

# Religion within the Limits of Bare Reason

Immanuel Kant

Copyright © Jonathan Bennett 2017. All rights reserved

[Brackets] enclose editorial explanations. Small ·dots· enclose material that has been added, but can be read as though it were part of the original text. Occasional •bullets, and also indenting of passages that are not quotations, are meant as aids to grasping the structure of a sentence or a thought. Every four-point ellipsis . . . . indicates the omission of a brief passage that seems to present more difficulty than it is worth. Longer omissions are reported between brackets in normal-sized type.—Some footnotes are presented in the main text instead of at the bottom of the page; this is because of formatting problems; the reasons are aesthetic, and have nothing to do with content.—Passages starting with † were added in the second edition (see page 6).

First launched: October 2013

## Contents

Preface to the First Edition . . . . .	1
Preface to the Second Edition . . . . .	6
<b>First Essay: The bad principle existing alongside the good, i.e. The radical evil in human nature</b>	<b>8</b>
1. The Original Predisposition to Good in Human Nature . . . . .	11
2. The Propensity to Evil in Human Nature . . . . .	13
3. Man is bad by Nature . . . . .	15
4. The Origin of Evil in Human Nature . . . . .	19
General remark: Restoring the Original Predisposition to Good to its Power . . . . .	22

<b>Second Essay: The conflict of the good with the bad principle for command over man</b>	<b>29</b>
1. The good principle's legal claim to dominion over man . . . . .	31
A. The personified idea of the good principle . . . . .	31
B. The objective reality of this idea . . . . .	32
C. Difficulties that oppose the reality of this idea, and their solution . . . . .	34
2. The bad principle's legal claim to dominion over man, and the conflict between the two principles . . . . .	43
General remark . . . . .	46
<b>Third Essay: The victory of the good over the bad principle, and the founding of a kingdom of God on Earth</b>	<b>50</b>
1. Philosophical account of the good principle's victory in founding a Kingdom of God on Earth . . . . .	51
A. The ethical state of nature . . . . .	51
B. Man ought to leave his ethical state of nature in order to become a member of an ethical commonwealth . . . . .	52
C. The concept of an ethical commonwealth is the concept of a people of God under ethical laws . . . . .	53
D. The only way humans can bring about a people of God is through a church . . . . .	54
E. The constitution of every church originates in some historically revealed faith (call it ecclesiastical faith), which is best based on a holy scripture . . . . .	55
F. Ecclesiastical faith has pure religious faith as its highest interpreter . . . . .	59
G. The gradual transition of ecclesiastical faith to the exclusive sovereignty of pure religious faith is the coming of the Kingdom of God . . . . .	62
2. Historical account of the gradual establishment of the dominion of the good principle on Earth . . . . .	68
General remark . . . . .	77
<b>Fourth Essay: Service and pseudo-service under the sovereignty of the good principle, or Religion and <i>Pfaffentum</i></b>	<b>83</b>
1. The service of God in religion as such . . . . .	84
A. The Christian religion as a natural religion . . . . .	86
B. The Christian religion as a scholarly religion . . . . .	90
2. The pseudoservice of God in a statutory religion . . . . .	92
A. The Universal Subjective Basis of the Religious Illusion . . . . .	93
B. The Moral Principle of Religion Opposed to the Religious Illusion . . . . .	94
C. <i>Pfaffentum</i> as a Government in the Pseudoservice of the Good Principle . . . . .	97
D. The Guide of Conscience in Matters of Faith . . . . .	103
General Remark . . . . .	106

## Glossary

**anschaulich:** This is left untranslated on page 30 because no English word or short phrase quite does the job. To make something *anschaulich* is to make it—in this case metaphorically speaking—solid, something we can grab onto, push around, manipulate.

**archetype:** Translates Kant's *Urbild*, and means 'model' or 'prime example'—something to be followed or copied..

**atonement:** This English word comes from the notion of two people—e.g. a sinner and God—being *at one*; that goes with Kant's mention on page 63 of *Versöhnung* = 'reconciliation', suggesting that the core notion concerns God's attitude to the sinner, almost how he feels about him. But the word translated—wrongly but unavoidably—as 'atonement' is *Genugthuung*, which comes from *genug* = 'enough', 'sufficient'; the thought is that of reparation, paying a penalty. That is the emphasis all through the third Essay: Kant speaks of it as *legally* undoing what you have done; his phrase *Bezahlung für seine Schuld* means 'reparation for his guilt' and equally well means 'payment of his debt'.

**change of heart:** This nearly always translated *Sinnesänderung*, literally = 'change in thinking' or 'change of mentality'. On pages 24, 38 and 42 it translates *Herzensänderung*, literally = 'change of heart'. There's no evidence that Kant intended a distinction here, and much that he didn't.

**chiliasm:** 'The belief that Christ will reign in bodily presence on earth for a thousand years' (OED).

**constitutive:** A constitutive principle, for Kant, is a principle saying that such-and-such is the case, rather than serving merely as advice or recommendation or the like. (Cf. 'regulative', below.)

**debt:** This translates *Schuld*, which also means 'guilt'. In many passages Kant clearly means both at once, with 'debt' as a kind of metaphor for 'guilt'.

**deduction:** In Kant's terminology, the 'deduction' of an idea is an intellectual process in which the idea is introduced and in some way defended or justified.

**determine:** The basic meaning of 'determine' is *settle, fix, pin down*; thus, to determine what to do next is to decide what to do next, to settle the question. When on page 9 Kant says that in a morally bad action the will can't be 'determined' by anything outside it, the word conveys the notion of *fixed*, which would rule out freedom.

**duty:** This translates *Pflicht*, which Kant uses as his all-purpose name for what one morally ought to do. Most English-language moral philosophers also use 'I have a duty to do A' to mean 'I morally ought to do A'; but that isn't what it means in good standard English, where the term 'duty' is tightly tied to jobs, roles, social positions. The duties of a janitor; the duties of a landowner.

**evil:** This as a noun translates *Böse* and means merely 'something bad'. (The corresponding adjective (*böse*) is translated here by 'bad', so as to avoid loading it with all the force 'evil' has in English when used as an adjective.) For the noun, 'evil' is used because we don't have 'bad' as a noun as we have 'good' ('friendship is a good'). This has become a standard philosophical usage—e.g. 'the problem of evil' means 'the problem posed by the existence of bad states of affairs'.

**idea:** In Kant's terminology an 'idea' is a concept that comes from or belongs to *reason*, as distinct from the concepts

belonging to *the understanding*, which are what we use in thinking about the contingent empirical world.

**ideal:** As a noun this translates Kant's *Ideal*, a technical term which he explains in the first *Critique* at B 595–7, and is still using in the same sense here. An ideal is an idea [see above] which is the idea of an individual thing. The idea of *perfect moral purity* is not an ideal, the idea of *God* is an ideal. Kant does think of ideals as things we can steer by, try to live up to, etc., but the core meaning is that of 'idea of an individual'. When this word first occurs here (on page 31) Kant moves rapidly between 'idea' and 'ideal'; but that is harmless, because any ideal is an idea.

**illumination:** 'A doctrine involving belief in or a claim to intellectual or spiritual enlightenment' (OED).

**man:** This translates Kant's *Mann* and (more often) his *Mensch*. The latter can be translated as 'human being', but in this version 'man' has been preferred as less fussy. On page 21 the biblical narrative of The Fall is of course really about a woman, Eve.

**personality:** In uses starting on page 12 the word refers to the condition of having respect for the moral law. In the uses starting on page 82 it involves the doctrine of the Trinity—one God, three persons. Kant's uses of *Persönlichkeit* on page 71 clearly concern personal identity, and are translated accordingly.

**Pfaffentum:** The nearest English is 'priesthood' but that doesn't capture the derogatory tone of it, which Kant explains on page 97. The corresponding down-putting word for priests is **Pfaffen**.

**principle:** Kant often uses *Princip* in a sense, once common but now obsolete, in which it means 'source', 'cause', 'driver', 'energizer', or the like. The same was true of the French

*principe*, the Latin *principia*, and the English 'principle'. On page 45 the phrase 'a realm in which the power is held by principles' seems to be using the word in both senses at once. And on page 72 (the last of the **how** items) Kant is clearly talking about a 'principle' as a cause or driver and yet, oddly, the word he uses is not *Princip* but *Grundsatz* = 'basic proposition', which is hardly ever used in that way.

**rational:** This translates Kant's *rational*, an adjective that occurs only four times in the whole work, once on page 1 and three times on page 65.

**regulative:** A regulative principle, for Kant, is a principle that serves as advice or recommendation or even command, but not as giving any information. (Cf. 'constitutive', above.)

**science:** The use of this to translate *Wissenschaft* is practically unavoidable, but it has to be taken broadly as covering all the learned disciplines, so that (e.g.) history and theology are 'sciences'.

**statutory:** A statutory law is one that comes from someone's *choosing* to make it a law. The idea on page 56 of God's laws as being 'merely statutory' is the idea of their being laws only because God has decreed them.

**subtle reasoning:** This weakly 'translates' the various cognates of the verb *vernünfteln*, a splendid off-shoot of the noun *Vernunft* = 'reason', meaning: to employ a parade of super-subtle possibly invalid reasoning, weaving webs, splitting hairs, and so on. Neither this nor the corresponding noun *Vernünftelei* has a compact English equivalent.

**thaumaturgy:** 'The performance of miracles or wonders; magic' (OED).

**theodicy:** Attempt to reconcile the existence of bad states of affairs with the goodness of God.

**vicarious:** Acting in place of someone else. A vicarious atonement for my sins is an act of atonement performed by someone other than myself. Kant's reference on page 42 to 'the vicarious ideal of the son of God' means the idea of the son of God as a stand-in for God. In this version the word

translates *stellvertretend* = 'place-taking'. The corresponding noun *Stellvertreter* is translated by 'proxy'.

**Weltwesen:** Literally 'world-being'; the ten occurrences of this word are left untranslated because the preparer of this version can't get a good sense of what Kant means by it.

## Preface to the First Edition

Morality is based on the conception of man [see Glossary] as a being who is free and who—just *because* he is free—binds himself through his reason to unconditional laws. So it has no need for

- the idea of some other being above him, for him to
- know what his duty is, or
- the idea of an incentive other than the law itself, for him to •do his duty.

Or if such a need occurs for a given man, that's his fault; and in that case the need can't be met by anything outside himself, because the deficiency of his morality can't be made up for by anything that doesn't come from himself and his freedom. —Thus morality itself has no need for religion

**what Kant says next:** whether objectively, as regards the will [*das Wollen*], or subjectively, as regards ability [*das Können*],

**what he is getting at:** whether as •telling us what we should aim at or as •motivating us to aim at it,

because its needs are entirely met by pure practical reason. Its laws set the standard that every other purpose has to satisfy, and there's no further standard that they have to

meet. What makes them binding is a sheerly **formal** feature of the maxims that are to be adopted in accordance with them, namely the feature of being *universal* laws. Morality has no need for anything **material** to direct our free choices, i.e. no need for any end or purpose, to tell us what our duty is or to get us to perform it.<sup>1</sup> When the question of duty comes up, morality can and should ignore all ends. Should I be truthful in my testimony in the witness box? Should I be faithful in returning to another man the property he has entrusted to me? There is no need for me to work out what my duty is by considering what end I can bring about by acting in either of those ways—ends don't come into it. Indeed, if when a man's avowal is lawfully demanded he looks around for some kind of *end*, that fact alone shows him to be worthless.

