

# Groundwork for the Metaphysic of Morals

Immanuel Kant

Copyright ©2010–2015 All rights reserved. Jonathan Bennett

[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 square brackets in normal-sized type.] In the title, ‘Groundwork’ refers not to the foundation that is laid but to the work of laying it.

First launched: July 2005

Last amended: September 2008

## **Contents**

|   |           |
|---|-----------|
| <b>Preface</b>  | <b>1</b>  |
| <b>Chapter 1: Moving from common-sense knowledge to philosophical knowledge about morality</b>  | <b>5</b>  |
| <b>Chapter 2: Moving from popular moral philosophy to the metaphysic of morals</b>              | <b>14</b> |
| <b>Chapter 3: Moving from the metaphysic of morals to the critique of pure practical reason</b> | <b>41</b> |

## Chapter 2:

### Moving from popular moral philosophy to the metaphysic of morals

Although I have derived our existing concept of duty from the ordinary commonsensical use of our practical reason, that doesn't at all imply that I have treated it as an empirical concept. On the contrary, if we attend to our experience of men's doings, we meet frequent and—I admit—*justified* complaints that we can't cite a single sure example of someone's being disposed to act from pure duty—not one!—so that although much is done that accords with what duty commands, it always remains doubtful whether it is done *from* duty and thus whether it has moral worth. That is why there have always been philosophers who absolutely denied the reality of this dutiful disposition in human actions, attributing everything that people do to more or less refined self-interest. This hasn't led them to question the credentials of the concept of morality. Rather, they have left that standing, and have spoken with sincere regret of the frailty and corruption of human nature, which is high-minded enough to accept the idea of duty—an idea so worthy of respect—as a source of commands, is too weak to follow this idea by obeying the commands, and employs reason, which *ought* to be its source of laws, only to cater to the interests that its preferences create—either singly or, at best, in their greatest possible harmony with one another.

It is indeed absolutely impossible by means of experience to identify with complete certainty a single case in which the maxim of an action—however much it might conform to duty—rested solely on moral grounds and on the person's thought of his duty. It sometimes happens that we make a considerable sacrifice in performing some good action, and can't find within ourselves, search as we may, anything

that could have the power to motivate this except the moral ground of duty. But this shouldn't make us confident that the true determining cause of the will was actually our sense of duty rather than a secret impulse of self-love masquerading as the idea of duty. For we like to give ourselves credit for having a more high-minded motive than we actually have; and even the strictest examination can never lead us entirely behind the secret action-drivers—or, rather, behind the pretended action-driver to where the real one secretly lurks—because when moral worth is in question it is not a matter of visible actions but of their invisible inner sources.

The claim that the concept of duty is an empirical one is not only false but dangerous. Consider the people who ridicule all morality as a mere phantom of human imagination overreaching itself through self-conceit: one couldn't give *them* anything they would like better than the concession that the concepts of duty have to come wholly from experience (for their laziness makes them apt to believe that the same is true of all other concepts too). This concession would give them a sure triumph. I am willing to admit—out of sheer generosity!—that most of our actions are in accord with duty; but if we look more closely at our thoughts and aspirations we keep encountering the beloved *self* as what our plans rely on, rather than the stern command of duty with its frequent calls for *self*-denial. One needn't be an enemy of virtue, merely a cool observer who can distinguish even the most intense wish for the good from actual good, to wonder sometimes whether true virtue is to be met with *anywhere* in the world; especially as one gets older and one's

power of judgment is made wiser by experience and more acute in observation. [Kant was 60 years old when he wrote this work.] What, then, can stop us from completely abandoning our ideas of duty, and preserve in us a well-founded respect for its law? Only the conviction that

- Even if there never *were* any actions springing from such pure sources, *that's* not the topic. Our concern is not •with whether this or that *was done*, but •with reason's commanding—on its own initiative and independently of all appearances—what *ought to be done*.

So our concern is •with •a kind of• actions of which perhaps the world has never had an example; if you go purely by experience you might well wonder whether there *could* be such actions; and yet they are sternly commanded by reason. Take the example of *pure sincerity in friendship*: this can be demanded of every man as a duty; the demand comes independently of all experience from the idea of reason that acts on the will on *a priori* grounds; so it isn't weakened in the slightest by the fact—if it is a fact—that there has never actually been a sincere friend.

When this is added:

- If we don't want to deny all truth to the concept of morality and to give up applying it to any possible object, we have to admit that morality's law applies so widely that it holds •not merely for men but for all rational beings as such, •not merely under certain contingent conditions and with exceptions but with absolute necessity •and therefore unconditionally and without exceptions•,

—when *this* becomes clear to us, we see that no experience can point us towards even the *possibility* of such apodictic laws. [This word, like the German *apodiktisch*, comes from Greek meaning, roughly, 'clearly demonstrated'. Kant uses it to mean something

like 'utterly unbreakable, unconditional, permitting no excuses or exceptions'.] For what could entitle us to accord *unlimited* respect to something that perhaps is valid only under contingent human conditions? And how could laws for •our will be held to be laws for •the will of any rational being (and valid for us only because we are such beings), if they were merely empirical and didn't arise *a priori* from pure though practical reason?

One couldn't do worse by morality than drawing it from examples. We can't get our concept of morality initially from examples, for we can't judge whether something is fit to be an example or model of morality unless it has *already* been judged according principles of morality. •This applies even to •the model that is most frequently appealed to•. Even •Jesus Christ must be compared with our ideal of moral perfection before he is recognized as being perfect; indeed, he says of himself 'Why callest thou me (whom you see) *good*? There is none good (the archetype •or model• of good) but one, i.e. God (whom you don't see)' [Matthew 19:17; the bits added in parentheses are Kant's]. But •don't think that with God the father we have at last found the example or model from which we can derive our concept of morality•. Where do we get the concept of *God as the highest good* from? Solely from the idea of moral perfection that reason lays out for us *a priori* and which it ties, unbreakably, to the concept of a free will. •Some have said that the moral life consists in 'imitating Christ', but imitation has no place in moral matters; and the only use of examples there is •for encouragement—i.e. showing beyond question that what the law commands *can* be done—and •for making visible •in particular cases• what the practical rule expresses more generally. But they can never entitle us to steer purely by examples, setting aside their true model which lies in reason.

Well, then, there are moral concepts that are established *a priori*, along with the principles of morality. *Would it be a good idea to set these out in abstract form?* Given that

- there is no genuine supreme principle of morality that *doesn't* rest on pure reason alone independently of all possible experience,

·and thus given that

- the *a priori* concepts and principles I have mentioned are the whole foundation for morality·,

I don't think there should be any question about whether they should be presented abstractly. At any rate, there should be no question about that if we want our knowledge of them to be distinguished from ordinary knowledge and to merit the label 'philosophical'. But these days the question may arise after all. For if we conducted a poll on the question:

Which would you prefer—•pure rational knowledge of morality, separated from all experience and bringing with it a metaphysic of morals, or •popular practical philosophy?

it is easy to guess on which side the majority would stand!

Catering to the notions of the man in the street is all very well *after* we have made a fully satisfactory job of ascending to the principles of pure reason—*first* providing a metaphysical basis for the doctrine of morals and *then* getting it listened to by popularizing it. But it's utterly absurd to aim at popularity [here = 'being accessible by the common man'] at the outset, where everything depends on the correctness of the fundamental principles. There is a real virtue—a rare one!—in genuine popularization of philosophy; but the procedure I have been describing, ·in which popularity is sought at the outset·, involves no such virtue. It is not hard

to be generally comprehensible if one does it by dropping all basic insight and replacing it with a disgusting jumble of patched-up observations and half-reasoned principles. Shallow-minded people lap this up, for it is very useful in coffee-house chatter, while people with better sense feel confused and dissatisfied, and helplessly turn away. Philosophers who see right through this hocus-pocus call people away from sham 'popularity' and towards the genuine popularity that can be achieved on the basis of hard-won insights; but they don't get much of a hearing.

When we look at essays on morality written in this beloved style, what do we find? Sometimes •human nature in particular is mentioned (occasionally with the idea of a rational nature in general); now •perfection shows up, and now •happiness; •moral feeling here, •fear of God there; a •little of this and a •little of that—all in a marvellous mixture. It never occurs to the authors to ask: *Can* the principles of morality be found in knowledge of human nature (knowledge that we can get only from experience)? If they can't—if the principles are *a priori*, free from everything empirical, and to be found in pure rational concepts with not a trace of them anywhere else—shouldn't we tackle the investigation of them as a separate inquiry, as *pure* practical philosophy or (to use the dread word) as a *metaphysic* of morals,<sup>2</sup> dealing with it on its own so as to bring it to completion and make the popularity-demanding public wait until we have finished?

