

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.¹⁵ 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

¹⁵ 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.

Chapter 3:

Moving from the metaphysic of morals to the critique of pure practical reason

The concept of freedom is the key to explaining the autonomy of the will

Will is a kind of causality that living beings exert if they are rational, and when the will can be effective independent of outside causes acting on it, that would involve this causality's property of *freedom*; just as *natural necessity* is the property of the causality of all non-rational beings, through which they are caused to act in specific ways by the influence of outside causes.

The account of freedom I have just given is *negative* (·it says there is freedom when the active will does *not* have external causes acting on it·), and so it isn't fruitful for insight into what freedom *is*; but there flows from it a concept of freedom that is *positive*, and accordingly richer and more fruitful. Although freedom is not a property of the will according to •laws of nature, it doesn't follow that freedom is lawless! It must in fact be a causality according to immutable •laws *of a special kind*. The 'concept' of lawless free-will would be an absurdity, because the concept of causality brings with it the concept of laws according to which if something we call a cause is given then something else, the effect, must occur. ·And since

freedom conceptually involves *causality*, and

causality conceptually involves *law*,

it follows that

freedom conceptually involves *law*

· We saw that· natural necessity is a •**heteronomy** of effective causes, because each effect can come about only through a law according to which •**something else** gets the cause to

exercise its causality. What can the freedom of the will be, then, but •**autonomy**, i.e. the will's property of •**itself** being a law? However, the proposition:

•The will itself is a law in all its actions

only expresses the principle:

•Act only on a maxim that can also have *itself as a universal law* for its object.

And this is just the formula of the categorical imperative and is the principle of morality. Therefore a •free will and a •will under moral laws are identical. [Earlier in this paragraph Kant said that all causality involves an *effect's* being caused by something else. What he is now treating as the mark not of causality as such but of heteronomous causality in particular is *the cause's* being stirred into action by something else. In the next paragraph, 'something else' occurs again in a manner that does *seem* to conflict with what Kant first said about causality in general.]

So if we start with freedom of the will, we get morality (together with its principle) from it merely by •analysing its concept. But the principle of morality:

An absolutely good will is one whose maxim can always include itself regarded as a universal law,

is a •synthetic proposition, because that property of the maxim can't be found by analysis of the concept of an absolutely good will. What makes such a synthetic proposition possible is there being two cognitions [= 'items of knowledge'] that are connected with each other through their both being contained in some third cognition. In the case of physical causes, the 'third cognition' that ties the cause to the effect is the •nature of the sensible world; but the concept of *that* conjoins the two concepts of something as cause in relation

to *something else* as effect; so it doesn't meet our present needs. In our present context, the 'third cognition' that does the job is the •positive concept of freedom. Two tasks present themselves: (1) To show what this third cognition is to which freedom directs us and of which we have an *a priori* idea, and (2) To make comprehensible the deduction of the concept of freedom from pure practical reason, and along with that the possibility of a categorical imperative. But I can't do either of these right here and now; first, some further preparation is needed.

Freedom must be presupposed as a property of the will of all rational beings

It isn't enough to ascribe freedom to •our will, on whatever grounds, if we don't also have sufficient grounds for attributing it to •all rational beings. For morality serves as a law for us only because we are rational beings, so it must hold for all rational beings; and morality must be derived solely from the property of freedom; so freedom must be shown to be a property of the will of all rational beings. And it doesn't suffice to do this on the basis of certain supposed empirical facts concerning human nature: we need an *a priori* proof (•which empirical facts can't provide•), and we need a result concerning absolutely all rational beings endowed with a will (•and not merely concerning human beings•). Now I say this:

Any being who can't act otherwise than *under the idea of freedom* is, just for that reason, really free in his conduct—i.e. all laws that are inseparably bound up

with freedom hold for him just as if his will were validly pronounced free in itself as a matter of theoretical philosophy.¹⁶

Now I maintain this:

We must necessarily equip every rational being who has a will with the idea of freedom, this being an idea under which he must act.

