

# Groundwork for the Metaphysic of Morals

Immanuel Kant

Copyright © Jonathan Bennett 2017. All rights reserved

[Brackets] enclose editorial explanations. Small ·dots· enclose material that has been added, but can be read as though it were part of the original text. Occasional •bullets, and also indenting of passages that are not quotations, are meant as aids to grasping the structure of a sentence or a thought. Every four-point ellipsis . . . . indicates the omission of a brief passage that seems to present more difficulty than it is worth. Longer omissions are reported between 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>

## Preface

Ancient Greek philosophy was divided into three branches of knowledge: •natural science, •ethics, and •logic. This classification perfectly fits what it is meant to fit; the only improvement it needs is the supplying of the *principle* on which it is based; that will let us be sure that the classification does cover all the ground, and will enable us to define the necessary subdivisions of the three broad kinds of knowledge. [Kant, following the Greek, calls the trio *Physik*, *Ethik* and *Logik*. Our word 'physics' is much too narrow for *Physik*, which is why 'natural science' is preferred here. What is lost is the surface *neatness* of the Greek and German trio, and of the contrast between natural science and metaphysics, *Physik* and *Metaphysik*]

There are two kinds of rational knowledge:

- material knowledge, which concerns some object, and
- formal knowledge, which pays no attention to differences between objects, and is concerned only with the *form* of understanding and of reason, and with the universal rules of thinking.

Formal philosophy is called •'logic'. Material philosophy—having to do with definite objects and the laws that govern them—is divided into two parts, depending on whether the laws in question are laws of •nature or laws of •freedom. Knowledge of laws of the former kind is called •'natural science', knowledge of laws of the latter kind is called •'ethics'. The two are also called 'theory of nature' and 'theory of morals' respectively.

•Logic can't have anything empirical about it—it can't have a part in which universal and necessary laws of thinking are derived from *experience*. If it did, it wouldn't be *logic*—i.e. a set of rules for the understanding or for reason, rules that are valid for *all* thinking and that must be rigorously proved. The •natural and •moral branches of knowledge, on the other

hand, *can* each have an empirical part; indeed, they must do so because each must discover the laws for its domain. For •the former, these are the laws of nature considered as something known through experience; and for •the latter, they are the laws of the human will so far as it is affected by nature. The two sets of laws are nevertheless very different from one another. The laws of *nature* are laws according to which everything *does* happen; the laws of morality are laws according to which everything *ought* to happen; they allow for conditions under which what ought to happen doesn't happen.

•*Empirical* philosophy is philosophy that is based on experience. •*Pure* philosophy is philosophy that presents its doctrines solely on the basis of *a priori* principles. Pure philosophy can in turn be divided into two: when it is entirely *formal* it is •logic; when it is confined to definite objects of the understanding, it is •metaphysics.

In this way there arises the idea of a two-fold metaphysic—a *metaphysic of nature* and a *metaphysic of morals*. Physics, therefore, will have an empirical part and also a rational part, and ethics likewise, though here the empirical part may be called more specifically 'practical anthropology' and the rational part 'morals' in the strict sense.

All crafts, trades and arts have profited from the division of labour; for when •each worker sticks to one particular kind of work that needs to be handled differently from all the others, he can do it better and more easily than when •one person does everything. Where work is not thus differentiated and divided, where everyone is a jack-of-all-trades, the crafts remain at an utterly primitive level. Now, here is a question worth asking: Doesn't pure philosophy in each of its parts require a man who is particularly devoted to that part? Some people regularly mix up the empirical with the rational, suiting their mixture to the taste of the public

without actually knowing what its proportions are; they call themselves *independent thinkers* and write off those who apply themselves exclusively to the rational part of philosophy as mere *ponderers*. Wouldn't things be improved for the learned profession as a whole if those 'independent thinkers' were *warned* that they shouldn't carry on two employments at once—employments that need to be handled quite differently, perhaps requiring different special talents for each—because all you get when one person does several of them is bungling? But all I am asking is this: Doesn't the nature of the science of philosophy require that we carefully separate its empirical from its rational part? That would involve putting

- a metaphysic of nature before real (empirical) natural science, and
- a metaphysic of morals before practical anthropology.