·The answer to that last question is 'Yes, we should', because· a completely self-contained metaphysic of morals, with no admixture of anthropology or theology or physics or . . . occult qualities. . . ., is not only an essential basis for all theoretically sound and definite •knowledge of duties,

<sup>2</sup> We can if we wish divide the philosophy of morals into 'pure' (metaphysics) and 'applied' (meaning 'applied to human nature'), like the divisions of mathematics and logic into pure and applied. This terminology immediately reminds us that moral principles are not based on what is special in human nature but must stand on their own feet *a priori*, and that they must yield practical rules for every rational nature, and accordingly for man.

but also a tremendously important help towards actually carrying out its precepts. For the *pure* thought of duty and of the moral law generally, unmixed with empirical inducements, has a stronger influence on the human heart purely through reason—this being what first shows reason that it *can* be practical—than all other action-drivers that may be derived from the empirical field; so much stronger that reason, aware of its dignity, despises the empirical inputs and comes to dominate them. In contrast with this, a mixed theory of morals— assembled from action-drivers involving feelings and preferences and from rational concepts—is bound to make the mind vacillate between motives that can't be brought together under any principle and that can lead to the good only by great good luck and will frequently lead to the bad.<sup>3</sup>

What I have said makes five things clear: that all moral concepts have their origin entirely *a priori* in reason, and this holds as much for the most ordinary common-sense moral concepts as for the ones used in high-level theorizing; that moral concepts can't be formed by abstraction from any empirical knowledge or, therefore, from anything contingent; that this purity or non-empiricalness of origin is what gives them the dignity of serving as supreme practical principles; that any addition of something empirical takes away just that much of their influence and of the unqualified worth of actions performed in accordance with them; and that not

only is it necessary in developing a moral *theory* but also important in our *practical* lives that we derive the concepts and laws of morals from pure reason and present them pure and unmixed, determining the scope of this entire practical but pure rational knowledge (the entire faculty of pure practical reason). [What follows is meant to flow on from that fifth point; Kant wrote this paragraph as one sentence.] This determination of *scope* is to be done *not* on the basis of principles of *human* reason that non-moral philosophy might allow or require, *but rather* (because moral laws are to hold for every rational being just because it is rational) by being derived from the universal concept of *rational being*. To apply morals to men one needs anthropology; but *first* morals must be completely developed as pure philosophy, i.e. metaphysics, independently of anthropology; this is easy to do, given how separate the two are from one another. For we know—and here I repeat the fifth of the points with which I opened this paragraph—that if we don't have such a metaphysic, it is not merely pointless to try to settle accurately, as a matter of theory, what moral content there is in this or that action that is in accord with duty, but impossible to base morals on legitimate principles even for ordinary practical use, especially in moral instruction; and that's what is needed for pure moral dispositions to be produced and worked into men's characters for the purpose of the highest good in the world.

<sup>3</sup> I have been asked. . . why teachings about virtue containing so much that is convincing to reason nevertheless achieve so little. . . . The answer is just this: the teachers themselves haven't brought their concepts right out into the clear; and when they wish to make up for this by hunting all over the place for motives for being morally good so as to make their medicine have the right strength, they spoil it. Entertain the thought of

an act of honesty performed with a steadfast soul, with no view towards any advantage in this world or the next, under the greatest temptations of need or allurements.

You don't have to look very hard to see that conduct like *this* far surpasses and eclipses any similar action that was affected—even if only slightly—by any external action-driver. It elevates the soul and makes one want to be able to act in this way. Even youngish children feel this, and one should never represent duties to them in any other way.

In this study I have already moved

- from common moral judgment to philosophical moral judgment,

and am now advancing by natural stages *·within* the realm of philosophical moral judgment, specifically:

- from popular philosophy to metaphysics.

Popular philosophy goes only as far as it can grope its way by means of examples; metaphysics is not held back by anything empirical, and, because it has to stake out the whole essence of rational knowledge of this kind, it will if necessary stretch out as far as ideas *·of reason·*, of which there can't be any examples. In making this advance we must track and clearly present the practical faculty of reason, right from *•the universal rules that set it up through to •the point where the concept of duty arises from it.*

Everything in nature works according to laws. Only a rational being has a *will*—which is the ability to act according to *the thought of* laws, i.e. to act on principle. To derive actions from laws you need reason, so that's what *will* is—practical reason. When *•reason* is irresistible in its influence on the will, the actions that a rational being recognizes as objectively necessary are also subjectively necessary; i.e. the will is an ability to choose *only* what reason recognizes, independently of preferences, as practically necessary, i.e. as good. But when *•unaided reason* isn't enough to settle the will, the will comes under the influence of subjective

conditions (certain action-drivers) that don't always agree with the objective conditions—in short, the will is not in complete accord with reason. In this case (which is the actual case with men) the actions that are recognized as *•objectively necessary* are *•subjectively contingent*, and if such a will is determined according to objective laws that is because it is *constrained*. . . i.e. is following principles of reason to which it isn't by its nature necessarily obedient.

When the thought of an objective principle constrains a will, it is called a 'command' (of reason), and its verbal expression is called an 'imperative'.

All imperatives are expressed with an 'ought', which indicates how an objective law of reason relates to a will that it *constrains*. An imperative says that it would be *good* to do or to refrain from doing something, but it addresses this to a will that doesn't always do *x* just because *x* is represented to it as *good* to do. Practical good is what determines the will by means of the thoughts that reason produces—and thus not by subjective causes but objectively, on grounds that are valid for every rational being just because it is rational. This contrasts with the thought that it would be *nice* to act in a certain way; the latter influences the will only by means of a feeling that has purely subjective causes, which hold for the senses of this or that person but not as a principle of reason that holds for everyone.<sup>4</sup>

<sup>4</sup> When the faculty of desire is affected by feelings, we speak of what the person *•prefers*, which always also indicates a *•need*. When a contingently determinable will is affected by principles of reason, we say that it has an *•interest*. Interests are to be found only in a dependent will, one that isn't of itself always in accord with reason; we can't make sense of the idea of God's will's having interests. But even the human will can *have* an interest without *acting on* it. The interest that one merely *has* is a practical interest in the *•action*; the interest on which one *acts* is a pathological interest in the *•upshot* of the action. [See the note on 'pathological' on page 9.] Whereas the former indicates only the effect on the will of principles of reason *•in themselves*, the latter indicates the effect on it of the principles of reason *•in the service of the person's preferences*, since *·in these cases·* all reason does is to provide the practical rule through which the person's preferences are to be satisfied. In the former case, my focus is on the action; in the latter, it is on whatever is pleasant in the result of the action. We saw in chapter 1 that when an action is done from duty, attention should be paid not to any interest in its upshot but only to the action itself and the law which is its principle in reason.

Objective laws of the good would *apply* to a perfectly good will just as much to as to any other; but we shouldn't think of them as *constraining* such a will, because it is so constituted that it can't be determined to act by anything except the thought of the good. Thus no imperatives hold for God's will or for any holy will. The 'ought' is out of place here, for the volition is of itself *necessarily* at one with the law. Thus, what imperatives do is just to express the relation of •objective laws of volition •in general to the •subjective imperfection of the will of this or that •particular rational being—the will of any human, for example.

All imperatives command either •hypothetically or categorically. The •former expresses the practical necessity of some possible action as a means to achieving something else that one does or might want. An imperative would be categorical if it represented an action as being objectively necessary in itself without regard to any other end.

Since every practical law represents some possible action as •good, and thus as •necessary for anyone whose conduct is governed by reason, what every imperative does is to specify some action that is

•necessary according to the principle of a will that has something good about it.

If the action would be good only as a means to something else, the imperative is **hypothetical**; but if the action is thought of as *good in itself* and hence as

•necessary in a will that conforms to reason, which it has as its principle,

the imperative is **categorical**.

The imperative thus says of some action I could perform that it would be good, and puts the practical rule into a relationship with my will; •and it is no less an imperative if I don't immediately perform the •commanded• action simply because it is good (I don't know that it is good, or I do know

this but •I don't care, because• my conduct is guided by other maxims that are opposed to the objective principles of practical reason).

A **hypothetical** imperative merely says that the action is good for some purpose that one *could* have or that one actually *does* have. In the •former case it is a problematic practical principle, in the •latter it is an assertoric one. The **categorical** imperative, which declares the action to be objectively necessary without referring to any end in view. . . .holds as an apodictic practical principle.

Anything that could come about through the powers of some rational being *could* be an end •or goal or purpose• for some will or other. So •there are countless possible ends, and therefore• countless hypothetical imperatives, i.e. principles of action thought of as necessary to attain a possible end in view. Every science has a practical segment in which

•some purpose is set forth as a problem, and

•imperatives are offered saying how that purpose can be achieved.

So we can give these imperatives the general label 'imperatives of skill'. The practical part of a science is concerned only with •what must be done to achieve a certain purpose; it doesn't address the question of •whether the purpose is reasonable and good. The instructions to a physician for how to make his patient thoroughly healthy, and to a poisoner for how to bring certain death to his victim, are of equal value in that each serves perfectly to achieve the intended purpose. Since in early youth we don't know what purposes we may come to have in the course of our life, parents •try above all to enable their children to learn *many kinds* of things, and •provide for *skill* in the use of means to *any chosen* end. For any given end, the parents can't tell whether it will actually come to be a purpose that their child actually has, but •they have to allow that• some day it *may* do so. They are

so focused on this that they commonly neglect to form and correct their children's judgment about the worthwhileness of the things that they may make their ends.

But there is one end that can be supposed as actual in all rational beings to which imperatives apply, i.e. all rational beings that are *dependent* [see footnote 4 above]; and thus one purpose that they not only *can* have but that we can assume they all *do* have as a matter of natural necessity. This purpose is *happiness*. The hypothetical imperative that declares some action to be practically necessary for the promotion of happiness is an *assertoric* imperative. We should describe it not as

•necessary to a problematic purpose, one that is merely possible,

but as

•necessary to a purpose that we can *a priori* and with assurance assume for each person, because it belongs to his essence.

Skill in the choice of means to one's own greatest welfare can be called 'prudence' in the narrowest sense.<sup>5</sup> Thus the imperative that refers to the choice of means to one's own happiness (i.e. the precept of prudence) is still only *hypothetical*; it commands the action not *outright* but only *as a means to another end*.

•After those two kinds of hypothetical imperative• we come at last to one imperative that commands certain conduct •immediately, and not •through the condition that some purpose can be achieved through it. This imperative is

*categorical*. It isn't concerned with what is to result *from* the conduct, or even with what will happen *in* the conduct (its •matter), but only with the •form and the principle from which the conduct follows. What is essentially good in the conduct consists in the frame of mind—the willingness to obey the imperative—no matter what the upshot is. This may be called 'the imperative of *morality*'.