But although morality doesn't need a representation of an end that must precede and contribute to the determining of the will, it may well have a necessary relationship to such an end, not as a *basis* for moral maxims but as an inevitable *consequences* of maxims adopted in conformity

<sup>1</sup> Those who aren't satisfied with the merely formal notion of *conformity to law* as the basis for settling what is one's duty will admit that such a basis can't be provided by self-love directed to one's own comfort. Then what can they say is the basis? They have two options:

- (1) a rational [see Glossary] basis, one's own *perfection*,
- (2) an empirical basis, the *happiness* of others.

There are two ways they could understand 'perfection' in this context. **(1a)** They could understand it as referring to moral perfection (i.e. having a will that is unconditionally obedient to the law) ; but in that case they would be explaining in a circle. Or **(1b)** they could take it to refer to natural perfection, considered as something that can be improved, and so it can in many different ways, e.g. skill in the arts and sciences, taste, bodily agility, and so on. But these are good only conditionally, because they are good only when their use doesn't conflict with the moral law (the only thing that commands unconditionally); so the aim to have natural perfection can't be the principle [see Glossary] of the concepts of duty. And that also holds for **(2)** the aim of producing happiness for others. Before an action is directed to the happiness of others it must first be weighed in itself, according to the moral law; so the most we can get from the purpose of bringing happiness to other people is a conditional duty, which means that this purpose can't serve as the supreme principle of moral maxims.

with morality's laws. In the absence of any reference to an end, no determination of the man's will can take place, because such a determination has to be followed by some effect, and the representation of the effect must be capable of being accepted—not as •the basis for the determination of the will and as an end antecedently aimed at, but—as •an end conceived of as the result of the will's determination through the law. Without an end of this sort,. . . .a will can't be satisfied: it is told *how* to act but not what it is to act *toward*. So although morality doesn't need an end to determine what conduct is right,. . . .an end does arise out of morality. For reason can't be indifferent to the answer to the question 'What will result from this right conduct of ours?', an answer pointing to an end that may not be entirely within our reach but can at least guide our doings and allowings. Hence the end is no more than

an idea of an object that contains and **unites**:

- the formal condition of all the ends that we *ought to* have, and
- whatever is in harmony with duty in all the ends that we *do* have,

that is, contains and **unites**:

- duty
- happiness in proportion as one is obedient to duty,

i.e. the idea of a highest good in the world.

For the possibility of this we must postulate a higher, moral, most holy, and omnipotent Being, the only thing that can **unite** the two elements of this highest good. But this idea, viewed practically, is not an empty one. . . .

[Why does Kant say 'But this idea is not empty'? Just by calling this item an 'idea' [see Glossary] he is implying that it can't be empirically cashed out in any way, i.e. that nothing could possibly count as perceiving or meeting up with something corresponding to it, so it has no place in

our scientific or metaphysical theorising about what is the case in the world. *But* this idea does in a disciplined way make a difference to *how we behave*. . . .]

. . . . because it does meet our natural need to conceive of some sort of final end, one that can be justified by reason, for all our doings and allowings taken as a whole; if we didn't have that conception of the highest good, our need for it would be a hindrance to our moral resolve. . . . So it makes a moral difference whether men form for themselves the concept of a final purpose of all things; adhering to that concept won't add to the number of their duties, but it will provide them with a special reference-point for the unification of all purposes; and that's the only way for objective, practical reality to be given to the combination of •the purposiveness arising from freedom with •the purposiveness of nature—a combination that we can't possibly do without. Consider this case:

A man honours the moral law, and can't help asking himself: 'If it were up to me to create a world that I would belong to, and if I did this under the guidance of practical reason, what sort of world would I create?'

He would select precisely the world that the moral idea of the highest good brings with it, and also he would will that such a world should somehow come into existence, because the moral law demands the realisation of the highest good we can produce. He would will this even if he saw that in that world he might pay a heavy price in happiness because he might not be adequate to the demands of the 'highest good.' idea, demands that reason lays down as conditions for happiness. He would feel compelled by reason to make this judgment •impartially, as though it were coming from someone else, and yet •as his own. . . .

So morality leads inescapably to religion, through which it extends itself to the idea of a powerful moral lawgiver,

outside of mankind, whose aim in creating the world is bring about the final state of the world that men can and ought to aim at also.

START OF LONG FOOTNOTE

If the proposition *There is a God, so there is a highest good in the world* is to arise as a dogma from morality alone, it is a synthetic *a priori* proposition. It is synthetic because although it is accepted only as an aid to conduct and not as a statement of fact, it goes beyond the concept of duty that morality contains. . . ., so it can't be extracted from morality by analysis. But how can such a synthetic proposition be *a priori*? The general moral concept of

duty

is indeed identical with the concept of

agreement with the bare idea of a being who gives moral laws to all men;

and as far as that goes the proposition commanding this agreement would be analytic. But the assumption of the law-giver's existence goes beyond saying merely that such a thing is possible. I think I know the solution of this problem, but in this place I can only point to the solution, not set it out fully.

An end or purpose is always the object of an *inclination*, i.e. of a desire to possess a thing through one's action, just as the law (which commands conduct) is an object of *respect*. An objective purpose (i.e. the end that we ought to have) is what sheer reason tells us to have. The end that includes the necessary and sufficient conditions of all other ends is the *final end*. The subjective final end of *Weltwesen* [see Glossary] that have reason—i.e. the purpose that they actually have—is their own happiness. . . .; and all practical propositions based on this final end are synthetic and also empirical rather than *a priori*. But the proposition that

*everyone's final end or purpose ought to be the highest good that is possible in the world* is a synthetic *a priori* practical proposition (an objectively practical one given by pure reason). It is synthetic because it goes beyond the concept of duties in this world and adds an upshot of the duties that isn't contained in the moral laws and so can't be extracted from them by analysis. These laws command absolutely, no matter what the upshot is; indeed, when we are considering a particular action the moral laws tell us to give no thought to what the consequences will be; and in this way they make duty an object of the greatest respect without presenting any end or upshot as an incentive to us to do our duty. Respect for duty is all the incentive anyone needs if he (as he should) attends only to what pure reason commands in the law. What need does anyone have to know what consequences will be drawn from his doings and allowings by the course of events in the world? All he needs is to know that he does his duty, even if there is no life after this one and in this life those who are happy are not the same group as those who deserve to be happy. But it's one of the inescapable limitations of man and of his faculty of practical reason (and perhaps of all other *Weltwesen* as well) that in every action he performs he looks to its upshot, wanting to find in it something that could serve as a purpose for him and could also prove the purity of his intention; this upshot comes last in the sequence of events but first in his thought and intention. In this purpose, even if it is directly presented to him by bare reason, he looks for something he can love; and the law pays some attention to this search. The law itself merely arouses his respect and not his love; and doesn't acknowledge this sought-for object of love as something man needs; but the law extends itself so as to bring it [i.e. the sought-for object of love] in, by including among its reasons for action *the moral final purpose of reason*. That is, the



proposition:

Have as your final purpose the highest good that is possible in the world!

is a synthetic *a priori* proposition that is introduced through the moral law itself, though practical reason in doing this stretches out beyond the law. This extension is possible because the moral law is being taken in relation to a natural characteristic of man, namely that for all his actions he has to think not only about the law but also about a purpose ·or upshot·. [The next few lines of Kant's text are horrendously difficult; we can safely pass them by. This footnote then ends:] If the strictest obedience to moral laws is to be considered the cause of the ushering in of the highest good (as upshot), then, since humans can't bring about happiness in the world proportionate to worthiness to be happy, an omnipotent moral being must be postulated as ruler of the world, under whose care this proportion is achieved. That is, morality leads inevitably to religion.

END OF LONG FOOTNOTE

Just as morality recognises the holiness of its law as an object of the greatest *respect*, so at the level of religion it presents the ultimate cause that fulfills those laws as an object of *worship*—and thus morality appears in its majesty. But everything—even the most sublime thing—dwindles in the hands of men who are turning the idea of it to their own use. Something that can truly be venerated only when it is *freely* respected is forced to ·lose that freedom and· adapt itself to forms that are authoritative—meaning that they are backed up with coercive laws; and something which if left to itself exposes itself to the public criticism of everyone has to submit to a criticism that has power, i.e. to a censorship.

But the command *Obey the authority!* is also moral, and obedience to it—as to all duty-commands—can be extended

to religion, so it's fitting that a treatise devoted to the determinate concept of religion should itself present an example of this obedience. It will, however, be obedience that is based not on

attention merely to the law governing *one* way things are ordered in the state while ignoring *all the others*, but rather on

a combined respect for all of them taken together.

Now the theologian who passes judgment on books may be appointed either as

**(a)** a cleric, who is to care only for the soul's welfare or as

**(b)** a scholar, who is to care also for the welfare of the sciences [see Glossary].

[In what follows, 'working part' translates *Glied* = 'limb', or 'member' in the sense in which arms and legs are members.] The **(b)** scholar is a working part of a public institution (called a 'university') that is charged with developing all the sciences and defending them against intrusions from the outside; so it's up to him to ensure that the pretensions of **(a)** the cleric are kept within bounds, so that his censorship doesn't harm the sciences. And if both of them are Biblical theologians, the **(b)** scholar should have the upper hand, as a working part of the university and as belonging to the department whose job it is to deal with theology. They both have the role of caring for souls, but **(b)** the theologian in role as university scholar also has a special function to perform in regard to the welfare of the sciences. If this rule isn't maintained, things are bound to end up in the state they were in at (for example) the time of Galileo. The Biblical theologian, wanting to humble the pride of the sciences without doing any actual *work* in this connection, might venture an invasion into astronomy or some other science (e.g. the ancient history of the earth) and confiscate and cancel all the endeavours of

human reason—like tribes who, finding that they don't have the means or the resolution needed to defend themselves against threatened attacks, adopt a 'scorched-earth' strategy in which they transform all about them into a wilderness.

In the territory of the sciences, biblical theology is a neighbour of philosophical theology, a domain that has been entrusted to another department. This must have complete freedom to expand as far as its science reaches, provided that it stays within the limits of bare reason alone. It is entitled to bring in history, sayings, books of all peoples, even the Bible, but only for confirming and expounding its own propositions, not aiming to carry these propositions into Biblical theology or to usurp the cleric's privilege of changing the latter's public doctrines. If it is shown that the philosophical theologian has really overstepped his limits and trespassed on biblical theology, the biblical theologian (in his role as a cleric) has an indisputable right of censorship. . . . He has this *as a working part of his department* which has been assigned to care for the community's second interest, namely, the prosperity of the sciences, an assignment that is just as valid as the first, namely the care of souls.