Each of these two branches of metaphysics must be carefully cleansed of everything empirical, so that we can know how much pure reason can achieve in each branch, and from what sources it creates its *a priori* teaching. The metaphysic of morals must be cleansed in this way, no matter who the metaphysicians of morals are going to be—whether they will include all the moralists (there are plenty of them!) or only a few who feel a calling to this task.

Since my purpose here is directed to moral philosophy, I narrow the question I am asking down to this:

- Isn't it utterly necessary to construct a pure moral philosophy that is completely freed from everything that may be only empirical and thus belong to anthropology?

That there must be such a philosophy is self-evident from the common idea of *duty* and *moral laws*. Everyone must admit that if a law is to hold *morally* (i.e. as a basis for someone's being *obliged* to do something), it must imply

absolute necessity; that the command: *You are not to lie* doesn't apply only to human beings, as though it had no force for other rational beings (and similarly with all other moral laws properly so called); that the basis for obligation here mustn't be looked for in people's natures or their circumstances, but must be found *a priori* solely in the concepts of pure reason; and that any precept resting on principles of mere experience may be called a practical rule but never a moral law. This last point holds even if there is something universal about the precept in question, and even if its empirical content is very small (perhaps bringing in only the motive involved).

Thus not only are moral laws together with their principles essentially different from all practical knowledge involving anything empirical, but all moral philosophy rests solely on its pure or non-empirical part. Its application to human beings doesn't depend on knowledge of any facts about them (anthropology); it gives them, as rational beings, *a priori* laws—ones that are valid *whatever* the empirical circumstances may be. (Admittedly experience comes into the story in a certain way, because these laws require a power of judgment that has been sharpened by experience—partly in order to pick out the cases where the laws apply and partly to let the laws get into the person's will and to stress that they are to be acted on. For a human being has so many preferences working on him that, though he is quite capable of having the *idea* of a practical pure reason, he can't so easily bring it to bear on the details of how he lives his life.)

A metaphysic of morals is therefore indispensable, for two reasons, one theoretical and one practical. One reason comes from our wish, as theoreticians, to explore the source of the *a priori* practical principles that lie in our reason. The other reason is that until we have the guide and supreme

norm for making correct moral judgments, morality itself will be subject to all kinds of corruption. Here is the reason for that. For something to be morally good, it isn't enough that it conforms to the moral law; it must be done *because* it conforms to the law. An action that isn't performed with that motive *may* happen to fit the moral law, but its conformity to the law will be chancy and unstable, and more often than not the action won't be lawful at all. So we need to find the moral law in its purity and genuineness, this being what matters most in questions about conduct; and the only place to find it is in a philosophy that is *pure* in the sense I have introduced—see page 1. So metaphysics must lead the way; without it there can't be any moral philosophy. Philosophy that isn't pure, i.e. that mixes pure principles with empirical ones, doesn't deserve the name of 'philosophy' (for what distinguishes philosophy from intelligent common sense is precisely that the former treats as separate kinds of knowledge what the latter jumbles up together). Much less can it count as 'moral philosophy', since by this mixing of pure with empirical it deprives morality of its purity and works against morality's own purposes.

I am pointing to the need for an entirely new field of investigation to be opened up. You might think that there is nothing new about it because it is already present in the famous Wolff's 'introduction' to his moral philosophy (i.e. in what he called 'universal practical philosophy'); but it isn't. Precisely because his work aimed to be *universal* practical philosophy, it didn't deal with any particular *kind* of will, and attended only to *will in general* and with such actions and conditions as that brings in; and so it had no room for the notion of a will that is determined by *a priori* principles with no empirical motives, which means that it had no place for anything that could be called a *pure* will. Thus Wolff's 'introduction' . . . concerns the actions and conditions of the

**human will** as such, which for the most part are drawn from empirical psychology, whereas the metaphysic of morals aims at a non-empirical investigation, namely investigating the idea and principles of a **possible pure will**. Without having the least right to do so, Wolff's 'universal practical philosophy' does have things to say about *laws* and *duty*; but this doesn't conflict with what I have been saying. For the authors of this intellectual project remain true to *their* idea of it in this part of its territory also: they don't distinguish

- motives that are presented completely *a priori* by reason alone and are thus *moral* in the proper sense of the word,

from

- motives that involve empirical concepts—ones that the understanding turns into universal concepts by comparing experiences.