Volition according to these three principles is plainly distinguished by the dissimilarity in the pressure they put on the will. As an aid to getting this dissimilarity clear, I believe we shall do well to call them, respectively,

*rules* of skill,

*advice* of prudence,

*commands* (laws) of morality.

For it is only *law* that carries with it the concept of a *necessity* ('This action *must* be performed') that is unconditional and objective and hence universally valid; and commands are laws that must be obeyed even when one would prefer not to. *Advice* also involves necessity, but it's a necessity that can hold only under a subjectively contingent condition (i.e. whether this or that man counts this or that as part of his happiness). Whereas the categorical imperative isn't restricted by •or made dependent on• any condition. As absolutely (though practically) necessary, it can be called a 'command' in the strict sense. We could also call the first imperatives 'technical' (relevant to arts and skills), the second 'pragmatic' (relevant to well-being), and the third 'moral' (relevant to any free conduct whatsoever, i.e. to morals).<sup>6</sup>

<sup>5</sup> The word 'prudence' may be taken in two senses, that of (1) 'worldly prudence' and that of 2 'private prudence'. (1) refers to a man's skill in influencing others so as to get them serve his purposes. (2) is the insight to bring all these purposes together to his own long-term advantage. Any value that (1) has ultimately comes from (2); and of someone who is 'prudent' in sense (1) but not in sense (2) we might better say that he is over-all *not* prudent but only clever and cunning.

<sup>6</sup> This seems to me to be the right meaning for the word 'pragmatic'. For constraints are called 'pragmatic' when they don't strictly flow from the law of states as necessary statutes but rather from provision for the general welfare. A history is composed 'pragmatically' when it teaches *prudence*—i.e. instructs the world how it could look after its advantage better (or not worse) than it has in the past.

The question now arises:

- How are all these imperatives possible?

This question doesn't ask, for any kind of imperative,

- How can the action that the imperative commands be performed?

Rather, it asks,

- How are we to understand the constraint that the imperative puts upon the will in setting it its task?

·We shall see that there is not much of a problem about this for the first of the three kinds of imperative, and the same is true—though with slight complications—of the second·.

(1) How an imperative of **skill** is possible requires no particular discussion. If someone wills an end, and if reason has decisive influence on his actions, then he also wills any steps he can take that are indispensably necessary for achieving that end. What this proposition implies about the will is *analytic*, and here is why:

When I will x as to-be-brought-about-by-me, I already have—as a *part* of that act of will·—the thought of the means to x, i.e. the thought of my causality in the production of x. And the imperative extracts from the concept of *willing x* the concept of *actions necessary for the achievement of x*.

(Of course, truths about *what* means are necessary for achieving x are synthetic propositions; but those are only about how to achieve x and not about the act of the will.) ·Here's an example of this interplay between analytic and synthetic propositions·. Mathematics teaches that

- to bisect a line according to an infallible principle, I must make two intersecting arcs from each of its extremities;

and this is certainly a *synthetic* proposition. But if I know that that's the only sure way to bisect the line, the proposition

- if I fully will the effect, I must also will the action necessary to produce it

is *analytic*. For •conceiving of something as an effect that I can somehow bring about is just the same as •conceiving of myself as acting in this way.

(2) If only it were as easy to give a definite concept of happiness, the imperatives of **prudence** would perfectly correspond to those of **skill** and would likewise be analytic. For then we could say that, with prudence as with skill, whoever wills the end wills also (necessarily according to reason) the only means to it that are in his power. Unfortunately, however, the concept of happiness is so indefinite that, although each person wishes to attain it, he can never give a definite and self-consistent account of *what* it is that he wishes and wills ·under the heading of 'wanting happiness'·. The reason for this is that

all the •elements of the concept of happiness are empirical (i.e. must be drawn from experience),

whereas

the •completed· idea of happiness requires ·the thought of· an absolute whole—the thought of a maximum of well-being in my present and in every future condition.

Now it is impossible for a finite being—even one who is extremely clear-sighted and capable—to form a definite ·and detailed· concept of what he really wants here ·on this earth·. ·Consider some of the things people say they aim for·! •Wealth: but in willing to be wealthy a person may bring down on himself much anxiety, envy, and intrigues. •Great knowledge and insight: but that may merely sharpen his eye for the dreadfulness of evils that he can't avoid though he doesn't now see them; or it may show him needs ·that he doesn't know he has, and· that add to the burden his desires already place on him. •Long life: but who can guarantee

him that it wouldn't be a long misery? •Health: but often enough *ill*-health has kept him from dissolute excesses that he would have gone in for if he had been perfectly healthy! In short, he can't come up with any principle that could with complete certainty lay down what would make him truly happy; for that he would need to be omniscient. So in his pursuit of happiness he can't be guided by detailed principles but only by bits of empirical advice (e.g. concerning diet, frugality, courtesy, restraint, etc.) which experience shows to be usually conducive to well-being. It follows from this •that imperatives of prudence can't strictly speaking *command* (i.e. present actions objectively as practically *necessary*); •that they should be understood as advice rather than as commands of reason; •that the problem:

Settle, for sure and universally, what conduct will promote the happiness of a rational being

is completely unsolvable. There couldn't be an imperative that in the strict sense *commanded* us to do what makes for happiness, because happiness is an ideal not of reason but of imagination, depending only on empirical grounds. This means that whether a person will achieve happiness depends on countless many particular facts about his future states; and there is absolutely no chance of picking out the actions that will produce the right infinite totality of consequences that will constitute happiness. If the means to happiness *could* be stated with certainty, this imperative of **prudence** would be an analytic practical proposition, for it would then differ from the imperative of **skill** only in the way described in paragraph (1) above, namely: the imperative of skill is addressed merely to a purpose that a person *may* have, while the purpose of the imperative of prudence—namely happiness—is *given* for every person. That leaves them the same in this respect: each commands the means to something that the person is assumed to have as a *willed* purpose,

so each commands the *willing of the means* to someone who *wills the end*; and so each is analytic. So there is no difficulty about how such an imperative is possible. [Both here, and at the start (2) of the discussion of imperatives of prudence, Kant makes it pretty clear that such imperatives are *not* actually analytic because of the indeterminateness about what happiness amounts to, though they *would be* analytic otherwise. He evidently thinks that if there is only *this* barrier to their being analytic, their status as nearly analytic (so to speak) makes them unproblematic.]

(3) On the other hand, the question of how the imperative of **morality** is possible *does* call for an answer, for this imperative is not hypothetical, and so what it presents as objectively necessary can't be based on any presupposed purpose as in the case of hypothetical imperatives. But don't lose sight of the fact that it can't be shown empirically—can't be shown by *producing an example*—that there *are* any imperatives of morality; perhaps every imperative that seems to be categorical is tacitly hypothetical. For example,

Someone says 'You oughtn't to promise anything deceitfully' and we take this to be categorical; we assume •that an action of this kind must be regarded as in itself bad and thus that the imperative prohibiting it is categorical. (The alternative is to think •that the necessity involved in this prohibition is mere advice about how to avoid something else that is bad, along the lines of 'You oughtn't to promise falsely, in case people find out about it and your credit rating is wrecked'.)

But we can't point with certainty to any example in which the will is directed by the law alone without any other action-drivers, i.e. in which the will obeys a categorical imperative. In a given case this may *appear* to be so, but it's always *possible* that a fear of disgrace and perhaps also a dim sense of other dangers may have had a secret influence

on the will. We can't rule this out on empirical grounds: who can prove by experience that something *doesn't have* a cause of a certain sort when experience can only show us that *we don't perceive* such a cause? In such a case—i.e. when other incentives are secretly affecting the will—the so-called 'moral imperative', which appears to be categorical and unconditional, is in fact only a pragmatic injunction that calls on us to attend to our own advantage.

With each of the other two kinds of imperative, experience shows us that imperatives of the kind in question do exist, and the inquiry into their possibility is the search only for an explanation of them, not for evidence that they exist. It is not so with categorical imperatives. Our investigation of *their* possibility will have to proceed purely *a priori*—starting with no empirical presuppositions, and in particular without the advantage of the premise that such imperatives actually exist. That they *do* exist is one of the things we may hope to establish through our inquiry into their possibility. (In the meantime—though this is an aside—this much at least may be seen: the categorical imperative is the only one that can be taken as a practical *law*, while all other imperatives may be called *principles* of the will [here = 'movers of the will'] but not *laws*. This is because what is merely necessary-for-attaining-some-chosen-end can be regarded as itself *contingent*, as can be seen from the fact that when we give up the end in question we get rid of the instruction stated in the imperative. In contrast with this, an unconditional command leaves the will no freedom to choose the opposite, so that it (and only

it) involves the *necessity* that we require of a law.)

I have spoken of one thing we are *up against* when trying to show the possibility of categorical imperatives, namely that we must do this *a priori*, without being able to appeal to any empirical evidence that such imperatives do actually exist. Now for a second point about getting insight into the possibility of a categorical imperative or law of morality, namely: there's a very solid reason why it will be hard to do this, because this imperative is an *a priori* synthetic practical proposition.<sup>7</sup>

We know already that it is hard to see that theoretical propositions of this sort—i.e. ones that are synthetic and known *a priori*—are possible, so we must be prepared for at least as much difficulty when it comes to practical ones.

In approaching this task, let us first ask:

Doesn't the mere *concept* of a *categorical imperative* provide us with the form of words expressing the proposition—the only kind of proposition—that can be a categorical imperative?

Don't think that answering Yes to this ends our task. For even when we know how the imperative *sounds*—i.e. how it is worded—the question of how such an absolute command *is possible* will require difficult and special labours to answer; I shall get into these in the final chapter.