For the thought of such a being includes the thought of a reason that is practical, i.e. has causality with respect to its object. Now we can't conceive of a reason that would consciously take direction (about how to judge) from outside itself, for then the person •whose reason it was• would think that what settled how he judged was not his reason but some •external• impulse. Reason must regard itself as the author of its principles, owing nothing to external influences; so it must—as practical reason, or as the will of a rational being—regard itself as free. That is to say, the will of a rational being can be *his* will only under the idea of freedom, so that from a practical point of view such a will must be ascribed to all rational beings.

Why should I be moral?

We have finally traced the determinate concept of morality back to the idea of freedom, but we couldn't prove freedom to be actual in ourselves and in human nature. We saw only that we must *presuppose* it if we want to think of a being as rational and as conscious of himself as the cause of his own actions, i.e. as endowed with a will; and so we find that on just those same grounds we must ascribe to each being

¹⁶ I start in a way that is sufficient for my purposes, with freedom as something posited by all rational beings merely *in idea* as the basis for their actions. I go about things in this way so as to avoid having to prove freedom also in its theoretical respect. Even if the latter—•actual, factual, theoretical freedom•—is left unproved, •it makes no difference, because• the laws that would hold for a being who was really free hold also for a being who cannot act except under the idea of his own freedom. Thus we escape the burden •of proof• that the theoretical assertion of freedom would impose upon us.

endowed with reason and will this property of settling for himself how he will act, doing this under the idea of freedom.

From the presupposition of this idea of freedom there flowed also the consciousness of a law of action:

The subjective principles of actions (i.e. maxims) must always be adopted in such a way that they can hold also as objective, i.e. hold as universal principles, thereby serving as universal laws that we give to ourselves.

But why ought I, just because I am a rational being, subject myself to this law? And why should all other beings endowed with reason do so? Admittedly no interest *impels* me to do this, for that wouldn't yield a categorical imperative; but I must still *take* an interest in it, and have insight into how it comes about. For this 'ought' is really a 'shall' that holds for every rational being whose reason isn't hindered in its generating of actions. For beings like ourselves, that necessity of action is expressed only as 'ought', and the subjective necessity is thus distinguished from the objective. By 'beings like ourselves' I mean ones who are affected not only by reason but also by action-drivers that come from the senses—beings who don't always do what reason would have done if left to itself.

So it seems that all we have done with respect to the moral law (i.e. the principle of the autonomy of the will) is to *presuppose* it in the idea of freedom, as though we couldn't independently prove its reality and objective necessity. Even that would bring some gain, because in doing it we would at least have defined the genuine principle more accurately than had been done before; but we wouldn't have made any progress regarding the validity of the moral law or the practical necessity of subjecting ourselves to it. If anyone asked:

- Why do our actions have to be based on maxims that could be universally valid as laws?
- What is the basis for the value that we ascribe to this way of acting—a value so great that no interest, anywhere, can outweigh it?
- How does it come about that a man believes that it is only through *this* that he feels his own personal value, in contrast to which that of a pleasant or unpleasant state is to be regarded as nothing?

we couldn't give him any satisfactory answer.

We do indeed find that we can take an interest in a personal quality that makes us fit to enjoy some condition if reason were to allot that condition to us, even though the personal quality doesn't automatically bring the condition with it. An example might be: taking an interest in *being the sort of person who would be a good spouse*, this being distinct from taking an interest in being married. That is, we can take an interest in *being worthy of happiness* without having *being happy* as a motive. But this judgment about value—this taking of an interest—is in fact only the effect of the importance we have already ascribed to the moral law (when through the idea of freedom we detach ourselves from every empirical interest). But we are confronted by the proposition that

We *ought* to detach ourselves from every empirical interest, to regard ourselves as free in acting and yet as subject to certain laws, in order to find right there within ourselves a value that would compensate for the loss of everything that could make our situation desirable.

How *this* is possible, and hence *on what grounds the moral law is binding*, can't be grasped through my procedure up to here.

It has to be admitted that there is a kind of circle here from which there seems to be no escape. We take ourselves to be free in the order of effective causes so that we can think of ourselves as subject to moral laws in the order of ends; and then we think of ourselves as subject to these laws because we have ascribed to ourselves freedom of the will. This is circular because •freedom and •self-legislation by the will are both •autonomy; so they are equivalent concepts, which is why neither of them can be used to explain or support the other. At most they can be used for the logical purpose of bringing apparently different representations of the same object under a single concept (like reducing the both of the fractions $\frac{51}{68}$ and $\frac{69}{92}$ to $\frac{3}{4}$).