In the absence of that distinction, they consider motives without regard to how their sources differ; they treat them as all being of the same kind, and merely *count* them; and on that basis they formulate their concept of *obligation*, so-called. This is as far from *moral* obligation as it could be; but in a philosophy that doesn't decide whether the origin of all possible practical concepts is *a priori* or *a posteriori*, what more could you expect?

Intending some day to publish a metaphysic of morals, I now present this groundwork, this exercise of foundation-laying, for it. There is, to be sure, no other basis for such a metaphysic than a critical examination of *pure practical reason*, just as there is no other basis for metaphysic than the critical examination of *pure speculative reason* that I have already published. [The unavoidable word 'speculative' (like its cognate 'speculation') is half of the dichotomy between practical and speculative. A speculative endeavour is one aimed at establishing truths about what *is* the case, implying nothing about what *ought* to be the

case; with no suggestion that it involves guesswork or anything like that. Two of Kant's most famous titles—*Critique of Pure Reason* and *Critique of Practical Reason*—are really short-hand for *Critique of Pure Speculative Reason* and *Critique of Pure Practical Reason*, respectively. That involves the speculative/practical contrast; there is no pure/practical contrast. The second of those two works, incidentally, still lay in the future when Kant wrote the present work.] However, I have three reasons for not plunging straight into a critical examination of pure practical reason. **(1)** It is nowhere near as important to have a critical examination of pure practical reason as it is to have one of pure speculative reason. That is because even in the commonest mind, human reason can easily be brought to a high level of correctness and completeness in moral matters, whereas reason in its theoretical but pure use is wholly dialectical [= 'runs into unavoidable self-contradictions']. **(2)** When we are conducting a critical examination of pure practical reason, I insist that the job is not finished until practical reason and speculative reason are brought together and unified under a common concept of *reason*, because ultimately they have to be merely different applications of one and the same *reason*. But I couldn't achieve this kind of completeness here without confusing the reader by bringing in considerations of an altogether different kind from the matter in hand. That is why I have used the title *Groundwork for the Metaphysic of Morals* rather than *Critique of Pure Practical Reason*. **(3)** A metaphysic of morals, in spite of its forbidding title, can be done in a popular way so that people of ordinary intelligence can easily take it in; so I find it useful to separate this preliminary work on the foundation, dealing with certain subtleties here so that I can keep them out of the more comprehensible work that will come later. [Here and throughout, 'popular' means 'pertaining to or suitable for ordinary not very educated people'. The notion of being widely liked is not prominent in its meaning.]

In laying a foundation, however, all I am doing is seeking and establishing the *supreme principle of morality*—a self-contained and entirely completable task that should be kept separate from every other moral inquiry. Until now there hasn't been nearly enough attention to this important question of the nature of and basis for the supreme principle of morality. My conclusions about it could be clarified by bringing the supreme principle to bear on the whole system of morality, and confirmed by how well it would serve all through. But I must forgo this advantage: basically it would gratify me rather than helping anyone else, because a principle's being easy to use and its seeming to serve well don't prove for sure that it is right. They are more likely merely to create a *bias* in its favour, which will get in the way of its being ruthlessly probed and evaluated in its own right and without regard to consequences.

[Kant has, and uses in the present work, a well-known distinction between 'analytic' propositions (known to be true just by analysing their constituent concepts) and 'synthetic' propositions (can't be known without bringing in something that the concepts don't contain). In this next sentence he uses those terms in a different way—one that goes back to Descartes—in which they mark off not two kinds of proposition but two ways of proceeding. In the analytic procedure, you start with what's familiar and on that basis work out what the relevant general principles are; synthetic procedure goes the other way—you start with general principles and derive familiar facts from them.]

In the present work I have adopted the method that is, I think, the most suitable if one wants to proceed analytically from common knowledge to settling what its supreme principle is, and then synthetically from examining this principle and its sources back to common knowledge to which it applies. So the work is divided up thus:

**Chapter 1** Moving from common-sense knowledge to philosophical knowledge about morality.

**Chapter 2** Moving from popular moral philosophy to the

metaphysic of morals.