When I have the general thought of a *hypothetical imperative*, I can't tell just from this thought what such an imperative will contain. To know *that*, I have to know what the condition is. But when I have the thought *categorical*

<sup>7</sup> When I affirm a categorical imperative, I connect the action with the will *a priori*, and hence necessarily, without making this conditional on the person's preferring to achieve this or that end. (Though I do this *objectively*, i.e. under the idea of a reason that has complete control over all its subjective motivators.) So this is a practical proposition that *doesn't* analytically derive the willing of an action from some other volition already presupposed (for we don't have the perfect will that would be needed for there always to be such a volition, namely a volition to obey the moral law.) Rather, the proposition connects the action *directly* with the concept of the will of a rational being as something that isn't contained in it so that the connection isn't analytic.

*imperative*, I know right away what it will contain. For all the imperative contains is

the law, and

the necessity that the maxim conform to the law;

and the law doesn't contain any condition limiting it (-comparable with the condition that is always part of a hypothetical imperative-). So there is nothing left for the maxim to conform to except the universality of a law as such, and what the imperative represents as necessary is just precisely that conformity of maxim to law.<sup>8</sup>

So there is only one categorical imperative, and this is it: 'Act only on that maxim through which you can at the same time will that it should become a universal law.'

Now if all imperatives of duty can be derived from this one imperative as a principle, we'll at least be able to show what we understand by the concept of duty, what the concept means, even if we haven't yet settled whether so-called 'duty' is an empty concept or not.

The universality of law according to which effects occur constitutes what is properly called *nature* in the most general sense, . . . i.e. the existence of things considered as determined by universal laws. So the universal imperative of duty can be expressed as follows: **Act as though the maxim of your action were to become, through your will, a universal law of nature.**

I want now to list some duties, adopting the usual division of them into •duties to ourselves and •duties to others, and into •perfect duties and •imperfect duties.<sup>9</sup>

(1) A man who has been brought by a series of troubles to the point of despair and of weariness with life still has his reason sufficiently to ask himself: 'Wouldn't it be contrary to my duty to myself to take my own life?' Now he asks: 'Could the maxim of my action -in killing myself- become a universal law of nature?' Well, here is his maxim:

For love of myself, I make it my principle to cut my life short when prolonging it threatens to bring more troubles than satisfactions.

So the question is whether *this* principle of self-love could become a universal law of nature. If it did, that would be a nature that had a law according to which a single feeling •created a life-affirming push and also •led to the destruction of life itself; and we can see at a glance that such a 'nature' would contradict itself, and so couldn't *be* a nature. So the maxim we are discussing *couldn't* be a law of nature, and therefore would be utterly in conflict with the supreme principle of duty.

(2) Another man sees himself being driven by need to borrow money. He realizes that no-one will lend to him unless he firmly promises to repay it at a certain time, and he is well aware that he wouldn't be able to keep such a promise. He is disposed to make such a promise, but he has enough conscience to ask himself: 'Isn't it improper and opposed to duty to relieve one's needs in that way?' If he does decide to make the promise, the maxim of his action will run like this:

<sup>8</sup> A *maxim* is a subjective principle of acting, and must be distinguished from the *objective* principle, which is the practical *law*. The maxim contains the practical rule that reason comes up with in conformity with the state the person (the subject) is in, including his preferences, his ignorances, and so on; so it is the principle according to which the subject *acts*. The law, on the other hand, is the objective principle valid for every rational being, and the principle by which the subject *ought to act*; that is, it is an imperative.

<sup>9</sup> Please note that I reserve the •serious, considered• division of duties for a future *metaphysic of morals*, and that the present division is merely one I chose as an aid to arranging my examples. . .

When I think I need money, I will borrow money and promise to repay it, although I know that the repayment won't ever happen.

Here he is—for the rest of this paragraph—reflecting on this: 'It may be that this principle of self-love or of personal advantage would fit nicely into my whole future welfare, so that there is no *prudential* case against it. But the question remains: would it be right? To answer this, I change the demand of self-love into a universal law, and then put the question like this: If my maxim became a universal law, *then* how would things stand? I can see straight off that it could never hold as a universal law of nature, and must contradict itself. For if you take a law saying that anyone who thinks he is in need can make any promises he likes without intending to keep them, and make it *universal* so that everyone in need *does* behave in this way, that would make the promise and the intended purpose of it *impossible*—no-one would believe what was promised to him but would only laugh at any such performance as a vain pretence.'

(3) A third finds in himself a talent that could be developed so as to make him in many respects a useful person. But he finds himself in comfortable circumstances, and would rather indulge in pleasure than take the trouble to broaden and improve his fortunate natural gifts. But now he asks whether his maxim of neglecting his gifts, agreeing as it does with his liking for idle amusement, also agrees with what is called 'duty'. He sees that a system of nature conforming with this law could indeed *exist*, with everyone behaving like the Islanders of the south Pacific, letting their talents rust and devoting their lives merely to idleness, indulgence, and baby-making—in short, to *pleasure*. But he can't possibly **will** that this should become a universal law of nature or that it should be implanted in us by a natural instinct. For,

as a rational being, he necessarily wills that all his abilities should be developed, because they serve him and are given to him for all sorts of possible purposes.

(4) A fourth man, for whom things are going well, sees that others (whom he could help) have to struggle with great hardships, and he thinks to himself:

What concern of mine is it? Let each one be as happy as heaven wills, or as he can make himself; I won't take anything from him or even envy him; but I have no desire to contribute to his welfare or help him in time of need.

If such a way of thinking were a universal law of nature, the human race could certainly survive—and no doubt *that* state of humanity would be better than one where everyone chatters about sympathy and benevolence and exerts himself occasionally to practice them, while also taking every chance he can to cheat, and to betray or otherwise violate people's rights. But although it is possible that that maxim should *be* a universal law of nature, it is impossible to **will** that it do so. For a will that brought *that* about would conflict with itself, since instances can often arise in which the person in question would need the love and sympathy of others, and he would have no hope of getting the help he desires, being robbed of it by this law of nature springing from his own will.

Those are a few of the many duties that we have (or at least think we have) that can clearly be derived from the single principle that I have stated on the preceding page. We must *be able to will* that a maxim of our action become a universal law; this is the general formula for the moral evaluation of our action. •Some actions are so constituted that their maxim can't even be *thought* as a universal law of nature without contradiction, let alone being *willed* to be such. It's easy to see that an action of that kind conflicts with stricter or narrower (absolutely obligatory) duty. •With

other actions, the maxim-made-universal-law is not in that way internally impossible (·self-contradictory·), but it is still something that no-one could possibly *will* to be a universal law of nature, because such a will would contradict itself. It's easy to see that an action of *that* kind conflicts with broader (meritorious) duty. Thinking of duties in terms not of the *object* of their action but rather of the *kind of obligation* they involve, what I have given is a complete display of all the kinds of duty, in terms of their dependence on a single principle.

If we attend to what happens in us when we act against duty, we find that we don't (because we *can't*) actually will that our maxim should become a universal law. Rather, we are willing that the opposite of the maxim on which we are acting should remain as a law generally, but we take the liberty of catering to our preferences by making an exception—'just for me, just this once!'. So if we weighed everything from a single standpoint, namely that of *reason*, we would find a contradiction in our own will: willing that a certain principle •be objectively necessary as a universal law and yet •subjectively not hold universally but rather admit of exceptions. However, ·we *don't* consider our actions in this unitary way; rather·, we regard our action ·at one time· from the point of view of a will wholly conformable to reason and then ·at another time· from the point of view of a will affected by preferences; so there is actually no contradiction, but rather the preference's *resisting* the command of reason. In this the •universality of the principle is changed into mere •generality—i.e. the move is made from *all* to *ever so many* or *almost all*—so that the practical principle of reason meets the maxim half-way. This procedure, whether or not it can be justified in our own impartial judgment, shows that we really do acknowledge the validity of the categorical imperative and allow ourselves (while keeping a wary eye on it) only a

few exceptions—ones that strike us as unimportant and as forced on us.

I have thus at least shown that if *duty* is a concept that is to have significance and actual law-giving authority for our actions, it has to be expressed in categorical imperatives, never in hypothetical ones. And along with that I have made clear—and ready for any use—the *content* that the categorical imperative must have if it is to contain the principle of all duty (if there is such a thing as duty). This is a substantial result; but I haven't yet reached the point where I can prove *a priori* that •this kind of imperative really exists, that •there is a practical law that of itself commands absolutely and without any action-drivers, and that •obedience to this law is duty.

If we want to reach that point, it is extremely important that we pay heed to this warning:

Don't slip into thinking that the reality of this principle can be derived from the *special constitution of human nature!*

For duty has to be practical-and-unconditional necessity of action; so it has to hold for all rational beings (the only beings to which an imperative has anything to say), and is a law for all human wills *only because they are rational beings*. In contrast with that, anything that is derived from

- the temperament of human beings in particular, from
- certain feelings and propensities ·of human beings·,
- or even from (if this is possible)
- a particular tendency of the human *reason* that might not hold for the will of every rational being,

—such a thing can yield a •*maxim* that is valid for us, but not a •*law*. That is, it can yield •a subjective principle on which we might act if our desires and dispositions take us that way, but not •an objective principle telling us how to act even if all our dispositions, preferences, and natural tendencies

were pulling us in the opposite direction. Indeed, the •fewer subjective causes there are *for* acting in a certain way and •the more there are *against*, the more clearly we can see the sublimity and intrinsic dignity of duty's command to act in that way. The pulls in the other direction don't weaken the constraint of the law or lessen its validity.

