial” reason, as I call it) is “the feeling of pleasure and displeasure.” That Kant focuses here on desire should not eclipse the fact that the overall concern of *Religion* is more comparable with that of *CJ*; see Introduction §1 and notes 1.124 (above) and 5.114.