**Chapter 3** Final step from the metaphysic of morals to the critical examination of pure practical reason.

## Chapter 1:

### Moving from common-sense knowledge to philosophical knowledge about morality

Nothing in the world—or *out of it!*—can possibly be conceived that could be called ‘good’ without qualification except a GOOD WILL. Mental *talents* such as intelligence, wit, and judgment, and *temperaments* such as courage, resoluteness, and perseverance are doubtless in many ways good and desirable; but they can become extremely bad and harmful if the person’s *character* isn’t good—i.e. if the *will* that is to make use of these •gifts of nature isn’t good. Similarly with •gifts of fortune. Power, riches, honour, even health, and the over-all well-being and contentment with one’s condition that we call ‘happiness’, create pride, often leading to arrogance, if there isn’t a good will to correct their influence on the mind. . . . Not to mention the fact that the sight of someone who shows no sign of a pure and good will and yet enjoys uninterrupted prosperity will never give pleasure to an impartial rational observer. So it seems that without a good will one can’t even be worthy of being happy.

Even qualities that are conducive to this good will and can make its work easier have no intrinsic unconditional worth. We rightly hold them in high esteem, but only because we assume them to be accompanied by a good will; so we can’t take them to be absolutely •or unconditionally• good.

•Moderation in emotions and passions, self-control, and calm deliberation not only are good in many ways but seem even to constitute part of the person’s *inner* worth, and they were indeed unconditionally valued by the ancients. Yet they are very far from being good without qualification—good in themselves, good in any circumstances—for without the principles of a good will they can become extremely bad: •for example•, a villain’s •coolness makes him far more dangerous and more straightforwardly abominable to us than he would otherwise have seemed.

What makes a good will *good*? It isn’t what it brings about, its usefulness in achieving some intended end. Rather, good will is good because of *how it wills*—i.e. it is *good in itself*. Taken just in itself it is to be valued incomparably more highly than anything that could be brought about by it in the satisfaction of some preference—or, if you like, the sum total of all preferences! Consider this case:

Through bad luck or a miserly endowment from step-motherly nature, this person’s will has no power at all to accomplish its purpose; not even the greatest effort on his part would enable it to achieve anything it aims at. But he does still have a good will—not as a

mere wish but as the summoning of all the means in his power.

The good will of this person would sparkle like a jewel all by itself, as something that had its full worth in itself. Its value wouldn't go up or down depending on how useful or fruitless it was. If it was useful, that would only be the *setting* of the jewel, so to speak, enabling us to handle it more conveniently in commerce (a diamond ring is easier to manage than a diamond) or to get those who don't know much about jewels to look at it. But the setting doesn't affect the value of the jewel and doesn't recommend it the experts.

But there is something extremely strange in this idea of the absolute worth of the will—the mere *will*—with no account taken of any use to which it is put. It is indeed so strange that, despite the agreement even of common sense (an agreement I have exhibited in the preceding three paragraphs), you're bound to suspect that there may be nothing to it but high-flown fancy, and that I have misunderstood what nature was up to in appointing reason as the ruler of our will. So let us critically examine the idea from the point of view of this suspicion.

We take it as an axiom that in the natural constitution of an organized being (i.e. one suitably adapted to life) no organ will be found that isn't perfectly adapted to its purpose, whatever that is. Now suppose that nature's real purpose for you, a being with reason and will, were that you should survive, thrive, and be happy—in that case nature would have hit upon a *very* poor arrangement in appointing your *reason* to carry out this purpose! For all the actions that you need to perform in order to carry out *this* intention of nature - and indeed the entire regulation of your conduct—would be marked out for you much more exactly and reliably by *instinct* than it ever could be by *reason*. And if nature had

favoured you by giving you reason as well as instinct, the role of reason would have been to let you contemplate the happy constitution of your nature, to admire it, to rejoice in it, and to be grateful for it to its beneficent cause; *not* to let you subject your faculty of desire to that weak and delusive guidance and to interfere with nature's purpose. In short, nature would have taken care that reason didn't intrude into *practical morality* and have the presumption, with its weak insight, to think out for itself the plan of happiness and how to get it. Nature would have taken over the choice not only of ends but also of the means to them, and with wise foresight she would have entrusted both to instinct alone. [Kant presents this paragraph in terms not of 'you' but of 'a being'.]