Here we see philosophy put into a precarious position, which has to be made firm even though there is nothing in heaven or on earth to hang it from or stand it on! Here philosophy has to show its purity as the •sustainer of its own laws, and not as the •herald of laws that are whispered to it by an implanted 'sense' or by who knows what guardian 'nature'! 'Laws' of the latter kind may always be better than nothing, but they can't yield fundamental principles. Such principles can only be dictated by *reason*: they must have an entirely *a priori* origin, getting none of their commanding authority from the preferences of mankind and all of it from the supremacy of the law and due respect for it. Otherwise—that is, if human nature were the only basis for morality—mankind would be condemned to self-contempt and inner disgust.

Thus if anything empirical were brought in as an ingredient in the principle of morality, it would not only •be utterly useless in this role but would also •do terrific harm to the purity of morality •in practice—for in morals the proper, priceless value of an absolutely good will consists precisely in action's being driven by something that is *free from* all influences from contingent grounds that only experience can make available. We can't too strongly or too often warn against this slack—indeed this *low*—cast of mind, that looks for its principles [here = 'the sources of moral energy'] among

empirical motives and laws. •The warning is constantly and urgently needed, because reason in its weariness is glad to rest on this pillow and dreamily. . . . substitute for •morality a botched-up •bastard assembled from limbs of very different species—it looks like anything you want to see in it, but not like virtue to anyone who has ever beheld her in her true form.<sup>10</sup>

So this is our question:

Is it a necessary law *for all rational beings* that they should always judge their actions by maxims that they themselves could will to hold good as universal laws?

If there is a such a law, it must already be connected—wholly *a priori*—with the concept of *the will of a rational being*. But in order to discover this connection, we must, however reluctantly, take a step into metaphysics; but it will be into the metaphysic of morals, not the region of metaphysics involved in •speculative philosophy. [The 'speculative'/'practical' contrast is explained on pages 3–4.] A •practical philosophy doesn't commit itself to explanations of *what happens* but to laws about what *ought to happen* even if it never does—i.e. objective-practical laws. [Kant means that they are 'objective' at least in the sense that they are not 'subjective', i.e. don't have anything to say about what the person, the *subject*, wants or hopes or prefers or fears or aims at. For another example of this contrast at work, see pages 7–8.] In practical philosophy, therefore, we needn't inquire into

- why something pleases or displeases,
- how merely sensory pleasure differs from taste,
- whether taste is different from a general satisfaction of reason,
- what the feelings of pleasure and displeasure depend on,

<sup>10</sup> To behold virtue in its proper form is simply to present morality with •nothing sensuous stirred into the mixture and •every spurious adornment of reward or self-love stripped off. Viewed in that way, it outshines everything that appears charming to the senses, as can easily be seen by anyone whose reason hasn't been spoiled for all abstraction.

- how such feelings give rise to desires and inclinations.
- how desires and preferences, with the co-operation of reason, give rise to maxims.

All of that belongs ·not to practical philosophy but· to empirical psychology. (If we think of *natural science* as *the philosophy of nature based on empirical laws*, then empirical psychology is the second part of it, ·empirical physics being the first·.) In contrast with that, our present concern is with objectively-practical laws and thus with how a will relates to itself when it determines itself only by reason, and ·in that inquiry· every empirical consideration automatically falls away. Why? Because if *unaided reason* determines conduct, it must necessarily do so *a priori*, ·and thus without bringing in anything empirical·. *Can* reason determine conduct in this way? That is what we are now to investigate.

The will is thought of as someone's capacity ·or ability· to control how he behaves *in conformity with the representation of certain laws*. [Slightly correcting what Kant wrote, which literally means: 'The will is thought of as a capacity to control itself. . .'] Such a capacity can be found only in rational beings. Now, what serves the will as the objective ground for its action upon itself is an *end*, and if it is given by reason alone it must be an end for all rational beings. On the other hand, what contains the ground of the possibility of the action that leads to the end is called the *means*. The subjective ground of desire is the *action-driver*, while the objective ground of volition is the  *motive*. And so we have a distinction between

- subjective ends resting on action-drivers, and
- objective ends depending on motives that are valid for every rational being.

Practical principles are *formal* when they abstract from all subjective ends; they are *material* when they are based on subjective ends and thus on certain action-drivers. All of the ends—*material* ends—that a rational being voluntarily

sets before himself as things to be achieved through his conduct are merely •relative, for their value comes solely from how they •relate to the particular way in which the subject's faculty of desire is constituted; and from this we can't get any practical *laws*, i.e. any universal and necessary principles that hold for all rational beings and for every act of the will. So the only imperatives that these relative ends support are hypothetical ones.

But suppose there were something *whose existence in itself* had absolute value, something which *as an end in itself* could support determinate laws. *That* would be a basis—indeed the *only* basis—for a possible categorical imperative, i.e. of a practical law.

·*There is* such a thing! It is a human being!· I maintain that man—and in general every rational being—exists as an end in himself and *not merely as a means* to be used by this or that will at its discretion. Whenever he acts in ways directed towards himself or towards other rational beings, ·a person serves as a *means* to whatever end his action aims at; but· he must always be regarded as *also an end*. Things that are preferred have only *conditional* value, for if the preferences (and the needs arising from them) didn't exist, their object would be worthless. ·That wouldn't count against the 'objects' in question if the desires on which they depend did themselves have unconditional value, but they don't!· If the preferences themselves, as the sources of needs, did have absolute value, one would want to have them; but that is so far from the case that every rational being must wish he were altogether free of them. So the value of any objects *to be obtained* through our actions is always conditional. Beings whose existence depends not on our will but on nature, if they are not rational beings, have only relative value as means, and are therefore called 'things' [*Sachen*]; whereas rational beings are called '*persons*', because their nature

already marks them out as ends in themselves (i.e. as not to be used merely as means)—which makes such a being •an object of respect, and •something that sets limits to what anyone can choose to do. Such beings are not merely subjective ends whose existence as a result of our action has value *for us*, but are objective ends, i.e. things [*Dinge*] whose existence is an end in itself. It is indeed an *irreplaceable* end: you can't substitute for it something else to which it would be merely a means. If there were no such ends in themselves, nothing of absolute value could be found, and if all value were conditional and thus contingent, no supreme practical principle for reason could be found anywhere.

So if there is to be a supreme practical principle, and a categorical imperative for the human will, it must be an objective principle of the will that can serve as a universal law. Why must it? Because it has to be drawn from the conception of something that is an end in itself and therefore an end for everyone. The basis for this principle is: *rational nature exists as an end in itself*. Human beings necessarily think of their own existence in this way, which means that the principle holds as a subjective principle of human actions. But every other rational being also thinks of his existence on the same rational ground that holds also for myself;<sup>11</sup> and so it is at the same time an *objective* principle—one that doesn't depend on contingent facts about this or that *subject*—a supreme practical ground from which it must be possible to derive all the laws of the will. So here is the practical imperative: **Act in such a way as to treat humanity, whether in your own person or in that of anyone else, always as an end and never merely as a means.** Let us now see whether this can be carried out.

To return to our previous examples: **(1)** Someone thinking of committing suicide will, if he is guided by the concept of necessary duty to oneself, ask himself

•Could my suicide be reconciled with the idea of humanity as *an end in itself*?

•And his answer to this should be No. If he escapes from his burdensome situation by destroying himself, he is using a person merely as *a means* to keeping himself in a tolerable condition up to the end of his life. But a man is not a *thing* [*Sache*], so he isn't something to be used *merely* as a means, and must always be regarded in all his actions as an end in himself. So I can't dispose of a man by maiming, damaging or killing him—and that includes the case where the man is myself. (This basic principle needs to be refined so as to deal properly with questions such as 'May I have one of my limbs amputated to save my life?' and 'May I expose my life to danger in order to save it?' I shan't go into these matters here; they belong to *morals* and not to the metaphysic of morals.)

**(2)** [Three times in this next paragraph, and nowhere else in this work, Kant writes of someone's 'containing' the end of an action by someone else. Presumably for *B* to 'contain' the end of *A*'s action is for *B* to have *A*'s end as *his* end also, to seek what *A* seeks.] As concerns necessary. . . duties to others, when someone *A* has it in mind to make someone else *B* a deceitful promise, he sees immediately that he intends to use *B* merely as a means, without *B*'s containing in himself the *end* of the action. For *B* can't possibly assent to *A*'s acting against him in this way, so he can't contain in himself the end of this action. This conflict with the principle about treating others as ends is even easier to see in examples of attacks on people's freedom

<sup>11</sup> Here I put this proposition forward as a postulate. The reasons for it will be given in the last chapter.

and property; for in those cases it's obvious that someone who violates the rights of men intends to make use of the person of others merely as means, without considering that as rational beings they should always be valued at the same time as ends, i.e. as beings who can contain in themselves the end of the very same action.<sup>12</sup>

(3) With regard to contingent (meritorious) duty to oneself [for 'meritorious' see page 26], it isn't sufficient that the action *not conflict* with humanity in our person as an end in itself; it must also *harmonize* with it. In human nature there are predispositions to greater perfection that are part of nature's purpose for humanity. . . .; to neglect these might perhaps be consistent with the *preservation* of humanity as an end in itself but not with the *furtherance* of that end.

(4) With regard to meritorious duty to others: Humanity might survive even if

- no-one contributed to the happiness of others, but also
- no-one intentionally took anything away from the happiness of others;

•and this is a likely enough state of affairs, because• the end or purpose that all men *naturally* have is *their own* happiness. This would put human conduct into harmony with humanity as an end in itself, but only in a *negative* manner. For a *positive* harmony with humanity as an end in itself, what is required is that everyone •positively• tries to further the ends of others as far as he can. For the ends of any person, who is an end in himself, must as far as possible be also *my* ends, if that thought •of him as an end in himself•

is to have its *full* effect on me.