There remains open to us, however, one way out of the circle, namely, to pursue an inquiry into these:

- thinking of ourselves as causes that are effective *a priori* through freedom, and
- thinking of ourselves in terms of our actions considered as empirically observable effects.

•For example: •thinking of myself as voluntarily raising my arm, and •thinking of myself as seeing my arm go up. The question is: Don't these involve different standpoints?

[In this next paragraph and later on, Kant refers to a certain 'world' that he calls *Verstandeswelt* = 'world of understanding'. Evidently this 'world' has at least as much to do with •reason as with •understanding, and we shall see on the next page that Kant distinguishes these from one another. So 'world of understanding' would be a misleading label, and 'intelligible world' is used instead. In some places where *Verstandeswelt* would be especially misleading, Kant instead uses *intelligibel Welt*.] What I am about to say requires no subtle reflection, and presumably even the most ordinary intellect could arrive at it (doing so in its own way, through an obscure exercise of judgment that it calls 'feeling'!). All mental representations

that come to us involuntarily (as do those of the senses) enable us to know objects only •as they affect us, which leaves us still ignorant of what they are •in themselves; and therefore representations of this kind, however closely and sharply we attend to them, can give us only knowledge of *appearances*, never knowledge of *things in themselves*. This distinction may be made just by noticing the difference between •representations that we passively receive from somewhere else and •ones that we actively produce out of ourselves. Once the distinction has been made somehow, it automatically follows that we must admit and assume behind the appearances something else that is not appearance, namely things in themselves. But we have to accept that we can't get any closer to them, and can't ever know what they are in themselves, because all we can know of them is how they affect us. This must yield a distinction, though a rough one, between

- a *sensible world*, which can be very different to various observers, because of differences in their sensibilities, and
- an *intelligible world*, which underlies the sensible world, and remains always the same.

A man shouldn't claim to know even *himself* as he really is by knowing himself through inner sensation—i.e. by introspection. For since he doesn't produce himself (so to speak) or get his concept of himself *a priori* but only empirically, it is natural that he gets his knowledge of himself through *inner sense* and consequently only through how his nature *appears* and how his consciousness is *affected*. But beyond the character of his own subject, which is made up out of these mere appearances, he necessarily assumes something else underlying it, namely his *I* as it is in itself. Thus in respect to mere perception and receptivity to sensations he must count himself as belonging to the sensible world;

but in respect to whatever pure activity there may be in himself (which reaches his consciousness •directly and not •by affecting the ·inner or outer· senses) he must count himself as belonging to the intellectual world—though he doesn't know anything more about it.

A thoughtful person must come to some such conclusion as this about all the things that present themselves to him. Even someone with a very ordinary mind is likely to have this thought, because we know that such people are strongly inclined to expect something invisibly active at work behind the objects of the senses; but they don't learn anything from this because they soon spoil it by trying to make the 'invisible something' perceptible, i.e. to make it an object of intuition. ·They do this by wondering 'What is it *like*, this unknown something that lurks behind the appearances of things?', when their only concept of what a thing can be *like* is made from the concept of how things *appear* to them·.

Now a human being really finds within himself a capacity by which he distinguishes

•himself from •all other things,

and that includes distinguishing

•himself ·as something active· from •himself considered as affected by objects.

This capacity or faculty is *reason*. Its pure, spontaneously active nature puts reason on a higher level even than understanding, ·and here is why·. Understanding ·is like reason in this: it· is spontaneously active, and does not—like the faculty of *sense*—merely contain representations that come from our being passively affected by things. But ·it is unlike reason in that· the only concepts it can produce through its activity are ones whose only role is to *bring the representations of sense* under rules. . . . ·The intellectual management of the data of the senses is the understanding's *only* task·. Without this use of sensibility the understanding

wouldn't have any thoughts at all. In contrast with this, reason shows in its ideas, as we call them (·ideas relating to reason as concepts do to the understanding·), a spontaneity so pure that it goes far beyond anything that sensibility can come up with. The highest occupation of reason is to distinguish the sensible world from the intellectual world, thereby marking out limits for the understanding itself.