In such a case, it is this department of biblical theology that is the authoritative censor, not the department of philosophical theology. Why? Because the former department has a legal right to certain doctrines, whereas the latter department doesn't: in its domain doctrines freely come and go. So only the former—the department of biblical theology—can formally complain that its exclusive rights have been violated. The two bodies of doctrine are close to one another, and it may be feared that the philosophical department will cross the boundary between them, but there's no need for anxiety about this: you just have to bear in mind that there's nothing wrong with the philosopher's **borrowing something from** biblical theology to use for his

own purposes—even if he gives the borrowed material a meaning that suits bare reason but doesn't please the biblical theologian! Biblical theology won't want to deny that it has much in common with the teachings of bare reason, as it does also with historical and philological lore, making it subject to the censorship of these disciplines. There's something wrong in the philosophical theologian's conduct only if he **carries something into** biblical theology, trying to push it in directions that it isn't built for. (Similarly, a professor of natural-rights isn't trespassing when he uses in his philosophical doctrine of rights many classical terms and formulae borrowed from the Roman codex, even if. . . .he doesn't use them in exactly the sense that they originally had (according to scholars of Roman Law). He's not open to criticism for these borrowings unless he tries to get practising lawyers and even judges to understand this material in his way.) If the philosophical theologian weren't entitled to such borrowings, we could turn the thing around and accuse the biblical theologian or the legal theorist of repeatedly trespassing on philosophy's territory, because both of those often have to borrow from philosophy. . . . If the biblical theologian decided to have, if possible, nothing to do with reason in religious matters, it's easy to see which side would lose from this; a religion that rashly declares war on reason won't be able to hold out against it for long.

I will even venture to suggest that it might be beneficial to complete each student's education in biblical theology with a final course of lectures on the purely philosophical theory of religion (which avails itself of everything, including the Bible). The text for the lectures could be this book, or a better one of the same kind if such can be found. For the sciences—I mean these two theological sciences—get pure benefit from separation, so far as each first constitutes a whole by itself; it's only when they have been so constituted

that we should try to survey them in combination. Let the biblical theologian be at one with the philosopher; or let him think he should refute him, but only if he *hears* him! It is only by listening that the biblical theologian can be forearmed against all the difficulties the philosopher might make for him. To conceal these, or to brush them aside as ungodly, is a paltry device that doesn't stand the test; while to mix the two sciences, with the biblical theologian merely glancing across to philosophy occasionally, is a lack of thoroughness that will end up with no-one's really knowing how he stands towards the doctrine of religion as a whole.

To show how religion relates to human nature (with its good predispositions and its bad ones) I shall in the four following essays represent the good and bad principles [see Glossary] as a pair of independent active causes influencing men. The first essay has already appeared in a learned journal, but it had to be included here because of how tightly the materials of all four hang together: the three essays I am now adding contain a complete development of the project. . . .

### **Preface to the Second Edition**

In this edition misprints are corrected and in a few places the wording has been improved; those are the only alterations. Some new material is added; it will occur in footnotes and other additions that start with a dagger (†).

Some readers have expressed concerns about this work, wanting to know what I am *up to* in my choice of title for it. We have to distinguish **(a)** revelation from **(b)** pure religion of reason [which aligns with the distinction between **(a)** biblical theology and **(b)** philosophical theology]. Now, **(a)** can include **(b)**, because a biblical revelation could include, say, a divine command to disregard certain philosophical arguments for God's exist-

tence; whereas **(b)** can't include any of the historical content of **(a)**. So I'll be able to regard **(a)** as the wider sphere of faith which includes within itself **(b)** the narrower one—like two concentric circles). The philosopher, as a teacher of pure reason (working only with *a priori* principles), must stay within the smaller circle and set aside anything learned from experience. From this standpoint I can also run a second **experiment**. After setting aside the pure religion of reason considered as a self-sufficient system, I can take some alleged revelation and conduct a piecemeal investigation of how it checks out against moral concepts, and then see whether it leads back to the pure religion of reason. The latter may be self-sufficient and adequate for the parts of genuine religion that concern the morality of conduct; and those parts are really the whole, because genuine religion is an *a priori* concept of reason with no empirical content, so that it exists only in this moral domain. [Kant builds into that sentence a contrast between 'the morality of conduct' and materials that concern non-moral theories about what is the case, including ones about how best to go about teaching.] If this **experiment** succeeds, we'll be able to say that reason is not only compatible with Scripture but unified with it, so that if (guided by moral concepts) you follow one you'll also conform to the other. If this weren't so, we would have either two religions in one person, which is absurd, or one *religion* and one *cult*. Because a cult is not an end in itself (as religion is) but only valuable as a means, the two would often have to be shaken up together to get them to combine for a while; though each time they would then separate from one another, like oil and water, with the purely moral one (the religion of reason) floating on top.

I noted in the first Preface that this unification, or the attempt at it, is something the philosophical investigator of religion is entitled to do, and doesn't encroach on the

exclusive rights of the biblical theologian. Since then I have found this assertion •cited in *Morality* by the late J. D. Michaelis, a man well versed in both fields, and •applied throughout that entire work; and the higher department [here = the department of biblical theology] didn't find in it anything prejudicial to its rights.

In this second edition I would have liked to respond to what has been said about this book by worthy men, named and unnamed; but I haven't been able to because this material (like all literary material from abroad) has been so slow to arrive in our parts. This is particularly true of the *Annotationes quaedam theologicae* etc.—more fully: 'Some theological remarks concerning Kant's philosophical doctrine of religion'—by the renowned Dr Storr in Tübingen, who has examined my book with his accustomed sharpness and with a diligence and fairness deserving the greatest thanks. I do plan to answer him, but can't *promise* to do so because of the peculiar difficulties that old age sets in the way of working with abstract ideas. [Kant was 70 when he wrote this.] But a review in *Latest Critical News* can be dealt with as briefly as the reviewer did the book itself. For the book, in this reviewer's judgment, is nothing but my answer to the question I asked myself: 'How are the concepts and

doctrines of the ecclesiastical system of dogmatic theology possible according to pure (theoretical and practical) reason?' This essay, ·he claims·, has nothing to say to those for whom Kant's system is non-existent—i.e. those who don't know and understand the system and haven't the least desire to do so. I answer thus: To understand the essential content of this work, all you need is common morality; there's no need to bring in the *Critique of ·Pure· Practical Reason*, still less the *Critique of Pure ·Theoretical· Reason*. For example, when virtue as skill in actions conforming to duty (according to their **legality**) is called '*phenomenal* virtue', and the same virtue as an enduring attitude towards such actions from duty (because of their **morality**) is called '*noumenal* virtue', these terms are used only in deference to the schools [here = 'to one group of academic philosophers'], but the distinction itself is contained, though in other words, in the most ordinary everyday children's instruction and in sermons, and is easy to understand. If only the same could be said for the mysteries of the divine nature that are included among religious teachings! They're introduced into the catechism as though they were perfectly ordinary and everyday, but they won't become comprehensible to everyone unless they are first transformed into moral concepts.

## First Essay: The bad principle existing alongside the good i.e. The radical evil in human nature

[On the word 'principle' as used here and in the titles of the other three Essays, see the Glossary.]

The complaint that 'the world lies in evil' is older than history, indeed as old as that oldest of all poetic endeavours, the religion of the priests. All religions agree that the world began in a good state, whether in a Golden Age, a life in Eden, or an even happier communion with celestial beings. But they soon let this happiness vanish like a dream and give place to a fall into evil (moral evil, always going hand in hand with physical evil), speeding mankind from bad to worse with accelerated descent; so that now (but this Now is as old as history) we live in the final age, with the Last Day and the destruction of the world knocking at the door. In some parts of India the Judge and Destroyer of the world, Rudra (sometimes called Siwa or Siva), is already being worshipped as the reigning God, because Vishnu, the Sustainer of the world became weary of his task some centuries ago and renounced the supreme authority he had inherited from Brahma, the Creator of the world.

More recent though far less prevalent is the opposite optimistic belief that the world is steadily (though almost imperceptibly) moving from bad to better, or at least that the predisposition to move in that way can be found in human nature. It's only philosophers who have held this view, and these days especially the teachers of philosophy. If this is a thesis about movement along the scale from moral badness to moral goodness (not simply about the process of civilisation), it certainly hasn't been derived from experience—the history of all times speaks too loudly against it! Presumably it is merely a well-meaning postulate of moralists from Seneca to Rousseau designed to encourage

us to cultivate the seed of goodness that lies in us—if there is one. Their thought is that since we take it for granted that man is naturally sound of body (as at birth he usually is), there's no reason why we shouldn't assume that his soul is also healthy and free from evil; so nature itself is inclined to help us on developing this moral predisposition to goodness. [Kant adds a quotation to that effect from Seneca.]

But it may be that both sides are wrong about this. Isn't it at least possible that the truth lies between them: man as a species is neither good nor bad, or any way as much the one as the other—partly good, partly bad. We call a man bad, however, not because his **actions** are bad (contrary to law) but because his actions show that he has bad **maxims** in him. Through experience we can observe actions that are contrary to law, and we can observe (at least in ourselves) that they're performed with awareness that they are unlawful; but a man's maxims aren't observable in this way (even by himself in many cases); so experience can't support a confident judgment that a given man is bad. To be entitled to call a man bad, you would have to be able to infer *a priori* from several consciously bad acts—or from just one—an underlying evil maxim; and further from this maxim to infer the presence in the man of an underlying general basis for all his particular morally evil maxims, a basis that is itself another maxim.

You may have trouble with the word 'nature' which is used in the title of this Essay. When an action is said to arise from nature, that usually means that the action is not free, which implies that it isn't either *morally* good or

*morally* evil. But I'm not using 'nature' in that way. What I call 'human nature' is the with the subjective basis of the exercise (under objective moral laws) of man's freedom, a basis which—wherever it lies—precedes every action that is apparent to the senses. But this subjective basis must also be an expression of freedom, because otherwise the resultant action. . . . couldn't be *morally* good or bad. So the basis of evil can't lie in •anything that determines [see Glossary] the will through inclination, or in •any natural impulse; it can lie only in a •rule that the will makes for itself, as something on which to exercise its freedom—i.e. a •maxim. We mustn't ask what the man's subjective ground is for adopting this maxim rather than of its opposite. If it were not ultimately a maxim but a mere natural impulse, the man's 'free' action could be tracked back to determination by natural causes, which contradicts the very notion of freedom. So when we say 'He is by nature good' or '. . . . bad', this means only:

There is in him a rock-bottom basis (inscrutable to us) for the adoption of good maxims or of bad ones (i.e. maxims contrary to law); and he has this just because he is a man, so his having it expresses the character of his species.<sup>1</sup>

So we shall characterise as *innate* the ·good or bad· character that distinguishes man from other possible beings that have reason; but that won't prevent us from maintaining that nature is not •to bear the blame if the character is bad or •to take the credit if it is good, and that man himself is its author. ·To see how this can be so, you have to grasp how 'innate'

is being used here. The rock-bottom basis for the adoption of our maxims must itself lie in free choice, so it can't be something we meet with in experience; therefore, the good or evil in man (as the ultimate subjective basis for the adoption of this or that maxim relating to the moral law) is termed 'innate' only in the sense of being posited as the basis for—and thus being earlier than—*every* use of freedom in experience (including ones in earliest youth, as far back as birth); so it is conceived of as present in man at the time of birth—though birth needn't be its cause.

### Comment

The conflict between the two hypotheses presented above is based on a disjunctive proposition: *Man is (by nature) either morally good or morally bad*. Is this disjunction valid? Mightn't it be that man is by nature neither good nor bad? or that he is both at once, good in some respects and bad in others? Experience actually seems to confirm the middle ground between the two extremes.