This principle concerning the status of each human being—and more generally of each rational creature—as an end in himself is the supreme limiting condition on the freedom of action of each man. (*Supreme* in the sense that it trumps everything else, e.g. prudential considerations.) It isn't drawn from experience; •there are two reasons why it *can't* be•. •One reason is the principle's universality: it applies to *absolutely all rational beings*, and experience doesn't stretch out that far. •The other is the fact that the principle isn't about humanity considered subjectively, as something that men *do* take to be an end, i.e. do choose to aim at, but rather about humanity considered as the objective end that *ought* to constitute the supreme limiting condition of all subjective ends, whatever they may be. •Experience can inform us about what subjective ends men *do* set before themselves, but not about what non-subjective end *ought* to trump every subjective end•. So this principle •can't arise from experience, and• must arise from pure reason.

According to the first principle, the •objective basis for all practical legislation lies in •the rule and the form of universality, which makes it capable of being a natural law [bold type on page 24]. Its •subjective basis is •the end; and according to the second principle the subject of all ends is every rational being as an end in itself. From this we now derive the third practical principle of the will, as the supreme condition of its harmony with universal practical reason, namely, **the idea of the will of every rational being as a will laying down universal law.**

<sup>12</sup> Don't think that the banal 'Don't do to anyone else what you wouldn't want done to you' could serve here as a guide or principle. It is only a consequence of the real principle, and a restricted and limited consequence at that. It can't •as it stands• be a universal law, because it doesn't provide a basis for •duties to oneself, or •benevolent duties to others (for many a man would gladly consent to not receiving benefits from others if that would let him off from showing benevolence to them!), or •duties to mete out just punishments to others (for the criminal would argue on this ground against the judge who sentences him). And so on.

By this ·third· principle, any maxim is rejected if it isn't consistent with the will's role as a giver of universal law. Hence the will is not merely subject to the law, but subject to it in such a way that it must be viewed as

- prescribing the law to itself,  
and for just that reason as

- being subject to the law,  
the law of which it sees itself as the author.

I have presented two kinds of ·categorical· imperative: one kind tells us to act in a manner that is lawful, like the lawfulness of the natural order; the other lays down that rational beings are in themselves supreme *ends*; and just because both of these are categorical, their commanding authority owes nothing to any action-driver involving one's *interests*. But so far I have been *assuming* them to be categorical—an assumption I had to make if I was to explain the concept of duty. But *are there* any such imperatives, practical propositions that command categorically? Back there I couldn't prove independently that there are, any more than I can prove it in this present chapter. But there's something that I *could* have done—namely to point out an inherent feature of an imperative that specifically marks it off as categorical rather than hypothetical. The feature I have in mind is the renunciation of all one's *interests* when one wills from duty. And now we have an example of this in the formulation of the principle ·of morality· that is now before us, the third, which involves the idea of the will of every rational being as a *will that gives ·or legislates· universal law*.

A will that is *subject to laws* can be bound to them by some interest that it has; but a will that *is itself the supreme law-giver* can't depend upon any interest for this role. Why

can't it? Because if it did, it would need *another* law saying that its interests could be satisfied only if the first law were universally valid; ·in which case the first law wouldn't be supreme, after all·.

Thus the principle of every human will as *a will giving universal law in all its maxims*,<sup>13</sup> provided it is otherwise correct, is very well suited to being a categorical imperative because of this feature: it involves the idea of giving universal law, so it isn't based on any interest, and thus it is the only possible imperative that can be unconditional. . . .

Look back on all the previous attempts to discover the principle of morality—no wonder they all failed! The searchers saw that

- man is bound by his duty to laws;  
but it didn't occur to them that
- all man is subject to are laws—universal laws—  
legislated by himself,  
and that
- all he bound to is to act in accordance with his own  
will,

a will designed by nature to be a giver of universal law. The thought of him only as *subject to some law or other* brings with it the need for some *interest* that will pull or push him to obey the law—his will has to be constrained to act thus and so by *something else*—because the law hasn't arisen from *his* will. This strictly valid inference means that all the work of looking for a supreme ground for duty was wasted labour; it never brought them to *duty* but only to the necessity for acting from a certain *interest*. It might be the person's own interest or someone else's; either way, the imperative always had to be conditional, and couldn't serve as a moral command. I shall call this principle— ·the third of

<sup>13</sup> I needn't clarify this principle with fresh examples, because the ones I have already used to illustrate the categorical imperative and its formulation can here serve the same purpose.

my three—the principle of **autonomy** of the will in contrast with every others, which I accordingly count as **heteronomy**. [From Greek: *auto/hetero* = self/other, and *nomos* = law. So Kant's terminology distinguishes *self-governed* from *other-governed*.]

The concept of •every rational being as one who must regard himself as giving universal law through all the maxims of its will, so as to judge himself and his actions from this standpoint, leads to the fruitful concept of •a realm of ends. [The German *Reich* mainly means 'kingdom' or 'empire', but the less highly charged 'realm' seems to fit well enough here.]

By 'realm' I understand the systematic union of different rational beings through shared laws. (The next sentence presents a thought-experiment; in conducting it, we have to abstract from personal differences of rational beings, and thus from all content of their private ends or purposes.) Because laws determine which ends have universal validity, we can think of a •unified• *whole* of all ends in systematic connection—a whole composed of •rational beings who *are* ends in themselves and of •ends that they may individually set for themselves. This is a realm of ends, which is possible on the principles stated above.

That is because all rational beings stand under the law that each of them should treat himself and all others *never merely as a means* but always *also as an end in himself*. This gives rise to a systematic union of rational beings through shared objective laws, i.e. a *realm*; and it may be called a realm of ends because what these laws have as their purpose is just the relation of these beings to each other as ends and means. (Admittedly this realm of ends is only an ideal.)

A rational being is a *member* of the realm of ends if he *gives* universal laws in it while also being *subject to* those laws. He is *sovereign* in the realm of ends if, as law-giving, he isn't subject to anyone else's will. A rational being must always regard himself as law-giving in a realm of ends that

is possible only through the freedom of the will, and this holds whether he belongs to the realm as a member or as sovereign. Being sovereign in the realm of ends isn't a matter of *choice*; to be sovereign a rational being must •be completely independent •of everything else•, •have no needs, and •have unlimited power adequate to his will.

So the morality of any action is constituted by how the action relates to the law-giving that is indispensable if there is to be a realm of ends. But this law-giving must be found in every rational being, being able to arise from his will. So the principle that drives his will is:

•never to act on a maxim that couldn't consistently be a universal law,

and thus

•to act only so that the will could regard itself as giving universal law through its maxim.

In the case of a rational being whose maxims don't by their nature already necessarily conform to this objective principle, the necessity of acting according to that principle is called practical compulsion •or constraint•, i.e. *duty*. The *sovereign* in the realm of ends doesn't have duties; all the mere *members* have duties, and are indeed burdened by duty to the same extent.

[Of the two versions of the end of this next sentence, the first fits Kant's thought better: the only things he has called ends in themselves are rational beings. The second doctrinally drops out of the blue. But it is what Kant wrote; to get the first reading we must replace Kant's *sie* by *sich*.] The •practical necessity of acting in accordance with this principle, i.e. •duty, doesn't rest at all on feelings, impulses, and preferences; its sole basis is the way rational beings relate to one another—a relationship in which the will of a rational being must always be regarded as law-giving,

**plausible reading:** otherwise it couldn't think of **itself** as an end in itself.

**literal reading:** otherwise it couldn't think of **duty** as an end in itself.

Reason accordingly checks out every maxim of your will, in its role as giver of laws, to see how it relates to everyone else's will and also to every action towards yourself. It doesn't do this from any external practical motive or future advantage, but rather from the idea of the *dignity* of a rational being who obeys no law except one that he himself gives while obeying it.

In the realm of ends everything has either a *price* or an *intrinsic value*. Anything with a price can be replaced by something else as its equivalent, whereas anything that is above all price and therefore admits of no equivalent has intrinsic value. [In this paragraph, 'intrinsic value' translates *Würde*, which is usually translated—here and in the Kantian literature generally—by 'dignity'. At the end of the next paragraph Kant explicitly equates those two meanings.]

Something that involves general human desires and needs has a *market price*. Something that doesn't involve anyone's *needing* anything but accords with a certain taste (i.e. with pleasure in the purposeless play of our feelings) has a *luxury price* [*Affektionspreis* = 'price related to the feelings']. But if something makes it possible—and is the *only* thing that makes it possible—for something to be an end in itself, then it doesn't have mere relative value (a *price*) but has intrinsic value (i.e. dignity). [Kant wrote: *hat... einen innern Wert, d. i. Würde.*]

Now, it is only through morality that

- a rational being can be a law-giving member in the realm of ends;

so it is only through morality that

- a rational being can be an end in himself.

So morality, and humanity so far as it is capable of morality, are the only things that have dignity. Skill and diligence in work have a •market price; wit, lively imagination, and humour have a •luxury price; but fidelity in promises and benevolence on principle (not benevolence from instinct) have •intrinsic value which I have called *dignity*. If you don't have these, neither nature nor art can supply anything that would make up for that lack in you; for their value doesn't lie in the effects that flow from them—their usefulness, the advantages they bring—but only in the attitudes, i.e. the maxims of the will, that are ready to express themselves in this manner through actions, even if the actions don't meet with success. For us to look on these actions with immediate favour and pleasure, we don't have to bring in any of our subjective states, any immediate *liking* for or attraction to such actions. The actions exhibit the will that generates them as the object of an immediate respect, since nothing but reason is required to get the will to act like that. (Note that reason *imposes* these actions on the will; it doesn't *coax* it into performing them, for that would flatly contradict the notion of duty.) This esteem lets the value of such a turn of mind be recognized as *dignity* or *intrinsic value*, and puts it infinitely above any *price*; to compare it with, or weigh it against, things that have price would be to violate its holiness, as it were.