What we find in fact is that the more a cultivated reason devotes itself to the enjoyment of life and happiness, the more the person falls short of true contentment; which is why many people—especially those who have made the greatest use of reason—have a certain hostility towards reason, though they may not be candid enough to admit it. They have drawn many advantages from reason; never mind about its role in the inventions that lead to ordinary luxuries; my interest is in the advantages of intellectual pursuits, which eventually seem to these people to be also a luxury of the understanding. But after looking over all this they find that they have actually brought more trouble on themselves than they have gained in happiness; and eventually they come not to *despise* but to *envy* the common run of people who stay closer to merely natural instinct and don't give reason much influence on their doings. So much for the drawbacks of well-being and happiness as one's dominant aim in life. As for those who play down or outright deny the boastful eulogies that are given of the happiness and contentment that reason can supposedly bring us: the judgment they are making doesn't involve gloom, or

ingratitude for how well the world is governed. Rather, it's based on the idea of another and far nobler purpose for their existence. It is for achieving *this* purpose, not happiness, that reason is properly intended; and this purpose is the supreme condition, so that the private purposes of men must for the most part take second place to it. Its being the supreme or highest condition means that it isn't itself conditional on anything else; it is to be aimed at *no matter what* else is the case; which is why our private plans must stand out of its way.

So reason isn't competent to act as a guide that will lead the will reliably to its objectives and will satisfy all our needs (indeed it *adds* to our needs!); an implanted instinct would do this job much more reliably. Nevertheless, reason is given to us as a *practical* faculty, that is, one that is meant to have an influence on the *will*. Its proper function must be to produce a will that is good *in itself* and not good *as a means*. Why? Because

- nature has everywhere distributed capacities suitable to the functions they are to perform,
- the *means* to good are, as I have pointed out, better provided for by instinct, and
- reason and it alone can produce a will that is good *in itself*.

This *good* will needn't be the sole and complete good, but it must be the condition of all others, even of the desire for happiness. So we have to consider two purposes: **(1)** the unconditional purpose of producing a good will, and **(2)** the conditional purpose of being happy. Of these, **(1)** requires the cultivation of reason, which - at least in this life—in many ways *limits* and can indeed almost *eliminate* **(2)** the goal of happiness. This state of affairs is entirely compatible with the wisdom of nature; it doesn't have nature pursuing its goal clumsily; because reason, recognizing that its highest

practical calling is to establish a good will, can by achieving that goal get a contentment of its own kind (the kind that comes from attaining a goal set by reason), even though this gets in the way of things that the person merely *prefers*.

So we have to develop •the concept of a will that is to be esteemed as good in itself without regard to anything else, •the concept that always takes first place in judging the total worth of our actions, with everything else depending on it, •a concept that is already lodged in any natural and sound understanding, and doesn't need to be *taught* so much as to be *brought to light*. In order to develop and unfold it, I'll dig into the concept of duty, which *contains* it. The concept of a good will is present in the concept of duty, •not shining out in all its objective and unconditional glory, but rather •in a manner that brings it under certain subjective •restrictions and •hindrances; but •these are far from concealing it or disguising it, for they rather bring it out by contrast and make it shine forth all the more brightly. I shall now look at that contrast.

•My topic is the difference between doing something from duty and doing it for other reasons. In tackling this, I shall set aside without discussion two kinds of case—one for which my question doesn't arise, and a second for which the question arises but is too easy to answer for the case to be interesting or instructive. Following those two, I shall introduce two further kinds of case. **(1)** I shan't discuss actions which—even if they are useful in some way or other—are clearly *opposed to* duty, because with them the question of doing them *from duty* doesn't even arise. **(2)** I shall also ignore cases where someone does *A*, which really is in accord with duty, but where what he directly *wants* isn't to perform *A* but to perform *B* which somehow leads to or involves *A*. For example: he (*B*) unbolts the door so as to escape from the fire, and in so doing he (*A*) enables