But it matters greatly to ethics to hold off as long as possible from anything morally intermediate, whether in actions or in human characters. That is because such ambiguity threatens all maxims with becoming vague and unstable. Those who favour this strict way of thinking are usually called *rigorists* (a name that is intended to carry reproach but actually praises); their opposites could be called *latitudinarians*. These divide into •latitudinarians of neutrality, whom we can call 'indifferentists' and •latitudinarians of coalition, whom we can call 'syncretists'.<sup>2</sup>

<sup>1</sup> That the ultimate basis for the adoption of moral maxims in inscrutable can be seen. . . .from the following. This adoption must itself be free; so the basis for it—the explanation for its favouring (e.g.) a bad maxim rather than a good one—can't come from any natural drive and must involve yet another maxim; this in turn must have a basis. . . . and so we are launched on an infinite series of ever earlier bases for choices.

<sup>2</sup> [Kant has here a footnote arguing that because **(a)** the moral law is a motivating force in us, **(b)** there is no middle position between going with the law and going against the law. The details of the argument are obscure.]

·What follows will involve the notion of an action as being ‘morally indifferent’. Before getting into that, I have some things to say about that concept·.

·PASSAGE ADDED IN SECOND EDITION·

† [This passage is directed towards the great poet, dramatist and critic Friedrich Schiller, who was an unpaid professor at the University of Jena.] A morally indifferent action would be one resulting merely from natural laws, and hence standing in no relation whatsoever to the moral law, which is the law of freedom;. . .with regard to such an action there is no place for, and no need for, command or prohibition or permission.

In his masterly treatise on grace and dignity in morality (published in the journal *Thalia*), Professor Schiller objects to this way of representing obligation, as carrying with it a monastic cast of mind. But he and I agree on the most important ·other· principles, so I have to think that we don’t disagree about this one either, if only we can make ourselves clear to one another. I freely grant that I can’t associate the concept of duty with *grace*, precisely because of the concept’s dignity. The concept involves absolute necessitation, and *grace* stands in direct contradiction to that. The majesty of the moral law (as of the law on Sinai) instils awe (not dread, which repels; and not charm, which invites familiarity); this awe arouses the subordinate’s *respect* for his master; and in this case, where the master resides within us, this respect awakens a sense of how sublime of our own destiny is, which enraptures us more than any beauty. But virtue—i.e. the firmly based disposition strictly to do our duty—has results that are more beneficent than anything nature or art can accomplish in the world; and the splendid picture of human virtue does allow the Graces to enter the picture, though they keep a respectful distance when is the sole topic. [Kant is using the (three) Graces—Roman goddesses of charm, beauty etc.—as

a metaphor for the gracefulness that Schiller was writing about.] . . .

What is the aesthetic character, the temperament, as it were, of virtue? Is it courageous and hence joyous or fear-ridden and dejected? An answer is hardly necessary. The slavish dejected frame of mind can’t occur without a hidden hatred of the law. Whereas a heart that is happy in the performance of its duty. . .is a mark of the genuineness of a virtuous disposition. And of the genuineness of piety, which does not consist in the self-inflicted torment of a repentant sinner (a very ambiguous state of mind, which ordinarily is nothing but regret at having infringed the rule of prudence), but rather in a firm resolve to do better in the future. This resolve, then, encouraged by good progress, must create a joyous frame of mind, without which man is never certain of having •really achieved a love for the good, i.e. of having •incorporated it into his maxim.

END OF THE ADDED PASSAGE

·We confronted the question ‘How do good/bad figure in human nature? Is it indifferent between them? or a bit of one and a bit of the other? or . . .?’· According to the rigoristic diagnosis, the answer to this is based on an observation that is highly important to morality, namely:

Freedom of the will is utterly unlike anything else in that no incentive can determine the will to an action unless the man has incorporated that incentive into his maxim, making it [= this determination] the general rule that he wills to conduct himself by. Those are the only terms on which any incentive can co-exist with the will’s absolute spontaneity, i.e. its freedom.

Now, reason judges that the moral law is in itself an incentive, and anyone who makes it his maxim is morally good. If someone performs an action to which the moral law is relevant and his will was not •determined by this law, then

it must have been •influenced by an incentive contrary to it; and since this can happen only when a man admits this incentive (and thus his deviation from the moral law) into his maxim (in which case he is a bad man), it follows that his disposition in respect to the moral law is never indifferent, never ‘neither good nor bad’.

Nor can a man be morally good in some ways while morally bad in others. His being •good in one way means that he has incorporated the moral law into his maxim; if he were at the same time •bad in another way, this would involve his having a maxim that creates *exceptions* to his *universal* maxim about obedience to duty; and that’s a contradiction.<sup>1</sup>

When I speak of one or other disposition as ‘inborn’ and ‘natural’ I don’t mean that it hasn’t been acquired by the man whose constitution it is, or that he didn’t create it; all I mean is that this didn’t happen over time—that he has always been good or bad from his youth onwards. But this disposition—the ultimate subjective basis for the adoption of maxims—. . . must have been adopted by •the man’s• free choice, because otherwise he couldn’t be subject to praise or blame for it. But the subjective basis or cause of this adoption can’t be known (though it’s inevitable that we ask about it), because knowing it would involve bringing in another maxim, which would in its turn have a basis. . . . and so on •backwards to infinity•. Since we can’t explain this disposition, or rather its ultimate basis, in terms of any fundamental act of the will *in time*, we call it a property of the will that belongs to it by nature (although actually the

disposition is based on freedom). However, this proposition—  
When we say of someone ‘He is by nature good (or bad)’ we have to be saying this about the whole species, for if we could say it about the individual man then one man could be considered as good by nature, another as bad

—can’t be proved until and unless anthropological research shows that the evidence that justifies us in saying of someone ‘He is innately good (bad)’ is such as to provide no basis for excepting anyone, and that our attribution therefore holds for the species.

### 1. The Original Predisposition to Good in Human Nature

This predisposition can conveniently be divided into three elements involving dispositions toward three different goals. These can be considered as elements in the structure of mankind.

- (1) As a living being man has a predisposition to animality;
- (2) As a living and reason-possessing being man has a predisposition to humanity;
- (3) As a reason-possessing and morally accountable being man has a predisposition to personality [see Glossary].

<sup>1</sup> The ancient moral philosophers, who said just about all there is to say about virtue, addressed our two questions. They expressed the first of them thus: *Must virtue be learned? Is man naturally indifferent as regards virtue and vice?* And they put the second thus: *Is there more than one virtue, so that man might be virtuous in some respects and vicious in others?* They answered both with rigoristic definiteness in the negative, and rightly so; for they were considering virtue *in itself*, as it is in the idea of reason (what man ought to be). But if we want to pass moral judgment on this moral being, man *as he appears*, i.e. as experience reveals him to us, we can answer both questions in the affirmative; for in this case we judge him not by the standard of pure reason before a divine tribunal but by an empirical standard before a human judge. I’ll say more about this later.



·START OF FOOTNOTE·

We can't regard **(3)** as included in the concept of **(2)**; it has to be regarded as a special predisposition on its own. From the fact that a being has reason it doesn't follow—as far as I can see—that this reason, simply by having the thought that its maxims are fit to be laid down as universal laws, can determine the will unconditionally and thus be self-sufficiently 'practical' [i.e. be able to get itself moving without any input from outside]. A **(2)** *Weltwesen* [see Glossary] extremely well equipped with reason might need certain incentives, originating in objects of desire, to determine his choice. He could

use the full force of his reason to decide which set of incentives adds up to the strongest, and to work out how to achieve the states of affairs that they aim at, without *suspecting the possibility* of **(3)** the absolutely imperative moral law which proclaims that it is itself an incentive, and indeed the highest incentive. If his law weren't given to us from within, we would never have been able by high-level reasoning to bring it into existence or subject our will to it; yet this law is the only thing that tells us that our will isn't under the control of other incentives (tells us of our freedom) and at the same time that we are morally accountable for all our actions.

·END OF FOOTNOTE·

**(1)** Man's predisposition to animality can be brought under the general heading of 'physical and purely mechanical self-love', for which reason isn't needed. It is threefold: •for self-preservation; •for the propagation of the species through sexual intercourse and the care of offspring arising from that; and •for community with other men, i.e. the social impulse. On these stems all kinds of vices can be grafted (but they don't spring from **(1)** this predisposition itself as a root).

They can be called vices of the coarseness of nature, and their extreme cases are called the 'bestial vices' of •gluttony and drunkenness, •lasciviousness and •wild lawlessness (in relation to other men).

**(2)** The predisposition to humanity can be brought under the general title of 'self-love' that is physical and yet makes comparisons (for which reason is required)—we judge ourselves happy or unhappy only by comparing ourselves with others. This self-love creates the inclination to become worthy in the opinion of others. This starts as a desire merely for equality, to allow no-one to rise above oneself, combined with a constant anxiety about whether others are trying to do just that; from which there gradually arises the unjustifiable craving to achieve superiority for oneself over others. Great vices can be grafted onto this twin stem of •jealousy and •rivalry, namely the vices of secret and open hostility towards everyone we see as alien to us. These vices don't sprout from nature as their root. They are merely inclinations aroused in us •to defend ourselves against the attempts of others to get superiority over us, •to get the upper hand as a preventive measure. [In the next sentence, 'culture' (*Kultur*) could refer to •gardening or to •literature, music etc. Perhaps Kant punningly means both.] That's not what Nature wanted.; it wanted to use the idea of such competitiveness (which in itself does not exclude mutual love) only as a spur to culture. So the vices that are grafted onto this inclination could be termed 'vices of culture'; in the highest degree of malignancy—e.g. in envy, ingratitude, *Schadenfreude* etc.—where they are simply the idea of a maximum of evil going beyond what is human, they are called 'diabolical vices'.

**(3)** The predisposition to personality is the ability to have respect for the moral law as an incentive that can unaided move the will. A capacity for mere *respect* for the moral law within us would be moral *feeling*, which is a goal of the

natural predisposition not *in itself* but only as something that moves the will. Since this is possible only when the free will incorporates such moral feeling into its maxim, the constitution of such a will is good character. This...is something that can only be *acquired*; but that couldn't happen unless our nature already included a predisposition to it, a predisposition onto which nothing bad can possibly be grafted. We can't rightly call the idea of the moral law, with the respect that is inseparable from it, a *predisposition to personality*; it is personality (the idea of humanity considered quite intellectually). But the subjective basis for the adoption of this respect into our maxims as a motivating force seems to be something additional to personality, and thus to deserve to be called a predisposition to it.

When we look at the requirements for the three predispositions I have listed, we find that **(1)** isn't based on reason, **(2)** is based on practical reason but only in the service of other incentives, while **(3)** is based on reason that is practical in itself, i.e. reason that dictates laws unconditionally. All these predispositions are not only good in negative fashion (they don't contradict the moral law), but also predispositions toward good (they further the observance of the law). They are original [here = 'basic', 'not derivative'], because they are bound up with the possibility of human nature. Man can indeed use **(1)** or **(2)** contrary to their ends, but he can't extinguish any of the three. By a being's 'predispositions' I mean both •its constituent elements that are necessary to it and also •the way they are put together to make it the

being that it is. They are •original if they are necessarily involved in the possibility of such a being, but •contingent if the being could exist without them. Notice that I am here treating only predispositions that are directly related to the faculty of desire and the exercise of the will.