And what is it, then, that justifies virtue, or a morally good frame of mind, in making such lofty claims for itself? It is its enabling the rational being to have a *share* in the giving of universal laws and thus to become fit to be a member in a possible realm of ends. (His nature has already marked him out for this role, as an end in himself and therefore as a law-giver in the realm of ends.) . . . For a rational being has no value except what the law confers on it. The law-giving that confers all value must therefore have *dignity* (i.e. an

unconditional and incomparable value); and the •esteem that a rational being must have for this is best described as ‘respect’ [*Achtung*; some of Kant’s uses of this suggest that ‘reverence’ would be a better translation]. Autonomy is thus the basis for the dignity of human nature and of every rational nature.

I remind you that I have presented the principle of morality in three ways:

- Act as though the maxim of your action were to become, through your will, a universal law of nature. [page 24]
- Act in such a way as to treat humanity, whether in your own person or in that of anyone else, always as an end and never merely as a means. [page 29]
- Act only so that your will could regard itself as giving universal law through its maxim. [page 32]

The above three ways of presenting the principle of morality are basically only so many different formulations of the very same law, and any two of them come together in the third. They *do* differ in a certain way, but the difference is subjective rather than objective—i.e. the three formulations don’t express different moral principles, but they offer different ways for our minds to come at morality. This difference is intended to introduce a certain analogy that will bring an idea of reason closer to intuition and thus nearer to feeling. All maxims have:

- (1) **A form**; specifically, they are all universal. That leads to this formulation of the moral imperative: maxims must be chosen as if they were to hold as universal laws of nature.
- (2) **A matter or content**, i.e. an end. That leads to this formulation: all merely relative ends, ones that people

*choose*, must be restricted by and subordinated to the status of rational beings, which are *not chosen* as ends but are ends by their very nature, and are therefore ends in themselves.

- (3) **A complete fixing** of all maxims through this formulation: all the maxims that come from your own law-giving should harmonize with a possible realm of ends as with a realm of nature.<sup>14</sup>

Moving through these three items is like moving through the Categories of Quantity, as set forth in my *Critique of Pure Reason*:

- the •Unity of the form of the will (its universality),
- the •Plurality of the matter (the objects, ends), and
- the •Totality of the system of ends.

In arriving at moral judgments one does better to go by just one of the three formulations, specifically to follow the strict method and base one’s thinking on the universal formulation of the categorical imperative: *Act in accordance with a maxim that can at the same time make itself a universal law*. But if one wants to enable the moral law to have *access* to a mind, it is very useful to bring one and the same action under the three concepts I have listed, and thus, so far as possible, to bring it nearer to intuition.

We can now end where we started, with the concept of an unconditionally good will. A will is absolutely good if it *can’t* be bad, and thus never adopts maxims that conflict with themselves when they are generalized into universal laws. So this principle is also its supreme law: Always act on maxims whose universality as laws you can at the same time will. That’s the only way a will can avoid ever coming into conflict with itself, and such an imperative is categorical.

<sup>14</sup> •Teleology considers nature as a realm of ends; •morals regards a possible realm of ends as a realm of nature. In the •former the realm of ends is a *theoretical* idea for explaining what exists. In the •latter it is a *practical* idea for bringing about something that doesn’t yet exist but will become real through our conduct, in conformity with this very idea.

Because the •validity of the will as a universal law for possible actions has an analogy with the •way existing things are inter-connected under universal laws, this being the formal aspect of nature in general, the categorical imperative can be put like this: •Act on maxims that can at the same time have themselves as universal laws of nature as their object•. That gives us the formula for an absolutely good will.

Rational nature is distinguished from the rest of nature in that it sets for itself an end. This end would be the material of every good will, but •its relation to the *absolutely* good will involves a little wrinkle which I shall now explain•. The idea of an absolutely good will doesn't have anything to do with this or that end that the will seeks to •bring about; the ends that a will aims to •bring about can only make it relatively good, not absolutely good; so we must understand the end of an absolutely good will

not as *an end to be* •brought about but as *an independently existing end*

which connects with the absolutely good will *negatively*—it is

an end which must never be acted against,

which implies that

it must never, in any act of the will, be valued merely as a means but always also as an end.

...The principle: •Act in relation to every •rational being (whether yourself or another) so that in your maxim •he is an end in himself• is thus basically identical with the principle: *Act on a maxim that involves its own universal validity for every rational being*. That's because the statement 'In my use of means to any end I should restrict my maxim to the condition of its universal validity as a law for every subject' is equivalent to the statement 'The subject of ends (i.e. the rational being itself) must be made the basis of every maxim of action and thus be treated never as a mere means but as

the supreme limiting condition in the use of all means—i.e. also as an end.

It •follows from this—no-one could question that it follows—that every rational being, as an end in himself, must be able to regard himself as a *giver* of universal laws that include any laws to which he may be *subject*. For what marks him off as an end in himself is just this fitness of his maxims for universal law-giving. It also •follows that this dignity that he has, his prerogative over all merely natural beings, involves his having to take his maxims from the point of view of himself and every other rational being as law-givers—which is why they are called 'persons'. In this way, a world of rational beings. . . is possible as a realm of ends, through the law-giving activities of all the persons who are its members. Consequently every rational being must act as if his maxims made him at all times a law-giving member of the universal realm of ends. The formal principle of these maxims is: Act as though your maxims were also to serve as universal law for all rational beings. A realm of ends is thus possible only by analogy with a realm of nature. The realm of ends is possible only through

•maxims, i.e. rules imposed on *oneself*,

while the realm of nature is possible only through

•laws governing how things are acted on by *other things*.

Despite this difference, nature as a whole, though looked on as a machine, is given the name '*realm of nature*'—to the extent that, and because, it has reference to rational beings as its ends. [This is one of the places where 'kingdom' might be better than 'realm'.] Such a realm of ends would actually come into existence through maxims whose rule is prescribed to all rational beings by the categorical imperative, *if every rational being followed them all the time*. A rational being who scrupulously follows this maxim can't expect every other

rational being to follow suit; nor can he expect the realm of nature. . . .to favour his expectation of happiness. Despite that, the law:

Act in accordance with the maxims of a universal-law-giving member of a merely possible realm of ends remains in full force because it commands categorically. And just here lies the two-part paradox: (1) the will is subject to an inflexible rule concerning the place of *humanity* in its deliberations, simply because of the *dignity* of humanity as rational nature without any end or advantage to be gained by being human, and thus out of respect for a mere idea; and (2) what makes some maxims sublime, and makes every rational subject worthy to be a law-giver in the realm of ends, is just precisely this independence of his maxims from all such action-drivers as chosen ends and possible advantages. If it weren't for this independence, the rational being would have to be seen as subject only to the natural law of his needs. Even if the realm of nature and the realm of ends were thought of as united under one sovereign, so that the realm of ends moved from being a mere idea to becoming a reality and gained reinforcement from a strong action-driver, still there would be no increase in its intrinsic value. For when we think about this possibility of a world in which a unique sovereign brings it about that principled actions do always lead to good consequences, we have to think of the sole absolute law-giver as judging the value of rational beings only on the strength of the disinterested conduct that they prescribe to themselves merely from the idea. The essence of things isn't changed by their external relations, and the absolute [= 'non-relational'] value of a man doesn't involve his relations to other things either; so whoever is estimating a man's absolute worth must set aside his external-relational properties—and this holds for *anyone* doing such an estimation, even the supreme being. Morality

is thus the relation of actions not to anything external to the person, but to the autonomy of the will. . . . Now here are definitions, in terms of the autonomy of the will, of five key terms in morality. •An action that can co-exist with the autonomy of the will is *permitted*. •One that clashes with autonomy of the will is *forbidden*. •A will whose maxims are necessarily in harmony with the laws of autonomy is a *holy* or *absolutely good* will. •If a will is not absolutely good, it is morally constrained by the principle of autonomy and its relation to that principle is *obligation* (so a holy will can't have obligations). •The objective necessity of an action from obligation is called *duty*.

From what I have been saying, it is easy to understand how this happens: although in thinking of duty we think of *subjection* to law, we nevertheless also ascribe a certain *sublimity* and *dignity* to the person who fulfils all his duties. There is nothing sublime about being *subject to* the moral law, but this person is also a *giver of* the law—that's why he is subject to it, and only to that extent is he sublime. Also, I have shown above how the only action-driver that can give an action moral value is *respect for the law*, not any kind of fear or desire. The proper object of respect is our own will to the extent that it tries to act only on maxims that could contribute to a system of universal legislation (such a will is ideally possible for us), and the dignity of humanity consists just in its capacity to *give* universal laws to which it is also *subject*.

### **The autonomy of the will as the supreme principle of morality**

A will's autonomy is that property of it by which it is a law to itself, independently of any property of the objects of its volition. So the principle of autonomy is:

Always choose in such a way that the maxims of your choice are incorporated as universal law in the same volition.

That this practical rule is an imperative, i.e. that the will of every rational being is necessarily bound to it as a constraint, can't be proved by a mere analysis of the concepts occurring in it, because it is a synthetic proposition. This synthetic proposition presents a command, and presents it *as necessary*; so it must be able to be known *a priori*. To prove it, then, we would have to go beyond knowledge of objects to a critical examination of the subject (i.e. to a critique of pure practical reason). But that is not the business of the present chapter. But mere analysis of moral concepts *can* show something to our present purpose, namely that the principle of autonomy that we are discussing is the sole principle of morals. This is easy to show, because conceptual analysis shows us that morality's principle must be a categorical imperative and that the imperative in question commands neither more nor less than this very autonomy. [See note on page 32 for 'autonomy' and 'heteronomy'.]