Because of this, a rational being must regard himself—in his role *as an intelligence*, setting aside his lower faculties—as belonging not to the sensible world but to the intelligible world. So he has two standpoints from which he can •consider himself and •recognize the laws for the use of his powers and hence for all his actions. **(1)** As belonging to the sensible world, ·he falls· under the laws of nature (heteronomy). **(2)** As belonging to the intelligible world, ·he is· under ·the moral authority of· laws that are independent of nature, and so are not empirical but based entirely on reason.

As a rational being and thus as belonging to the intelligible world, a human being can never think of the causality of his own will except under the idea of freedom; because reason must always take itself to be independent of the determining causes of the sensible world, and that independence is what freedom *is*. Now we have the idea of freedom inseparably connected with the concept of autonomy, which is bound up with the universal principle of morality, which is ideally the ground of all actions of rational beings, just as the law of nature is the ground of all appearances.

That allays the suspicion (the one that I stirred up earlier) that there might be a hidden circle in our reasoning from •freedom to •autonomy and from •that to •the moral law—that we might have laid down the idea of freedom for the sake of the moral law so that we could later derive the law from freedom! That would have made us unable to give

any basis for the law. . . . But now we see that when we think of ourselves as free, we carry ourselves into the intelligible world as members of it and recognize •the autonomy of the will and •the morality that autonomy brings with it; whereas when we think of ourselves as under an obligation, we regard ourselves as belonging to the sensible world and at the same time also to the intelligible world.

How is a categorical imperative possible?

A rational being counts himself, as an intelligence, as belonging to the intelligible world, and only as an effective cause belonging to *this* world does he call his causality a 'will'. On the other side, though, he is conscious of himself as a bit of the sensible world in which his actions are encountered as mere *appearances* of that causality •of his will•. But we aren't acquainted with that causality •of his will•, so there's no way we can grasp how these actions can arise from it; and so they must instead be regarded as caused by other appearances, namely, desires and preferences belonging to the sensible world. Considered only as a member of the intelligible world, my behaviour would completely accord with the principle of the autonomy of the pure will; considered as a bit of the sensible world, my behaviour would have to be assumed to conform wholly to the natural law of desires and preferences and thus to the heteronomy of nature. (The former behaviour would rest on the supreme principle of morality, and the latter on that of happiness.) But *the intelligible world contains the ground or basis of the sensible world and therefore of its laws*, and so the intelligible world is (and must be conceived as) directly law-giving for my will, which belongs wholly to the intelligible world. Therefore I see myself, in my status as an intelligence, •as subject to the law of the intelligible world, i.e. the law of reason which

is contained in the idea of freedom, and •as subject to the autonomy of the will. Therefore the laws of the intelligible world must be regarded as imperatives for me, and actions that conform to them must be regarded as duties. All this holds, despite the fact that on the other side I am a being that belongs to the sensible world.

So this is how categorical imperatives are possible: The idea of freedom makes me a member of an intelligible world; if I were a member *only* of that world, all my actions *would* always conform to the autonomy of the will; but since I confront myself also as a member of the world of sense, my actions *ought* to conform to it. This categorical 'ought' presents *a priori* a synthetic proposition. •It is synthetic because •in it• (1) my will affected by my sensuous desires has *added to it* the idea of (2) something that reason says contains its supreme condition, namely •that very same will considered as pure, self-sufficiently practical, and belonging to the intelligible world. •It is a genuine addition; there's no way you could extract (2) from (1) by sheer analysis•. [Kant adds a not very helpful comparison of this with his doctrine, expounded in his *Critique of Pure Reason*, about *a priori* synthetic propositions that are essential to our knowledge of any system of nature.]