## 2. The Propensity to Evil in Human Nature

By 'propensity' I understand the subjective ground of the possibility of an inclination (habitual desire, craving) which mankind in general is liable to have.<sup>1</sup> What distinguishes a propensity from a predisposition is the fact that although it can be innate it doesn't have to be seen in that way; it can also be regarded as having been acquired (if it is good) or inflicted on the man by himself (if it is bad). But my topic here is only the propensity to what is genuinely bad, i.e. morally bad; for since such evil [see Glossary] is possible only as a state of the free will, and since the will can be judged as good or bad only by means of its maxims, this

- propensity to evil must consist in
- the subjective basis for the possibility of the maxims' deviating from the moral law.

If this propensity can be seen as belonging to mankind in general (and thus as being part of the character of the species), it can be called man's 'natural propensity to evil'. . . .

We can distinguish three different levels in this capacity for evil, i.e. in man's natural propensity to evil: **(1)** the

<sup>1</sup> † A propensity is really only a predisposition to want an enjoyment which, once it has been experienced, arouses in the subject an *inclination* to it. Thus all uncultured people have a propensity for intoxicants; many of them know nothing of intoxication and therefore have no desire for intoxicants, but once they have sampled one there is aroused in them an almost inextinguishable desire for it. Between •inclination, which presupposes acquaintance with the object of desire, and •propensity there •is instinct, which is a felt want to do or to enjoy something that one doesn't yet have any conception of (such as the sexual impulse, or the impulse in beavers to build dams). Beyond inclination there is finally a further stage in the faculty of desire, namely •passion, . . . which is an inclination that excludes the mastery over oneself.

weakness of the human heart in the general observance of adopted maxims, i.e. the **frailty** of human nature; **(2)** the propensity for mixing immoral with moral incentives (·which is bad· even when it is done with good intent and under maxims of the good), i.e. **impurity**; **(3)** the propensity to adopt bad maxims, i.e. the **wickedness** of human nature or of the human heart.

**(1)** The frailty of human nature is expressed even in the lament of the Apostle Paul: ‘What I will to do I don’t do!’ [Romans 7:18] In other words, I take up the good (the law) into the maxim of my will, but this good, which objectively in its ideal conception is an irresistible incentive, is subjectively—when it comes to actually following the maxim—weaker than ·the contrary· inclination).

**(2)** The impurity of the human heart consists in this: although the maxim is indeed good in respect of its object (the intended observance of the law) and may even be strong enough to lead to action, it isn’t purely moral; i.e. it hasn’t adopted *the law alone* as its all-sufficient incentive, and instead—usually or perhaps always—needs other incentives beyond this to get the will to do what duty demands. In short, actions called for by duty are done not purely for duty’s sake.

**(3)** The wickedness—or, if you like, the corruption—of the human heart is the propensity of the will to ·act on· maxims in which the incentives springing from the moral law are upstaged by others that aren’t moral. It can also be called the ‘perversity’ [Verkehrtheit] of the human heart, because it reverses [umkehrt] the moral order among the incentives of a free will; and although conduct that conforms to the law can occur in its presence, the cast of mind is corrupted at its root (so far as the moral disposition is concerned), so the man is described as ‘bad’.

You’ll notice that this propensity to evil is here attributed (as regards conduct) to men in general, even to the best

of them; this must be done if it’s to be proved that the propensity to evil is universal in mankind, i.e. that it is woven into human nature.

As regards conformity of conduct to the moral law there need be no difference between **(a)** a man of good morals and **(b)** a morally good man, except that **(a)**’s conduct doesn’t always—perhaps doesn’t ever—have the law as its sole and supreme incentive, while **(b)**’s conduct always does.

**(a)** He obeys the law according to the *letter* (i.e. his conduct conforms to what the law commands).

**(b)** He obeys the law according to the *spirit* (the spirit of the moral law consisting in this, that the law is sufficient in itself as an incentive).

... When incentives other than the law itself (e.g. ambition, self-love in general, even a kindly instinct such as sympathy) are needed to get the will to ·pursue· lawful actions, it is merely accidental that these actions conform to the law, for those incentives could just as well have led to its violation. So the man’s maxim, the goodness of which shows his moral worth, is contrary to the law; and **(a)** the man, despite all his good deeds, is nevertheless bad.

To pin down the concept of this propensity I need to explain something. Every propensity is either **(i)** physical, i.e. belonging to the will of the man as a natural being, or **(ii)** moral, i.e. belonging to the will of the man as a moral being. In **(i)** there is no propensity to moral evil, for such a propensity must spring from freedom; and a physical propensity (based on sensuous impulses) towards *any* use of freedom—good or bad—is a contradiction. Hence a propensity to evil can adhere only to the moral capacity of the will. But the only things that are morally bad (i.e. are things we can be held accountable for) are our own actions. On the other hand, the concept of a propensity is taken to apply to a subjective determining basis of the will

that *precedes all actions* and is therefore not itself an action. Hence the concept of *simple propensity to evil* would be self-contradictory if it weren't possible to take this expression [= 'action'] in two meanings, both of which can be reconciled with the concept of freedom. The word 'action' can apply to either of two uses of freedom: •that in which the supreme maxim (in conformity with the law or contrary to it) is adopted by the will, and •that in which the actions themselves (considered materially, i.e. in terms of what the man is trying to do) are performed in accordance with that maxim. The propensity to evil is both **(1)** an 'action' in the first sense, and at the same time the formal basis of any **(2)** unlawful 'action' in the second sense. . . . It can happen that

- a **(1)** bad action is performed, and yet
- every time a **(2)** bad action might result it is headed off by some incentive that doesn't involve the law.

In that case, the guilt for the **(1)** bad action remains. It's an intelligible action, knowable through bare reason, and not known as happening at some particular time or through any particular period; the **(2)** action is perceivable through the senses, empirical, given in time. The **(1)** action, particularly when compared with the **(2)** action, is called a bare propensity and *innate*. There are two reasons for this. It is because •the propensity can't be eradicated (because that would have to be done by the highest maxim which would have to be that of the good—whereas this propensity has already adopted something bad as the highest maxim); and especially because •although the corruption by evil of this highest maxim is our own action, we can't assign a further cause for it, any more than we can assign a cause for any basic feature of our nature (•e.g. a cause for our having reason•).

From what I have just said, you can see why in this section right from the outset I looked for the three sources of the morally evil solely in matters having to do with the

supreme basis for the adoption or or obedience to our maxims, and not in anything involving the senses. . . .

### 3. Man is bad by Nature

According to what I have said, the proposition *Man is bad* can only mean that *he is conscious of the moral law but has nevertheless allowed occasional departures from it into his maxim*. 'He is bad *by nature*' means that badness can be predicated of man as a species. It doesn't say that this quality can be inferred from the concept of *man* as such, for that would make it a necessary truth. All it means is that from what we know of man through experience we can't judge him otherwise, or that we can take it that evil is subjectively necessary to every man—i.e. built into each individual man as a separate fact about *him*—even to the best. Now this propensity must itself be considered as morally bad, so not as a natural predisposition but rather as something the man can be held accountable for; and consequently it must consist in unlawful maxims of the will. And because of freedom—i.e. because we are free to obey the maxims or disobey them—these maxims must be regarded as contingent; but that doesn't square with the universality of this evil unless the rock-bottom basis of all maxims is, somehow or other, entwined with and rooted in humanity itself. [Kant is openly declining to say *how* evil is rooted in humanity. He has already said that the rooting is not conceptual.] So we can call this a *natural propensity to evil*, and because we must always accept the guilt for it we can call it a *radical innate evil in human nature*, though one we have brought upon ourselves.

That such a corrupt propensity must indeed be rooted in men needn't be formally *proved*, given the multitude of glaring examples we see by observing men's actions. Some philosophers have hoped to encounter humanity's natural

goodness in the so-called *state of nature*, but just look at it! Look at the scenes of unprovoked cruelty in the murder-dramas enacted in Tofoa, in New Zealand, and in the Navigator Islands, and the ceaseless cruelty that is reported to happen in the wide wastes of northwestern America—cruelty from which no-one gets the smallest benefit.<sup>1</sup> And we have vices of barbarity of our own that are more than sufficient to draw us from the opinion of man's goodness in the state of nature. Perhaps you are drawn to the opinion that human nature can better be known in the civilised state (in which its predispositions can develop more completely), then listen to the long melancholy litany of indictments against humanity:

- secret falsity even in the closest friendship, so that even among the best friends it is always thought prudent to limit one's trust in others;
- a propensity to hate someone to whom one is indebted—something that benefactors must always be prepared for;
- heartfelt well-wishing that doesn't falsify the remark of La Rochefoucauld that 'in the misfortunes of our best friends there is something that is not altogether displeasing to us';

and many other vices concealed under the appearance of virtue, not to mention the vices of those who don't conceal them because we're content to regard as *good* a man who is *bad in a way that everyone is bad*. And we'll have enough vices of culture and civilisation (which are the worst of all) to make prefer to stop looking at the conduct of men lest we ourselves contract another vice, misanthropy.

If you are still not convinced that man is disposed to evil both in the state of nature and in the civilised state, consider the state that is oddly composed of both the others, i.e. the international situation, where civilised nations relate to each other in the way individuals do in the primitive state of nature (a state of continuous readiness for war), and are firmly resolved never to relate differently. That will make you aware of the fundamental principles of the great societies called *states* [see below], principles that flatly contradict their public pronouncements but can't be laid aside, and that no philosopher has yet been able to bring into agreement with morality.

•ADDED PASSAGE COMMENTING ON 'STATES'.

† When we look at the history of these merely as the visible upshot of the inner predispositions of mankind that are mostly concealed from us, we become aware of a certain mechanical movement of nature toward ends that are nature's own rather than those of the peoples. •Each separate state, so long as it has a neighbour that it hopes to conquer, works to enlarge itself at the expense of the neighbour, thus taking a step towards world-monarchy, a political system in which all freedom disappears, along with its consequences—virtue, taste, and learning. •But this monster (in which laws gradually lose their force), after swallowing all its neighbours eventually breaks up—through rebellion and disunion—into many smaller states. •These, instead of working for a union of states (a republic of federated free peoples), begin the same game over again, each for itself, making sure that war, that

<sup>1</sup> † Thus the perpetual war between the Arathapesca Indians and the Dog Rib Indians has no purpose but slaughter. Bravery in war is, in the savages' opinion, their highest virtue. Even in a civilised state it is admired, and is a basis for the special respect given to the profession in which bravery is the sole merit. There's a reason for this: we see a certain nobility in the natural disposition of someone who can make his honour an end to be valued more than life itself. . . . But the complacency with which victors boast of their mighty deeds (massacres, butchery without mercy, and the like) shows that what they really take satisfaction in is merely their own superiority and the destruction they can wreak, with no other objective.

scourge of mankind, never comes to an end. War is not in fact as incurably bad as

- that *tomb*, universal autocracy

or even as bad as

- a confederacy that exists to ensure that despotism doesn't disappear in any single state.

Still, the ancient writer was correct when he said that 'war creates more bad men than it destroys'.

·NOW BACK TO THE MAIN THREAD·

Nor (sad to say) has any philosopher been able to propose better principles that can be brought into harmony with human nature. The result is that the philosophical millennium, which hopes for a state of perpetual peace based on a league of peoples, a world-republic, is sneered at by everyone as fanaticism—as is the theological millennium, which waits for the completed moral improvement of the entire human race.