others to escape also. There is no need to spend time on such cases, because *in them* it is easy to tell whether an action that is in accord with duty is done •from duty or rather •for some selfish purpose. **(3)** It is far harder to detect that difference when the action the person performs—one that is in accord with duty—is what he directly wanted to do, rather than being something he did only because it was involved in something *else* that he directly wanted to do. Take the example of a shop-keeper who charges the same prices for selling his goods to inexperienced customers as for selling them to anyone else. This is in accord with duty. But there is also a prudential and not-duty-based motive that the shop-keeper might have for this course of conduct: when there is a buyers' market, he may sell as cheaply to children as to others so as not to lose customers. Thus the customer is honestly served, but we can't infer from this that the shop-keeper has behaved in this way from duty and principles of honesty. His own advantage requires this behaviour, and we can't assume that in addition he directly wants something for his customers and out of love for them he charges them all the same price. His conduct of his policy on pricing comes neither from *duty* nor from *directly wanting* it, but from a selfish purpose. [Kant's German really does say first that the shop-keeper *isn't* led by a direct want and then that he *is*. His point seems to be this: The shop-keeper *does* want to treat all his customers equitably; his intention is aimed at precisely that fact about his conduct (unlike the case in **(2)** where the agent enables other people to escape but isn't aiming at that *at all*). But the shop-keeper's intention doesn't *stop there*, so to speak; he wants to treat his customers equitably not because of what he wants for *them*, but because of how he wants them to behave later in *his* interests. This involves a kind of indirectness, which doesn't assimilate this case to **(2)** but does distinguish it from a fourth kind of conduct that still isn't morally worthy but not because it

involves the 'indirectness' of **(2)** or that of **(3)**.]

**(4)** It is a duty to preserve one's life, and moreover everyone directly wants to do so. But because of the power of that want, the often anxious care that most men have for their survival has no intrinsic worth, and the maxim *Preserve yourself* has no moral content. Men preserve their lives *according to* duty, but not *from* duty. But now consider this case:

Adversities and hopeless sorrow have completely taken away this unfortunate man's relish for life. But his fate has not made him passively despondent or dejected. He is strong in soul, and is exasperated at how things have gone for him, and would like actively to do something about it. Specifically, he *wishes for death*. But he preserves his life without loving it, not led by any want or fear, but acting from duty.

For this person the maxim *Preserve yourself* has moral content.

We have a duty to be charitably helpful where we can, and many people are so sympathetically constituted that without any motive of vanity or selfishness they find an inner satisfaction in spreading joy and take delight in the contentment of others if they have made it possible. But I maintain that such behaviour, done in that spirit, has no true moral worth, however amiable it may be and however much it accords with duty. It should be classed with actions done from other wants, such as the desire for honour. With luck, someone's desire for honour may lead to conduct that in fact accords with duty and does good to many people; in that case it deserves praise and encouragement; but it doesn't deserve high esteem, because the maxim on which the person is acting doesn't have the moral content of an action done not because the person *likes* acting in that way but from duty. [In this context, 'want' and 'liking' and 'desire' are used

to translate *Neigung*, elsewhere in this version translated as 'preference'; other translations mostly use 'inclination'.]

Now consider a special case:

This person has been a friend to mankind, but his mind has become clouded by a sorrow of his own that has extinguished all feeling for how others are faring. He still has the power to benefit others in distress, but their need leaves him untouched because he is too preoccupied with his own. But now he tears himself out of his dead insensibility and acts charitably purely from duty, without feeling any want or liking so to behave.

Now, for the first time, his conduct has genuine moral worth. Having been deprived by nature of a warm-hearted temperament, this man could find in himself a source from which to give himself a far higher worth than he could have got through such a temperament. It is just here that the worth of character is brought out, which is morally the incomparably highest of all: he is beneficent not from preference but from duty.

To secure one's own happiness is a duty (at least indirectly), because discontent with one's condition—bundled along by many cares and unmet needs—could easily become a great *temptation to transgress against duties*. But quite apart from duty, all men have the strongest and deepest desire [*Neigung*] for happiness, because in the idea of happiness all our desires are brought together in a single sum-total. But the injunction 'Be happy!' often takes a form in which it *thwarts* some desires, so that a person can't get a clear and secure concept of •the sum-total of satisfactions that goes under the name 'happiness'. So it isn't surprising that the prospect of •a single satisfaction, definite as to what it is and when it can be had, can outweigh a fluctuating idea •such as that of happiness•. For example, a man with the gout

[a painful ailment made worse by alcohol and rich food] can choose to enjoy what he likes and put up with the consequences, because according to his calculations (this time, anyway) he hasn't sacrificed present pleasure to a possibly groundless expectation of the 'happiness' that health is supposed to bring. But even for this man, whose will is not settled by the general desire for happiness and for whom health plays no part in his calculations, there still remains—as there does for everyone—the law that *he ought to promote his happiness, not from wanting or liking but from duty*. Only by following this could his conduct have true moral worth.