The practical application of common-sense confirms the correctness of this deduction. When we present examples of honesty of purpose, of steadfastness in following good maxims, and of sympathy and general benevolence (even with great sacrifices of advantage and comfort), there is no man, not even the most malicious villain (provided he is otherwise accustomed to using his reason), who doesn't wish that he also might have these qualities. It's merely because of his preferences and impulses that he can't make himself be like this; but he would like to be free of the burden of such preferences. He thus shows himself as having a thought in which he, with a will free from all impulses of sensibility,

transfers himself into an order of things altogether different from that of his desires in the field of sensibility. In this thought he doesn't look for any gratification of desires or any state of affairs that would satisfy any desire that he has or can imagine having; for if *that* were his aim, the very idea that elicits this wish from him would lose its pre-eminence. All he can be looking for is a greater intrinsic value as a person. He believes himself to be this better person when he shifts himself to the standpoint of a member of the intelligible world, to which he is automatically taken by the idea of freedom (i.e. of not being acted on by causes in the sensible world). And in this standpoint he is conscious of a •good will which on his own confession constitutes the law for his •bad will as a member of the sensible world. He recognizes the status of the law even while he breaks it. The moral 'ought' is therefore his own necessary will as a member of the intelligible world, and it is conceived by him as an 'ought' only because he regards himself at the same time as a member of the sensible world.

Concerning the outermost boundary of all practical philosophy

All human beings think of themselves as having free will. That is the source of all judgments that acts that *weren't* performed *ought to have been* performed. But this freedom isn't •something of which we have• an experiential concept; it *can't* be, because even when

experience shows the opposite of things that are represented as •necessary on the supposition of freedom, freedom still remains, •which shows that it can't be defeated by facts of experience because it isn't in the same arena, so to speak, as they are•. On the other hand it is equally

•necessary that everything that happens should be inexorably caused in accordance with natural laws; and this *natural necessity* [a single word in German] is also not •something of which we have• an experiential concept, because it brings with it the concept of necessity and thus of *something that can be known a priori*, •i.e. without consulting experience•. But this concept of a •system of• nature is confirmed by experience, and it has to be presupposed if experience is to be possible—experience being knowledge of the objects of the senses interconnected by universal laws. So •freedom is only an *idea of reason*, whose objective reality in itself is doubtful, whereas •nature is a *concept of the understanding*, which does and necessarily must exhibit its reality in examples drawn from experience.

From this there arises a dialectic of reason—a seeming conflict of reason with itself—because the freedom ascribed to the will seems to contradict natural necessity, •and reason finds itself drawn to each side of the apparent conflict• at this parting of the ways. •For speculative purposes •such as the pursuit of scientific theories•, reason finds the road of •natural necessity more well-trodden and usable than that of freedom. •But for practical purposes—thinking about *what to do and what not to do*—the only way of bringing reason to bear is along the path of •freedom; which is why even the subtlest philosophy can't argue freedom away, any more than the most ordinary common-sense can. So philosophy has to assume that no real contradiction will be found between freedom and natural necessity as applied to the very same human actions, for it can't give up the concept of nature any more than it can that of freedom.

We'll never be able to grasp *how* freedom is possible, but in the meantime we should at least eradicate in a convincing way this apparent contradiction. For if the very thought of freedom contradicted itself or contradicted nature (which is

equally necessary), freedom would have to be surrendered in favour of natural necessity.

But this contradiction couldn't be escaped if the subject who seems to himself to be free were thinking of himself *in the same sense* or *in the same relationship* when he •calls himself free as when he •takes himself to be subject to natural law with respect to the very same action. So speculative philosophy can't be excused from its task of showing at least this much: •that •the ways of talking that produce the illusion of contradiction come from our thinking of the person in a different sense and relationship when we call him free from that in which we consider him as a part of nature and subject to its laws; and •that these two •standpoints not only *can* very well coexist but *must* be thought of as necessarily united in one and the same subject. If this much is not shown, we are left with no basis for burdening reason with an idea •as troublesome as that of freedom—an idea which, though it can *without contradiction* be united with the well-established concept •of natural necessity, nevertheless entangles us in troubles that sorely embarrass reason in its theoretical use. It is only theoretical philosophy that has this duty; its purpose is to clear the way for practical philosophy. So it isn't up to the philosopher to *decide* whether to remove the apparent contradiction or rather to leave it untouched; for if he doesn't remove it, the theory about it would be a no-man's-land which the fatalist would be entitled to take over, as a squatter, driving all morality out.

But we still haven't reached the boundary of •practical philosophy. For the settling of the controversy •over freedom doesn't belong to •it. The situation is just that practical philosophy demands that theoretical reason put an end to the discord in which •it entangles itself in theoretical questions, so that practical reason may have peace and security from outward attacks that could put into dispute

the land on which it wants to build.