(1) It is commonly thought that the basis of this evil lies in man's sensibility [*Sinnlichkeit*] and the natural inclinations arising from it; but this can't be right. These inclinations aren't directly related to evil; rather, they provide an opportunity for the moral sense [*moralische Gesinnung*] to show virtue). And there's another reason: these natural inclinations are implanted in us from the outset—we aren't their authors—so we aren't responsible for their existence; but we *are* accountable for the propensity to evil. Why? Because it affects the man's morality, so it is present in him as a freely acting being, and it must be possible to hold him accountable for it as the offender—despite this propensity's being so deeply rooted in the will that we're forced to say that it is to be found in man by nature. Thus, as a basis for the morally bad in man, sensibility contains **too little**, because when the incentives that can arise from freedom are taken away, man is reduced to a **merely animal** being.

(2) Nor can the basis of this evil lie in a corruption of reason, the giver of the moral law—as if reason could destroy the authority of its own law, or deny the obligation arising from it! This is absolutely impossible. Also: a freely acting being can't be determined by natural laws; so if it were released from the moral law it would have to be operating without *any* law, and this is self-contradictory. Mightn't it operate on the basis of *opposition to the law* as an incentive? That would involve a thoroughly bad will, which contains **too much** for the purpose at hand—it would require the man to be a **diabolical** being. 'Merely animal', 'diabolical'—neither of these is applicable to man.

But even if the existence of this propensity to evil in human nature can be shown by empirical proofs of the opposition of man's will to the law—this being a real happening in time—such proofs don't teach us the real constitution of that propensity or the basis of this opposition. That constitution has to do with how •the will relates to •the moral law as an incentive; because the will is free the concept of it isn't empirical, and the concept of the moral law is also purely intellectual; so our grasp of how the propensity to evil is constituted—or as much grasp of it as is possible under the laws of freedom (of obligation and accountability)—must come to us *a priori* through the concept of evil. I now offer a development of that concept.

Even the lowest man doesn't. . . .repudiate the moral law, renouncing obedience to it like a rebel. The law indeed forces itself on him irresistibly by virtue of his •intrinsic• moral predisposition; and if no other incentive acted against it he would adopt it into his supreme maxim as the sufficient [here = 'sole, unaided'] determining basis of his will; i.e. he would be **morally good**. But by virtue of an equally innocent natural predisposition that he has, he clings to the incentive that relates to his senses and (in accordance with the subjective

principle of self-love) adopts it also into his maxim. [Kant speaks of this as a single incentive, but presumably he is referring to the whole set of incentives that kick off from things we encounter through our senses. Perhaps the phrase ‘the maxim of self-love’ is supposed to pull them together into a single cluster.] If he took this into his maxim as all he needs to determine the will, ignoring the moral law that he also has in him, he would be **morally bad**. So we have these premises:

- He naturally adopts both the moral law and the sense-related incentive into his maxim, and
- He would find either of these, if it were all he had, adequate in itself to determine his will.

Does it follow that

- He is at once good and bad?

No! That (as we saw on page 11) is a contradiction. It *would* follow, if the crucial difference between the maxims concerned their **content**, i.e. what incentive each recognises. But it doesn’t. Whether a man is morally good or morally bad depends on the **form** of his maxim, specifically on *which of the two incentives he makes the condition of the other*. So what makes a man bad—and even the best man *is* bad—is that he reverses the moral order of the incentives when he adopts them into his maxim. He does indeed adopt the moral law along with the law of self-love; but when he becomes aware that they can’t maintain parity with each other and that one must be subordinated to the other as its supreme condition, he makes the incentive of self-love and its inclinations the condition of obedience to the moral law; whereas the moral law ought to have been •treated as the supreme condition of the satisfaction of the incentive of self-love and thus •adopted into the universal maxim of his will as the sole incentive.

But even with this reversal of the ethical order of the incentives in his maxim, a man’s actions may still turn out

to conform to the law as much as if they had arisen from genuine basic principles. This happens when the incentives of inclination are drawn together. . . .into a single maxim under the name of *happiness*; for example, a basic principle enjoining truthfulness furthers our happiness by delivering us from the anxiety of making lies agree with one another. In this case, the man’s empirical character is good but his intelligible character is still bad.

Now if a propensity to this ‘reversal of incentives’ does lie in human nature, there is in man a natural propensity to evil; and this propensity itself is morally bad, because it must ultimately be sought in a will that is free and therefore a source of moral accountability. This evil lies deep, because it corrupts the basis of all maxims; as a natural propensity it can’t be wiped out by human powers, because that would have to be done through good maxims, and we’re discussing a situation where the ultimate subjective basis of all maxims is corrupted. Yet it must be possible for it to be outweighed because it is found in man, a being whose actions are free.

So the wickedness of human nature shouldn’t be called *malice* if that word is used in its strict sense as ‘naming’ a disposition. . . .to adopt evil *as evil* into our maxim; because that is diabolical. We should rather term it the *perversity* of the heart, which is called a bad heart because of what follows from it. It can coexist with a generally good will; it arises from two features of human nature:

- its frailty—the man’s not having the strength to follow the principles he has chosen for himself; and
- its impurity—the man’s failure to distinguish the incentives (even of well-intentioned actions) from each other by a moral standard, so that. . . .what he cares about is whether his actions conform to the law rather than whether they are motivated by the law and nothing else.

This doesn't always lead to unlawful acts and a propensity to them, . . . but the way of thinking that ignores the incentives in the maxim and attends only conformity with the letter of the law itself deserves to be called a deep-lying perversity in the human heart.

[Kant wrote this paragraph in three sentences.] This guilt is called 'innate' because it can be seen in man back when his use of freedom first appears, but it must have arisen from freedom and hence is subject to moral accountability. It has three levels, of which the first two (those of frailty and impurity) can be regarded as unintentional guilt; but at the third level it is deliberate guilt, involving something fraudulent in the human heart, in which the man •deceives himself about his own good and bad attitudes and •regards himself as justified before the law so long as his actions don't have bad consequences—which they easily could do, given the maxims that were at work in them. This is the source •of the peace of conscience of so many men—*conscientious* men, they think—when in a course of action where they didn't bring the law into their thinking, or at least didn't give it the dominant role there, they escape bad consequences by sheer good luck; and even •of their self-congratulatory sense of *merit* in not feeling themselves guilty of any such offences as they see others burdened with. They don't look into whether good luck should have the credit; or look deep enough into themselves to discover (as they could if only they would) an innermost cast of mind that would have led them

to similar morally bad conduct if they hadn't been prevented by inability, temperament, upbringing, and circumstances of time and place—none of which are things for which they are morally accountable. This dishonesty, which throws dust in our eyes and thwarts the establishing of a genuine moral attitude in us, then spreads out into falsehood and deception of others. If it isn't be called wickedness, it at least deserves the label 'worthlessness'. It is an element in the radical evil of human nature, which messes up one's capacity to make moral judgements about what a man should be taken for, and makes our attributions of responsibility—ours or those of others—wholly uncertain. It's a foul stain on our species; as long as we don't clean it out, it prevents the seed of goodness from developing as it otherwise would.

A member of the English Parliament once exclaimed, in the heat of debate, 'Every man has his price, for which he sells himself.' If this is true, . . . if there's no virtue that can't be overthrown by some temptation, . . . then certainly it holds true of men universally, as the apostle said: 'There is no difference: they are all sinners—none of them acts well in the spirit of the law, no, not one.' [*Romans 3:12*]<sup>1</sup>

#### 4. The Origin of Evil in Human Nature

An origin (a first origin) is the derivation of an effect from its first cause, i.e. the cause that isn't an effect of another cause of the same kind. It can be considered either as

<sup>1</sup> The real proof of this sentence of condemnation by morally judging reason is given in the preceding section, not in this one, which merely confirms it by experience. But experience can't reveal the root of evil in the supreme maxim of the free will relating to the law, the root which as an intelligible act precedes all experience. That there is *one* supreme maxim and *one* law to which it refers shows us why man's pure •intellectual judgment must be based on the principle that there's no middle case between good and bad; yet •empirical judgment on sensible conduct (actual doing and allowing) can be based on the principle that there is a middle between these extremes. •In fact there are two of them: a negative middle of *indifference* prior to all education, and a positive middle that is a mixture, partly good and partly bad. But this empirical judgment is merely a judgment on the morality of mankind as appearance, and in the final judgment it must submit to the pure intellectual judgment.



- an origin in reason, which explains something's existence (·not its coming into existence·) or as
- an origin in time, i.e. the cause of some event.

If an effect is referred to a cause that doesn't detract from its freedom (e.g. moral evil), then the will's being led to produce this effect is conceived of in terms not of events in time but of timeless reasons; such an effect can't be derived from any preceding state whatsoever. But that sort of derivation is always required when a bad action, as an event in the world, is related to its natural cause. To seek the temporal origin of free acts as such (as though they were natural effects) is thus a contradiction, as is the search for the temporal origin of a man's moral character, . . . because 'his moral character' means 'the basis for his exercise of freedom', and thus it must—like the determining basis of free will generally—be sought solely in representations of reason.

How are we to explain the spread and continuation of moral evil through all members and generations of our species? The clumsiest explanation is that *we inherited it from our first parents!* That's because we can say about moral evil precisely what the poet ·Ovid· said about good: 'Birth and ancestry and anything else that we didn't do ourselves I hardly consider to be ours.'<sup>1</sup> But notice this: in our search for the origin of this evil, we don't start with the propensity to evil, but focus on the inner possibility of the

actual evil of particular actions—on what factors must come together in the will if evil is to be performed.

[In this next paragraph and just once more, 'reason-origin' is used to translate *Vernunftursprung*, with no pretence of knowing what it means.] In our search for the reason-origin of bad actions, every such action must be regarded as though the individual had fallen into it directly from a state of innocence. Whatever his previous conduct may have been like, and whatever natural causes—internal or external—may have been influencing him, his action is still free and not determined by any of these causes; hence it can and must always be judged as an *original* use of his will, ·i.e. a use which is a cause that isn't caused·. Whatever his circumstances and entanglements, he *ought to* have refrained from that action; no cause in the world can deprive him of his status as a freely acting being. A man is rightly said to be accountable for the *consequences* of contrary-to-the-law actions that he has freely performed; but this merely means that there's no need to dodge around enquiring whether those consequences were free, because the admittedly free action that was their cause contains a sufficient basis for holding him accountable. A man is about to perform a free action: it doesn't matter how bad he has been up to this moment (so that evil has become habitual to him, his second nature); just as it **was then** his duty to be better, so also it **is now** his duty to better himself. It must be

<sup>1</sup> The three so-called 'higher faculties' (in the universities) would develop this notion of inherited evil each in terms of its own specialty. **(1)** For the Medical School it is an inherited **disease**, something like the tapeworm. No tapeworms have been met with anywhere but in us, not even (of this particular kind) in other animals; so some natural scientists actually believe that it must have existed in our first parents. **(2)** The Law School would regard this evil as an inherited **debt**—the legitimate consequence of inheriting the estate bequeathed us by our first parents (for being born is inheriting the use of earthly goods so far as we need them for our continued existence). This inheritance is encumbered by a serious crime, and we have to go on paying the fine until eventually death expels us from the estate. How just legal justice is! **(3)** The School of Theology would regard this evil as an inherited **sin**. They hold that our first parents played a personal part in the fall of a condemned rebel, and maintain either that we also took part •then (although we aren't now conscious of having done so) or that •now, born under the rule of the rebel (as prince of this world), we prefer the world's favours to the supreme command of the heavenly ruler, and don't have enough faith to free ourselves from this; so that we must eventually share the rebel's doom.

within his power to do his, ·because you can't have a duty to do something that is impossible for you·; so if he yet again acts badly, he is accountable. . . .right then for what he does right then. If he had been endowed with a predisposition to good. . . .and had at this moment stepped out of a state of innocence into evil, he wouldn't have been more accountable than he is ·for his bad action that is just one small addition to a lifetime of bad actions·. So if we're trying to determine and if possible explain the general subjective basis for our adopting something bad into our maxim, we should inquire not into the •temporal origin of such an action but only into its •reason-origin.