No doubt this is how we should understand the scriptural passages that command us to love our neighbour and even our enemy. We can't be commanded to *feel* love for someone, or to simply *prefer* that he thrive. There are two sorts of love: •practical love that lies in the will and in principles of action, and •pathological love that lies in the direction the person's feelings and tender sympathies take. [Kant uses 'pathological' simply to mean that this is a *state* that the person is in; from Greek *pathos* = 'that which happens to a person'; no suggestion of abnormality. His point is that being a *loving* person is no more morally significant than being a *stupid* person or a *right-handed* person.] The latter of these cannot be commanded, but the former can be—and that is a command to do good to others from duty, even when you don't want to do it or like doing it, and indeed even when you naturally and unconquerably *hate* doing it.

•So much for the **first proposition** of morality:

•For an action to have genuine moral worth it must be done from duty. •

The **second proposition** is:

•An action that is done from duty doesn't get its moral value from the purpose that's to be achieved through it but from the maxim that it involves, •giving the *reason* why the person acts thus. •

So the action's moral value doesn't depend on whether what is aimed at in it is actually achieved, but solely on the principle *of the will* from which the action is done, irrespective of anything the faculty of *desire* may be aiming at. From what I have said it is clear that the purposes we may have in acting, and their effects as drivers of the will towards desired ends, can't give our actions any unconditional value, any *moral* value. Well, then, if the action's moral value isn't to be found in

- the will in its relation to its hoped-for effect,

where *can* it be found? The only possible source for it is

- the principle on which the will acts—and never mind the ends that may be achieved by the action.

For the will stands at the crossroads, so to speak, at the intersection between •its *a priori* principle, which is formal, and •its *a posteriori* driver—the contingent desire that acts on it—which is material. In that position it must be determined by something; and if it is done from duty it must be determined by the formal principle of the will, since every material principle—every contingent driver of the will—has been withdrawn from it.

The **third proposition**—a consequence of the first two—I would express as follows:

- To have a duty is to be required to act in a certain way out of respect for law.

(1) As for what will result from my action, I can certainly *prefer* or *be drawn to* it, but I can't have *respect* for it; to earn my respect it would have to be something the will *does*, not merely something that its doings *lead to*. (2) Similarly, I can't •respect any want or preference: if the preference is mine, the most I can do is to •endorse it; if it is someone else's I can even •love it—i.e. see it as favourable to my interests. What can get respect and can thus serve as a command is •something that isn't (1) a consequence of my

volition but only a source for it, and isn't (2) in the service of my preferences but rather overpowers them or at least prevents them from being considered in the choice I make; •this something is, in a word, *law* itself. Suppose now that someone acts from duty: the influence of his *preferences* can't have anything to do with this, and so facts about what he might achieve by his action don't come into it either; so what is there left that can lead him to act as he does? If the question means 'What is there objectively, i.e. distinct from himself, that determines his will in this case?' the only possible answer is *law*. And if the question concerns what there is *in the person* that influences his will—i.e. what subjectively influences it—the answer has to be *his respect for this practical law*, and thus his acceptance of the maxim *I am to follow this law even if it thwarts all my desires*. (A maxim is a subjective principle of volition. The objective principle is the practical law itself; it would also be the subjective principle for all rational beings if reason fully controlled the formation of preferences.)

So an action's moral value doesn't lie in •the effect that is expected from it, or in •any principle of action that motivates it because of this expected effect. All the expected effects—something agreeable for me, or even happiness for others—could be brought about through other causes and don't *need* •the will of a rational being, whereas the highest good—what is unconditionally good—can be found only in •such a will. So this wonderful good, which we call *moral* goodness, can't consist in anything but *the thought of law* in itself *that only a rational being can have*—with the will being moved to act by this thought and not by the hoped-for effect of the action. When the person acts according to this conception, this moral goodness is already present •in him; we don't have to look for it •in the upshot of

his action.<sup>1</sup> [In passages like this, ‘thought’ translates *Vorstellung* = ‘mental representation’.]