The common-sense claim to have freedom of the will is based on the person's consciousness of something that has also been conceded as a presupposition, namely that

- reason is independent of causes that determine a person's psychological state—causes that are all of the sort that sensation can inform us about, and that can be brought under the general name 'sensibility'.

A human being, who in this way regards himself as an intelligence,

- when he thinks of himself as an intelligence with a will, and consequently with causality,

puts himself in a different order of things and in a relationship to determining grounds of an altogether different kind from what comes into play

- when he perceives himself as a phenomenon in the world of sense (as he really is also), and subjects *his* causality to external causal factors according to natural laws.

Now he soon realizes that both can exist together—indeed, that they *must*. For there is not the slightest contradiction between **(1)** *a thing in appearance* (belonging to the sensible world) being subject to certain laws from which as **(2)** *a thing in itself* it is independent. That he must think of himself in this twofold way rests on **(1)** his consciousness of himself as an object affected through the senses, and **(2)** his consciousness of himself as an intelligence (i.e. as independent of sensible impressions in the use of reason), and thus as belonging to the intelligible world.

[The indented portion of this paragraph is a first-person rendition of something that Kant writes using 'he' rather than 'I'.] That's how it comes about that a human being •claims to have a will that doesn't make him accountable for what belongs only to his desires and preferences, but •thinks of this same

will as making possible—indeed necessary—actions that he can perform only by disregarding all his desires and sensuous attractions. The causality of these actions lies in •him as an intelligence and in •an intelligible world's principles concerning effects and actions. All he knows about this intelligible world is this:

In this sensible world the law is given only by reason, and indeed pure reason independent of sensibility. Moreover, since it is only as an •intelligence that I am a *genuine self* (as a •human being I am only an *appearance of myself*), those laws apply to me immediately and categorically; so that nothing that I am pushed into doing by preferences or impulses—thus, nothing caused by the sensible world—can count against the laws of my volition as an intelligence.

Indeed, he doesn't hold himself responsible for those •preferences and impulses• or attribute them to his *genuine self* (i.e. to his will); though when he •allows them to influence his maxims in ways that go against the rational laws of his will, he holds his will to account for •that.

By *thinking itself* into an intelligible world, practical reason doesn't at all step across that world's boundaries, but it would do so if it tried to *see* or *feel* its way into it. The •thought of the• intelligible world is only a negative thought with respect to the sensible world; it doesn't give reason any laws for determining the will. The only positive thing about it is this:

Freedom as a negative determination—•i.e. as something that involves **not** being interfered with by sensible causes•—is also connected with a **positive** power and even a causality of reason, a causality that we call a 'will'.

...But if practical reason were to borrow an *object of the will* (i.e. a motive) from the intelligible world, it would be

overstepping its boundaries and pretending to be acquainted with something of which it knows nothing. So the concept of an intelligible world is only a *standpoint* that reason sees itself as having to take, outside appearances, *in order to think of itself as practical*. Reason *couldn't* be practical if the influences of sensibility settled how the human being behaved, but it *must* be practical unless his consciousness of himself as an intelligence, and thus as a rational and rationally active cause (i.e. a cause acting in freedom), is to be contradicted. This thought certainly brings with it the idea of an order and a law-giving different from that of the mechanism of nature, which has to do with the sensible world; and it necessitates the concept of an intelligible world, i.e. the totality of rational beings as things in themselves; but without the slightest pretence to have any thoughts about it that go beyond its formal condition—i.e. the universality of the maxim of the will as law, and thus the will's autonomy which is required for its freedom. All laws that are fixed on an *object* make for heteronomy, which belongs only to natural laws and can apply only to the sensible world.

We can explain things only by bringing them under laws governing things that *could* be confronted in experience. But freedom is only an idea •of reason•; there is no way its objective reality could be shown through natural laws or, therefore, through any experience. Because it can't be illustrated even in an analogical way with examples, we can't ever *grasp* it or even *see into it* a little. It holds only as a necessary presupposition of reason in a being who believes himself conscious of having

a will, i.e. a faculty or capacity different from that of mere *desire*—a capacity to get himself to act as an intelligence, and thus to act according to laws of reason and independently of natural instincts.