This squares with the Bible's way of presenting the origin of evil in the human species as having a *beginning*, in a narrative in which

- what in the nature of the case must be considered as *first* (but not in a temporal sense)

appears as

- coming first in time.

According to this account, evil doesn't start from an underlying propensity to evil (if it did, the beginning of evil wouldn't have its source in freedom); rather it starts from *sin*, meaning the transgressing of the moral law as a divine command. The state of man prior to any propensity to evil is called the state of *innocence*. The moral law was presented to mankind as a *prohibition* (*Genesis* 2:16–17); it had to be presented in that way to beings who were not not pure but tempted by desires. Now instead of straightforwardly following this law as an adequate incentive (the only unconditionally good incentive, the only one that there's no room for doubt about), the man

[see Glossary] looked around for other incentives (*Genesis* 3:6), ones that can be good only conditionally, specifically on the condition that they don't infringe the law. He then made it his maxim (I'm thinking here of his action as consciously springing from freedom) to follow the law of duty not ·solely· •as duty but also, in cases of need, •as furthering other ends. That started him wondering whether the commandment's exclusion of the influence of all other incentives was really meant so strictly; and his next move was to use subtle reasoning [see Glossary] to downgrade •obedience to the law to •the merely conditional character of a *means* (subject to the principle of self-love);<sup>1</sup> and finally he admitted into his maxim of conduct the ascendancy of the sensuous impulse over the incentive arising from the law—and thus sin occurred (*Genesis* 3:6). *Mutato nomine de te fabula narratur* [quoted from the Latin poet Horace; it means 'With the name changed, it's *your* story they are telling']. This clearly implies that we daily act in the same way, and that therefore 'in Adam all have sinned' and still sin; except that in us an innate propensity to transgression is presupposed, whereas the first man is credited with no such inborn propensity but rather with *a period of time in which* he is innocent. So his transgression is called a *fall* into sin; whereas our sin is represented as resulting from an innate wickedness in our nature. But all this means is that if we try to explain ·our· evil in terms of its beginning in time, we have to look for the cause of each deliberate transgression in a previous period of our lives, ·eventually being led· right back to a time when we didn't yet have the use of reason, and thus to see the source of evil in a propensity to evil (as a natural basis) that is therefore called

<sup>1</sup> All homage paid to the moral law is hypocritical if in one's maxim one doesn't grant to the law, as an incentive that is sufficient in itself, a higher rank than all the other determining bases of the will; and the propensity to do this is inward deceit, i.e. a propensity to lie to oneself in the interpretation of the moral law, to its detriment (*Genesis* 3:5). Accordingly, the Christian part of the Bible calls the author of evil (who is within us) 'the liar' right from the outset (*John* 8:44), and thus characterises man in terms of what seems to be the chief basis of evil in him.

‘innate’. But we don’t need to—and anyway we *can’t*—trace ·in this way· the causes of evil in the case of the first man, who is depicted as already having full command of the use of his reason; because on such an account the propensity to evil would have to have been created in him. That is why his sin is depicted as generated directly from innocence. But we mustn’t look for an origin in time of a moral character for which we are to be held accountable; though we can’t help doing so when we want to explain the sheer fact that we have this character. When the Bible depicts the origin of evil in this temporal way, perhaps it is allowing for this weakness of ours.

But the reason-origin of this propensity to evil—i.e. this perversion of our will in which it gives lower incentives dominance in its maxims—remains inscrutable to us. Here is why. [This indented passage mainly expands a super-compressed half-sentence of Kant’s, in ways that the ‘small dots’ convention can’t easily indicate.]

This propensity is something for which we are accountable; so a straightforwardly graspable explanation for it would have to involve our having adopted a bad maxim as its basis—evil must have sprung from something morally bad; it couldn’t have come from mere limitations in our nature. And this bad something wouldn’t give us what we wanted unless it was itself *basic*, and not an upshot of something earlier or deeper for which we were accountable and which would start off a new search for an explanation. But the basic human predisposition is a predisposition to *good*! Might it not have become corrupted? If the man

is accountable for the corruption then *he* must have done it, and our search for explanation starts up all over again.

So there is for us no conceivable basis from which the moral evil in us could originally have come. This inconceivability, together with a closer specification of the wickedness of our species, the Bible expresses in an historical narrative<sup>1</sup> which finds a place for evil at the creation of the world—not in man but in a *spirit* that was originally destined for something much higher. Thus the beginning of all evil is represented as inconceivable to us (for where did the evil in that spirit come from?); but man is represented as having fallen into evil only through seduction, and hence as being not basically corrupt. . . .rather as still capable of an improvement; unlike the seducing spirit, a being whose guilt can’t be lessened by pleading temptation of the flesh. For man, therefore, who despite a corrupted heart has a good will, there remains hope of a return to the good from which he has strayed.

### **General remark: Restoring the Original Predisposition to Good to its Power**

[On page 27 Kant says that this General remark could be entitled ‘Works of Grace’.]

Whatever a man’s moral condition •is or •will be, whether good or bad, that must be something that he •has brought on himself or •is now bringing on himself. It must be an effect of his free choice, for otherwise he wouldn’t be accountable for it and therefore he wouldn’t be *morally* good or bad. When it is said that ‘Man is created good’, this can only mean

<sup>1</sup> I don’t offer this as biblical exegesis, which lies outside the realm of bare reason. We can explain how to put an historical account to a moral use without deciding whether that’s what the author intended or is merely something inserted by us, provided this meaning is •true in itself (never mind how it squares with history) and also is •the only one that will let us get something salutary from a passage that would otherwise be only an inert addition to our historical knowledge. . . .

that he is created *for* good, and his original predisposition is good; but whether he is already actually good or bad depends entirely on his free choice of whether to admit into his maxim the incentives that this predisposition ‘to good’ carries with it. Even if some supernatural cooperation was needed for him to become good (or better)—some positive assistance or reducing of obstacles—his own free will must come into play in two ways: he must ‘first make himself worthy to receive this help, and ‘then he must *accept* it (which is no small matter), i.e. he must admit this positive increase of power into his maxim. If he weren’t involved in this way he wouldn’t be responsible for his goodness and wouldn’t be known as a good man.

How can a naturally bad man turn himself into a good man? No answer to that is within our conceptual reach, for how can a bad tree bear good fruit? . . . But we are accepting that the descent from good into evil occurs, and it is no more comprehensible than the climb back up from evil to good; each of them originates in freedom. So it can’t be disputed that the climb back up is possible. For despite the fall, the command telling us *We ought to become better* resounds in full strength in our souls; so it must be within our power to do this, even if what we’re able to do isn’t in itself sufficient to achieve this, and only makes us receptive to an inscrutable higher assistance. It must be taken for granted that through all this a seed of goodness has remained in its entire purity, incapable of being eliminated or corrupted; and this seed certainly can’t be *self-love*, which if accepted as the principle of all our maxims, is the very source of evil.

·REPORT ON AN OMITTED FOOTNOTE·

[Kant has at this point a difficult footnote which starts by sorting out an ambiguity in the term *self-love*. He distinguishes

(a) *benevolentia* [Latin]; my having this self-love is my **wishing myself well**, wanting things to go well for me, having myself as an object of my benevolence,

from

(a) *complacentia* [Latin]; my having this self-love is my **liking myself**.

Kant’s labels for these have an overlap:

(a) *Wohlwollen* — (b) *Wohlgefallen*

which could be clumsily Englished as

(a) ‘well-wanting’ — (b) ‘well-liking’;

but the overlap of *Wohl* in the German names can’t be put into civilised English that captures the intended meaning.

Here are the main points that Kant makes in the course of this footnote.

[It is natural to (a) wish oneself well; and reason can come into this in two ways: in connection with the choice of ‘the best and most durable kinds of well-being and ‘the best means to them. This use of reason doesn’t involve morality; in it reason is only ‘the maid-servant to natural inclination’ [compare Hume’s ‘the slave of the passions’]. But if the principle of wishing oneself well is made ‘the unconditional principle of the will’ it is the source of an intense antagonism to morality.

[I might (b) like myself because of how well I have done—success in business, nice family, etc.—but that kind of (b) isn’t significantly different from (a), and Kant sets it aside. That leaves him with *unconditional* liking for oneself (ULFO), a liking that owes nothing to any facts about how happy one’s life has been etc. This ULFO, Kant says, is possible only for someone whose maxims of action completely agree with the moral law; anyone conscious of having maxims that don’t square with the moral law within him will inevitably have a bitter *dislike* of himself. (The only exception would be someone to whom morality was indifferent, i.e. whose attitude to morality was ‘I can take it or leave it’.) Notice that

Kant has put a certain condition **on** the ULFO (you can have it only *if*. . .), but he doesn't put any condition **into** it.

[This ULFO could be called 'the reasonable love of oneself', Kant says, because it prevents the man from giving play to other incentives, ones aiming at this or that state of affairs under the label of 'happiness'. He then discusses the role of the concept of happiness in the ULFO, and concludes that my ULFO aims only at *my being worthy of happiness*; that can be basic, underived, and unconditional, and none of that is anywhere near to being true of any aim I may have to be 'happy' in the sense of achieving my non-moral aims.]

·END OF REPORT·

So restoring our original predisposition to good is not acquiring a *lost* incentive for good; that incentive—which consists in respect for the moral law—is something we have never been able to lose, and if we could and did lose it we could never get it back. What is restored is the moral law's *purity* as the ultimate basis of all our maxims, so that it doesn't merely collaborate with other incentives (inclinations), let alone being subordinated to any of them as its conditions, but is adopted in its entire purity as an incentive that is adequate *in itself* to determine the will. What is originally [here = 'basically', 'ultimately'] good is the holiness of maxims in doing one's duty merely for duty's sake. When a man allows this purity into his maxim, that doesn't make him holy (there's a great gap between the maxim and the deed!), but he puts himself on the road of endless progress towards holiness. When the firm resolve to do his duty has become habitual with him, he is said to have 'the virtue of conformity to law'—this conformity is virtue's **empirical character**. This virtue has as its steadfast maxim *Act in conformity to the law*; and there's nothing here about the incentives the will needs to get it to follow that maxim. Virtue in this sense is achieved

a little bit at a time; and in some cases a man requires long practice in observing the law, during which he passes from •a tendency to vice, through •gradual reform of his conduct and strengthening of his maxims, to •an opposite tendency. This doesn't need a change of heart [*Herzensänderung*]*—only a change of conduct. The man regards himself as virtuous if he feels that he has a firm hold on maxims of obedience to his duty even if these maxims don't arise from the ultimate basis for all maxims, namely from duty itself. For example,*

- the immoderate man turns to temperance for the sake of health,
- the liar to honesty for the sake of reputation,
- the unjust man to civic righteousness for the sake of peace or profit,

and so on—all according to the precious principle of *happiness!* But how can he become not merely law-abiding but morally good (pleasing to God)? come to be someone endowed with virtue in its **intelligible character**? someone who when he knows that it's his duty to do x doesn't need any other incentive to go ahead and do it? This can't be brought about through gradual reformation so long as the basis of the maxims remains impure, ·i.e. as long as non-moral incentives are part of his motivational mix·; it has to happen through a revolution in the man's attitude, a going over to the maxim of the attitude's holiness. He can become a new man only by a kind of rebirth, as it were a new creation, and a change of heart.