### **The heteronomy of the will as the source of all spurious principles of morality**

A will is looking for a law that will tell it what to do: if it looks anywhere except in the fitness of its maxims to be given as universal law, going outside itself and looking for the law in the property of any of its objects, *heteronomy* always results. For in that case the law is not something the will gives to itself, but rather something that the external object gives to the will through its relation to it. This relation, whether it rests on preference or on conceptions of reason, admits of only *hypothetical* imperatives: 'I should do  $x$  because I want  $y$ '. The moral or categorical imperative, on the other hand,

says that 'I should do  $x$  whether or not I want anything else. For example, the hypothetical says that 'I shouldn't lie if I want to keep my reputation. The categorical says that 'I shouldn't lie even if lying wouldn't bring the slightest harm to me. So the categorical imperative must abstract from every object thoroughly enough so that no object has any *influence* on the will; so that practical reason (the will), rather than catering to interests that are not its own, shows its commanding authority as supreme law-giving. Thus, for instance, I ought to try to further the happiness of others, but not in the spirit of 'it matters to me that these people should be happier, because...' with the blank filled by a reference to some *preference* of mine, whether directly for the happiness of the people in question or indirectly via some satisfaction that is related to their happiness through reason. Rather, I should try to further the happiness of others solely because a maxim that excludes this can't be included as a universal law in one and the same volition.

### **Classification of all possible principles of morality that you'll get if you take heteronomy as the basic concept**

Why the interest in *all possible* principles that come from this underlying mistake? Because in the absence of a critical examination of the pure use of reason, human reason always—including here—tries *every possible* wrong way before it succeeds in finding the one true way!

If you start with the idea of heteronomy—i.e. of how the will can be directed *from outside itself*—you will be led to principles of one of two kinds: empirical and rational. **(1)** The empirical ones have to do with *happiness*, and are based on the thought of the will as being influenced by either **(1a)** physical feelings concerning one's own happiness or **(1b)** moral feelings. **(2)** The rational ones have to do with

*perfection* and are based on the thought of the will as being influenced by either **(2a)** the rational concept of perfection as a possible result of our activities or **(2b)** the concept of an independently existing perfection (the will of God). You can see that all four of these have the will being influenced from outside itself. Let us now look into them in detail.

**(1)** Empirical principles are not at all fit to serve as the basis of moral laws. For moral laws should be *universal*, valid for all rational beings without distinction, that being what makes them *unconditionally* practically necessary; but this universality is lost if moral laws are derived from the specific constitution of human beings—a constitution that may not be shared by other rational beings—or the particular circumstance in which human beings happen to live. **(1a)** But the principle of *one's own happiness* is the most objectionable of the empirical bases for morality. There are at least three reasons for this, of which the third is the weightiest. •This basis for morality is just false: experience contradicts the allegation that well-being is always proportional to good conduct. •The principle contributes nothing to the establishment of morality, because making a man happy is very different from making him good, and making him prudent and sharp in seeing what is to his own advantage is far from making him virtuous. •Above all: this principle supports morality with action-drivers that undermine it and destroy all its sublimity, for it puts the motives to virtue and those to vice in the same class, obliterating the difference of *kind* between them, and teaching us merely to make a better job of calculating what will make us happy. **(1b)** Now for the supposed special sense, moral feeling.<sup>15</sup> There are endless differences in degree between different kinds of feeling, so

that feelings can't give us a uniform standard of good and bad; and anyway one can't validly judge for *others* by means of *one's own* feeling. So the appeal to moral feelings is superficial. Those who believe that feelings can help them to grasp universal laws are people who can't *think!* Despite all this, the moral-feeling approach is nearer to morality and the *dignity* of morality, because it honours virtue by ascribing *immediately to her* the satisfaction and esteem we have for her, and does not, as it were, **(1a)** tell her to her face that what attaches us to her is not her beauty but only our advantage! [The use of 'her' to refer to virtue is based not on Kant's using a personal pronoun (German doesn't have them) but just on the content of the metaphor that he uses.]

**(2)** Among the rational principles of morality—the ones based on reason—there is **(2a)** the ontological concept of perfection. It is empty, indefinite, and consequently useless for finding in the immeasurable field of possible reality the greatest possible sum of perfections that is suitable to us. Also, when we try to say what marks off *this* reality—perfection—from all other realities, we inevitably tend to move in a circle and can't avoid tacitly presupposing the morality that we are trying to explain. Nevertheless, this is better than **(2b)** the theological concept, which derives morality from a most perfect divine will. There are two reasons for the inferiority of the theological concept; or, more accurately, they are two halves of a single reason which constitutes a *dilemma* confronting the theological approach to morality. The perfection of the divine will is not something that is *given* to us in intuition analogous to how items are given to us through the senses; so we have to derive it from our own concepts. •Foremost among these is our concept

<sup>15</sup> I bring *moral feeling* under the heading of *happiness* because every empirical interest promises to contribute to our well-being by the agreeableness that a thing affords, either **(1a)** indirectly, through the thing's contributing to our happiness, or **(1b)** directly, through our finding the thing itself agreeable without any thought of our own future advantage. . . .

of morality; if we let this generate our concept of God's perfection, and then use the latter as a basis for morality, we are guilty of a flagrantly circular explanation. •And if we don't get at God's perfection in that way, our only remaining concept of it is made up of

the attributes of desire for glory and dominion,  
combined with  
the awe-inspiring conceptions of power and vengefulness;

and any system of ethics based on these would be directly opposed to morality.

The **(1b)** concept of the moral sense and **(2a)** that of perfection in general have this to be said for them, that they don't weaken morality; but neither is capable of serving as its foundation. Still, if I had to choose between them, I would opt for **(2a)** perfection in general, because it takes the decision about the basis for morality away from the realm of sensibility and submits it to the court of pure reason. It doesn't get a *decision* there, but at least it preserves the indefinite idea of a will that is good in itself, without falsifying it—saving a place for it until it can be more narrowly defined.

You won't mind if I don't grind through a long refutation of all these doctrines. There's no need for me to do that, because it is so easy to do and so well understood—even by those whose official positions require them to declare for one of these theories because their hearers wouldn't tolerate suspension of judgment. What matters more to us here is to know this: All these principles try to base morality purely on heteronomy of the will, so they are bound to fail.

[An addition to this paragraph is marked by \* rather than ·, because of its length.] Whenever an *object* of the will has to be laid down as prescribing the rule that is to tell the will what to do, the rule is none other than heteronomy. In such a case

the imperative is conditional—

*If or because* you want such and such an object, you ought to act thus and so

—so it can't command morally, i.e. categorically. The object's influence on what my will does may go through my *preference* (as in the principle of my own happiness) or through my *reason* directed to objects of my possible volitions (as in the principle of perfection); but the will in these cases never determines itself *directly* by the conception of the action itself. It is always directed by an object *through something other than the will*, namely through the action-driver that is stirred up in the will by the prospect of getting a certain result:

I ought to do *x* because I will *y*;

and then another law must be planted in me, a law saying that *I must will y*; and this law in its turn would require an imperative to restrict this maxim—\*i.e. an imperative of the form

I ought to will *y if I want z*.

Why? Because if instead we had simply

I ought to will *y*,

that involves no appeal to anything outside the will; it is a categorical imperative, and doesn't involve heteronomy of the will. But that puts it outside the scope of the present discussion, which is of the consequences of trying to base morality on heteronomy, i.e. on the influence on the will of factors outside it. Relying on heteronomy has one bad consequence that I haven't yet mentioned\*. With a hypothetical imperative such as we get with heteronomy of the will, the aim is for •the thought of a result to be obtained by one's own powers to stir up in the will •an impulse of a certain kind (the thought of achieving *y* is to stir up an impulse to do *x*); but whether and how that thought generates that impulse depends on the natural constitution of the person

concerned—i.e. depends either on his sensibility (preference and taste) or on his understanding and reason. Now, what the person's sensibility or intellect makes of any intended upshot—e.g. whether it takes pleasure in it—depends on the details of what kind of sensibility or intellect nature has endowed the person with; which implies that strictly speaking the source of this law is *nature*. As a law of nature, this would have to be known and proved by experience, which means that it would be contingent and therefore unfit to be a necessary practical rule such as the moral rule must be. This is still *heteronomy* of the will: the law is given to the will not by •the will itself but by •an impulse from outside it, an impulse that influences the will because the person's nature makes him susceptible to it.

So an absolutely good will, the principle of which must be a categorical imperative, doesn't specify any object, and contains only *the form of volition as such*, and this form is *autonomy*. That is,

- the sole law that the will of every rational being imposes on itself

is just

- the fitness of the maxims of every good will to turn themselves into universal laws;

and there is no need for *this* to be supported by any action-driver associated with an interest.

*How can there be such a synthetic practical a priori proposition, and why it is necessary?* The solution of that problem doesn't lie within the boundaries of the metaphysic of morals; and I haven't here affirmed its truth, let alone claimed to have a proof of it in my power. All I have done is to show, by spelling out the generally accepted concept of morality, that an autonomy of the will is unavoidably connected with morality—is indeed its foundation. So anyone who holds that morality is *something* and not *a chimerical idea without truth* must accept, along with morality, the principle that I have derived here. Consequently, this chapter like the first was merely analytic; it reached its conclusions by analysing, spelling out the content of, the generally accepted concept of morality. If the categorical imperative, and with it the autonomy of the will, is true and absolutely necessary as an *a priori* principle, it follows that morality isn't a phantom; but to *prove* that it isn't we must be able to make **a synthetic use of pure practical reason**. But we mustn't venture on this use without first making a *critique* of this faculty of reason. In this next chapter—the last—I shall give the chief features of such a critique, in enough detail for our purpose.