But when we come to an end of causation according to

natural laws, we are at an end of all *explanation*, and all that is left for us to do is to *defend*—i.e. to refute objections from those who purport to have seen more deeply into the essence of things and who boldly declare freedom to be impossible. We can only show them that the supposed contradiction they have discovered in the idea of freedom lies simply in this:

They have to regard a human being as •appearance in order to bring natural laws to bear on his actions; and now when we require them to think of him-as-intelligence as a •thing in itself, they still persist regarding considering him as appearance.

Separating his causality (his will) from all natural laws of the sensible world *does* indeed involve a contradiction if this is the very same subject that we previously brought under natural laws; but the contradiction will disappear if they will think again, and admit that behind appearances things in themselves must stand as their hidden ground, and that we can't insist that the laws of operation of these grounds must be the same as those that govern their appearances.

The subjective impossibility of *explaining* the freedom of the will is the same as the impossibility of discovering and making graspable an interest which a human being can take in moral laws.¹⁷ Yet he *does* actually take an interest in them, and our name for the foundation of this is 'moral feeling'. Some have wrongly offered this moral feeling as our standard for moral judgment, whereas really it should be seen as the *subjective* effect that the law has on the will; the *objective* grounds •for moral judgment• come •not from

feeling but• from reason.

If a sensuously affected rational being is to will an action that reason alone prescribes as what he ought to do, reason must of course be able to instil a feeling of pleasure or satisfaction in the fulfilment of duty, and hence must have a causal power to affect sensibility in accordance with its own principles. But it is wholly impossible to conceive *a priori* how a mere thought with nothing sensuous in it produces a sensation of pleasure or unpleasure. For that is one particular kind of causality which, like every kind of causality, we can learn about only by consulting experience, not *a priori*. But •we can't understand *this* causality through experience either, because• we can do that only for cause-effect pairs where *both* items are objects of experience, whereas here the effect •does lie within experience but the cause—namely, reason acting through mere ideas, which furnish no object for experience—•does not. So it is completely impossible for us human beings to explain how and why we have an interest in *the universality of the maxim as law* and thus an interest in morality. Only this much is certain: **(1)** It is *not* the case that the law holds for us because we have an interest in it (for that would be heteronomy, making practical reason depend on sensibility in the form of an underlying *feeling*, which could never yield a moral law); and **(2)** It is the case that we have an interest in the moral law because it holds for us as human beings, because it has arisen from our will as intelligence, and hence from our genuine self. That source for the moral law is what gives it its authority, what

¹⁷ It is by *interest* that reason becomes practical, i.e. becomes a cause acting on the will. That is why it is only of a being with reason that we say 'He takes an interest in' something; non-rational creatures don't have interests—only sensuous impulses. Reason takes an •immediate interest in actions only in cases where what has moved the will in that direction is the universal validity of the action's maxim. That's the only kind of interest that is pure [= 'non-empirical']. In contrast with that, reason takes an indirect or •mediated interest in an action if it acts on the will only through the intervention or •mediation of another object of desire or under the supposition of some particular feeling that the subject has; and since such objects of desire and particular feelings can't be found out by reason itself, unaided by experience, this mediated kind of interest is only *empirical* and not a *pure* interest of reason. . . .

makes it hold for us, because **reason necessarily makes what belongs to mere appearance subordinate to the character of the thing in itself.**

So the question *How is a categorical imperative possible?* can be answered to this extent: We can cite the only presupposition under which it is possible, namely the idea of freedom; and we can have insight into the necessity of this presupposition. That is all we need for the *practical use* of reason (i.e. to be convinced of the categorical imperative's validity and hence also of the moral law). But how this presupposition itself is possible can never be grasped by any human reason. However, . . . the presupposition of the freedom of the will is quite possible, as speculative philosophy can prove, for it doesn't involve itself in a contradiction with the principle that natural necessity interconnects all the appearances in the sensible world. More than that, it is unconditionally necessary for any rational being. I mean that it is *practically* necessary for him, meaning that he needs it for his consciousness of his causality through reason, . . . needs the idea of it as the fundamental condition of all his voluntary acts. But the question still stands:

How can pure reason, all by itself without any outside help from other action-drivers, be practical? *How* can the mere principle of the universal validity of its maxims as laws. . . create, unaided, an action-driver and produce an interest that would be called 'purely moral'? In short: **How can pure reason be practical?**

All human reason is wholly incompetent to explain this, and it is a waste of trouble and labour to try.