But if a man is corrupt in the basis of his maxims, how can he possibly bring about this revolution, using his own powers to become a good man? ·It seems impossible·, yet duty tells us to do it, and duty doesn't demand anything that we can't do. The only way to reconcile these is to say that because it is necessary for man it must be possible for him to undergo

- a ·total· revolution in his way of thinking, but only
- a gradual reform in his way of sensing (which places obstacles in the way of the former).

That is, if by a single unchangeable decision he reverses the ultimate basis of his maxims whereby he was a bad man (and thus ‘puts on the new man’ [Kant is quoting from St Paul here]), that makes him, so far as his principle and way of thinking are concerned, •someone who is receptive to goodness; but only in continuous labour and growth is he •a good man. Because of the purity of the principle he has adopted as the supreme maxim of his will, and because of its stability, he can hope to find himself on the good (though narrow) path of continual *progress* from bad to better. For him who sees through to the intelligible basis of the heart (i.e. of all the maxims of the will), and for whom this unending journey towards being a good man is a single step, i.e. for God, this amounts to his actually being a good man (pleasing to God); and to that extent this change can be regarded as a revolution. But in the judgment of men, who can assess themselves and the strength of their maxims only by how well they come, over time, to dominate ·the inclinations generated by· their way of sensing, this change must be regarded as but an everlasting struggle toward the better, and thus as a gradual reform of that perverted cast of mind, the propensity to evil.

It follows that a man’s moral growth has to start not by •improving his conduct but by •transforming his way of thinking and •laying the foundations of his character. Yet customarily people tackle this differently, fighting against vices piecemeal while leaving undisturbed their common root. But even the most limited man is capable of being struck by respect for an action conforming to duty—a respect that is greater the more he isolates it in thought from self-interested incentives that might influence the maxim

of conduct. Even children can detect the smallest trace of an improper incentive; they see an action thus motivated as instantly losing all moral worth. There’s no better way of developing this predisposition in the young than by getting them •to attend to examples of actual conduct—it can be the conduct of men who are *good* in the sense that their conduct conforms to law—and •to judge the impurity of the maxims that led to them. This works its way into their way of thinking, so that *duty for its own sake* begins to have a noticeable weight in their hearts. But teaching a pupil to *admire* virtuous actions doesn’t favour his feeling for moral goodness, even if the actions have involved great sacrifice. However virtuous a man is, all the good he can ever do is merely his simple duty; and doing his duty is nothing more than doing what is in the common moral order and hence not something to be admired. Admiration will lower our feeling for duty, as if doing one’s duty were something extraordinary and meritorious.

But there’s one thing in our soul that we can’t stop from regarding with the highest wonder [*Verwunderung*], when we view it properly, and for which admiration [*Bewunderung*] is both legitimate and even uplifting—I’m talking about the fundamental moral predisposition in us. [Kant says that what we are to admire is that predisposition *überhaupt*—not any detailed facts about this predisposition but just the fact that we have it at all.]

We are beings whose needs make us dependent on nature in ever so many ways, yet we are also raised so far above these needs. . . .that we count them as *nothing*, and count ourselves as unworthy of existence if we put •satisfying them ahead of •conforming to the law—a law through which our reason commands us powerfully yet without making promises or threats; and all this despite the fact that what makes life worth desiring is the satisfaction of those needs.

**What is it within us that produces this result?** This question must weigh on the mind of any man, however unintelligent, who has been taught the holiness that inheres in the idea of duty but who hasn't yet advanced to an inquiry into the primary output of this law, the concept of freedom.<sup>1</sup> And the very incomprehensibility of this predisposition, which announces a divine origin, works on the mind. . . .and strengthens it for whatever sacrifice a man's respect for his duty may demand of him. An especially good way to awaken a man's moral sentiments is to arouse in him, *often*, this feeling of the sublimity of his moral destiny. Why? Because this works directly •against the innate propensity to pervert the incentives in the maxims of our will and •toward the re-establishment in the human heart of an unconditional respect for the law as the ultimate test of which maxims are to be adopted, i.e. of the original [here = 'fundamental'] moral order among the incentives, and so of the predisposition to good in all its purity.

But doesn't this restoration through one's own exertions slam up against the thesis of the innate corruption of man that unfits him for all good? It does indeed, as far as the conceivability. . . .of such a restoration is concerned. This is true of everything that is to be represented as an event

in time,. . . .and thus as necessary under the laws of nature, while its opposite is to be represented as possible through freedom under moral laws. But the corruption thesis doesn't conflict with the possibility of this restoration itself. If the moral law commands that we *ought* now to be better men, it unavoidably follows that we *can* now be better men. The innate-evil thesis is useless in moral •dogmatics, whose precepts have the same content and the same force whether or not we have an innate tendency toward transgression. But in moral •self-discipline this postulate has more to say, though only this much more:

In the moral development of the predisposition to good implanted in us, we must start not from an innocence that is natural to us but from the assumption of our will's wickedness in adopting its maxims contrary to the original moral predisposition; and since this propensity ·to evil· is ineradicable we must fight against it incessantly.

Because this leads only to a never-completed progress from bad to better, it follows that the ·total· conversion of a bad man's disposition into that of a good man has to be identified with a change resulting in

<sup>1</sup> The concept of the freedom of the will doesn't precede our consciousness of the moral law within us; it is inferred from the fact that this law can determine our will as an unconditional command. To be convinced of this, ask yourself: 'Am I certainly and directly conscious of power to overcome, by a firm resolve, every incentive to transgression, however great?' You have to admit that you *don't know* whether in such a case you wouldn't be shaken in your resolve. Yet duty commands you unconditionally: you *ought* to remain true to your resolve; and from this you rightly conclude that •you must be able to do so, and that therefore •your will is free. [Kant adds that some philosophers have contended that free will is perfectly comprehensible, doing this—helped by a certain concept of 'determinism'—by attacking a 'problem' that hasn't bothered anyone. He concludes:] The real problem concerns *predeterminism*, according to which voluntary actions are events whose determining bases lie back in earlier time (which, with what happened in it, is no longer within our power). How can *this* be consistent with freedom, according to which doing A and not doing A are both within the subject's power at the moment of of action? *That* is what we want to understand, and never shall.

† There's no problem about reconciling the concept of freedom with the idea of God as a necessary Being. What is needed for God's freedom is the absolute spontaneity of his actions, and the only threat to this would have to come from *predeterminism*, where the determining basis of the action is in earlier time; but God doesn't exist in time, so this difficulty vanishes.

the conformity to the moral law of the ultimate inward basis for the adoption of *all* his maxims,

given that this new basis (the new heart) can't be changed in its turn. A man can't naturally get assurance that such a conversion has occurred, whether by immediate consciousness or from the evidence of changes in his way of life; because the depths of the heart (the subjective ultimate basis of his maxims) are inscrutable to him. But he must be able to *hope* •to reach the road that leads to it—the road pointed out to him by a fundamentally improved disposition—and •to do this through his own efforts, because. . . .he can count as morally good only by virtue of actions he is accountable for, actions performed by himself.

Against this demand for self-improvement, reason brings in all sorts of ignoble religious ideas (including the false ascription to God himself of the principle of happiness as the supreme condition of his commandments). [Kant builds into that sentence the claims •that reason 'is by nature averse to the work of moral reconstruction', and •that it enlists bad religious ideas 'under the pretext of natural incapacity', i.e. on the plea that it isn't up to doing the job itself. We'll soon see that this isn't a joke.]

All religions can be divided into

(a) favour-seeking religion (mere worship) and

(b) moral religion, i.e. the religion of the good way of life.

In (a) the man flatters himself by believing that of course God can make him eternally happy (through remission of his sins) without his having to become a better man, or at least. . . .that God can make him a better man without his having to do anything but ask for it; which amounts to doing nothing at all, because asking an all-seeing Being for something is equivalent to merely *wanting* it. . . . But in (b) Christianity (the only moral religion there has ever been) it is a basic principle that each person must do everything

in his power to become a better man, and that what is not within his power will be made up for through cooperation from above—but only if. . . .he has worked on becoming a better man through his basic predisposition to good. It's not absolutely necessary for him to know what this cooperation consists in. . . .; but it is essential for him to know what *he* must do in order to become worthy of this help.

† This General Remark is the first of four that are appended to the Essays in this work, one each. They could carry the titles:

(1) Works of Grace,

(2) Miracles,

(3) Mysteries,

(4) Means of Grace.

These are, as it were, accessories to religion within the bounds of pure reason; they don't fall within that territory but they bump up against it from the outside. Reason, conscious of its inability to satisfy its moral need, stretches out to high-flown ·religious· ideas that can make up for this, but it doesn't expand its domain so as to take them *into* it. Without disputing that the objects of these ideas are possible, or even that they are real, reason simply can't admit them into its maxims of thought and action. It holds that if in the inscrutable realm of the supernatural there's something that it can't understand but that may be needed to make up for its moral insufficiency, this unknown something will be available to its good will. Its attitude to the possibility of this supernatural supplement might be called **reflective** belief, in contrast with **dogmatic** belief, proclaims itself as a form of knowledge and strikes reason as dishonest or presumptuous. . . . If we try to introduce these morally high-flying ideas *into* religion, the upshots are:



- (1) from the supposed inward experience of works of grace, *fanaticism*;
- (2) from the alleged outer experience of miracles, *superstition*;
- (3) from a supposed enlightening of the understanding with regard to supernatural mysteries, *illumination* [see Glossary]. . . .;
- (4) from daring attempts to affect the supernatural so as to get means of grace, *thaumaturgy* [see Glossary].

These are all sheer aberrations of a reason that goes beyond its proper limits, doing this for a purpose that it fancies to be moral (pleasing to God).—Focusing now on (1) works of grace: calling works of grace to our aid is one of those aberrations,

and can't be admitted into the maxims of reason if it is to stay within its limits; nor can *anything* supernatural, simply because in the realm of the supernatural all use of reason ceases. Can't we come to know them **theoretically**, by finding evidence that they are works of •grace and not inner •natural effects? No, because we can't extend the concept of cause and effect beyond matters of experience or, therefore, beyond nature. And the hypothesis of a **practical** application of this idea is self-contradictory. [Kant explains why: If we are to deserve any credit for becoming good (or better), this must have happened through something we did; whereas relying for this on works of grace is trying to get moral credit by doing nothing. He concludes:] So we can admit a work of grace as something incomprehensible, but we can't admit it into our maxims either for theoretical or for practical use.