It is just the same as if I tried to find out how freedom itself as the causality of a will is possible, for in making that attempt I would be leaving the philosophical basis of explanation behind, and I have no other. I *would* still have the intelligible world, the world of intelligences, and I could

drift around in that; but it couldn't supply the desired explanation, because although I have a well-founded *idea* of that world I don't have the least *knowledge* of it—and I can't have such knowledge, however hard I exercise my natural faculty of reason. This intelligible world signifies only a *something*—a whatever-it-is—that is left after I have excluded from the factors acting on my will everything belonging to the sensible world, which I did merely so as to shut the principle of motives out of field of sensibility. I did this by limiting this field and showing that it doesn't contain absolutely everything, and that outside it there is still more; but that's all I know about this 'more', namely that it lies outside the sensible world. It is pure reason that has this idea, that is the thought of this ideal entity, the intelligible world; it has been deprived of all *matter* (i.e. all knowledge of objects); so all that I am left with in trying to make sense of pure reason is

- the *form*, namely the practical law of the universal validity of maxims, and
- the possible role of pure reason as an effective cause acting on the will in accordance with that form.

There is no room here for any external action-driver. If we insist on there being one, then the action-driver—i.e. that in which reason directly takes an interest—would have to be this idea of an intelligible world. But to understand how *this* could drive action is precisely the problem we can't solve.

Here, then, is the outermost boundary of all moral inquiry [assuming that Kant wrote *oberste* = 'highest' when he meant to write *äußerste* = 'outermost', as in the heading on page 47]. It's very important to locate it accurately, because if we don't, either of two disasters may occur. On the one hand, reason may search for the supreme moral motive in the sensible world, in a way harmful to morals On the other hand, reason may impotently flap its wings in the space—so far as reason

is concerned it's an *empty* space!—of. . . the intelligible world, without being able to move from its starting point and so losing itself among phantoms. For the rest, the *idea* of a pure intelligible world, as a whole of all intelligences to which we ourselves belong as rational beings (though on the other side we also belong to the sensible world), is always a useful and permissible idea for the purpose of a rational •belief, even though all •knowledge terminates at that world's boundary. Its service is that of awakening in us a lively interest in the moral law through the noble ideal of a universal realm of *ends in themselves* (rational beings) to which we can belong as members only when we scrupulously conduct ourselves by maxims of freedom as if they were laws of nature.

Concluding remark

The speculative use of reason *with respect to nature* leads to the absolute necessity of some supreme cause of *the world*. The practical use of reason *with regard to freedom* leads also to an absolute necessity, but only of *the laws of actions* of a rational being as such. Now, it is an essential principle of all use of our reason to push its knowledge to an awareness of its necessity, for otherwise it wouldn't be rational knowledge. But it is also an equally essential *limitation* of this very same reason that it can't see that

*necessarily x exists or y happens, or
necessarily z ought to happen,*

except on the basis of some *condition* that applies to *x* or *y* or *z*. •But the obtaining of a condition won't make something necessary unless the condition itself is necessary; and so •if reason keeps searching for conditions it only pushes its satisfaction further and further into the future. So reason, restlessly seeking the unconditionally necessary, sees itself as *having to assume it*, though it has no way of making it comprehensible to itself; it is happy enough if it can merely discover the concept that is *compatible* with this presupposition. According to my account of the supreme principles of morality, reason can't render comprehensible the absolute necessity of an unconditional practical law (such as the categorical imperative must be). If you want to complain about this, don't blame my account—blame reason! •Not that *blame* is appropriate: reason can't be blamed for being unwilling to explain the moral law through a condition—i.e. by making some interest its basis—for a law explained in that way would no longer be if it did, the law would cease to be moral and would no longer be the supreme law of freedom. So we truly don't comprehend the unconditional practical necessity of the moral imperative; but we do comprehend its *incomprehensibility*, which is all that can fairly be demanded of a philosophy that in its principles forces its way out to the boundaries of human reason.