So we have a law the thought of which can settle the will without reference to any expected result, and *must* do so if the will is to be called absolutely good without qualification; what kind of law can this be? Since I have robbed the will of any *impulses* that could come to it from obeying any law, nothing remains to serve as a ‘guiding’ principle of the will except *conduct’s universally conforming to law as such*. That is, **I ought never to act in such a way that I couldn’t also will that the maxim on which I act should be a universal law**. In this context the ‘guiding’ principle of the will is conformity to law as such, not bringing in any particular law governing some class of actions; and it *must* serve as the will’s principle if *duty* is not to be a vain delusion and chimerical concept. Common sense in its practical judgments is in perfect agreement with this, and constantly has this principle in view.

Consider the question: May I when in difficulties make a promise that I intend not to keep? The question obviously has two meanings: is it •prudent to make a false promise?

does it conform to •duty to make a false promise? No doubt it often *is* •prudent, ·but not as often as you might think·. Obviously the false promise isn’t made prudent by its merely extricating me from my present difficulties; I have to think about whether it will in the long run cause more trouble than it saves in the present. Even with all my supposed cunning, the consequences can’t be so easily foreseen. People’s loss of trust in me might be far more disadvantageous than the trouble I am now trying to avoid, and it is hard to tell whether it mightn’t be more *prudent* to act according to a universal maxim *not ever to make a promise that I don’t intend to keep*. But I quickly come to see that such a maxim is based only on fear of consequences. Being truthful from •duty is an entirely different thing from being truthful out of •fear of bad consequences; for in •the former case a law is included in the concept of the action itself (·so that the right answer to ‘What are you doing?’ will include a mention of that law·); whereas in •the latter I must first look outward to see what results my action may have. [In the preceding sentence, Kant speaks of a ‘law for me’ and of results ‘for me’.] To deviate from the principle of duty is certainly bad; whereas to be unfaithful to my maxim

<sup>1</sup> It might be objected that I tried to take refuge in an obscure feeling behind the word ‘respect’, instead of clearing things up through a concept of reason. Although respect is indeed a feeling, it doesn’t come from outer influence; rather, it is a •feeling that a rational concept creates unaided; so it is different in kind from all the •feelings caused from outside, the ones that can come from desire or fear. When I directly recognize something as a law for myself I recognize it *with respect*, which merely means that I am conscious of submitting my will to a law without interference from any other influences on my mind. The will’s being directly settled by law, and the consciousness of this happening, is called ‘respect’; so *respect* should be seen as an *effect* of the law’s operation on the person’s will, not as a *cause* of it. Really, respect is the thought of a value that breaks down my self-love. Thus it is not something to be either desired or feared, though it has something analogous to both ·desire and fear·. The only thing that can be respected is law, and it has to be the law that we •impose on ourselves yet •recognize as necessary in itself.

•As a law it makes us subject to it, without consulting our self-love; which gives it some analogy to fear.

•As imposed on us by ourselves, it is a consequence of our will; which gives it some analogy to preference.

·This is really the only *basic* sense of the term ‘respect’·. Any •respect for a *person* is only •respect for the law (of righteousness, etc.) of which the person provides an example. Our respect for a person’s talents, for instance, is our recognition that we ought to practice until we are as talented as he is; we see him as a kind of example of a •law, because we regard it as our •duty to improve our talents. ·So respect for persons is a disguised form of respect for law·. All moral *concern* (as it is called) consists solely in respect for the law.

of prudence may be very advantageous to me, though it is certainly safer to abide by it. How can I know whether a deceitful promise is consistent with duty? The shortest way to go about finding out is also the surest. It is to ask myself:

- Would I be content for my maxim (of getting out of a difficulty through a false promise) to hold as a *universal law*, for myself as well as for others?

•That is tantamount to asking:

- Could I say to myself that anyone may make a false promise when he is in a difficulty that he can't get out of in any other way?

Immediately I realize that I could will •the lie but not •a universal law to lie; for such a law would result in there being no promises at all, because it would be futile to offer stories about my future conduct to people who wouldn't believe me; or if they carelessly did believe me and were taken in •by my promise•, would pay me back in my own coin. Thus my maxim would necessarily destroy itself as soon as it was made a universal law.

So I don't need to be a very penetrating thinker to bring it about that my will is morally good. Inexperienced in how the world goes, unable to prepare for all its contingencies, I need only to ask myself: *Can you will that your maxim become a universal law?* If not, it must be rejected, not because of any harm it might bring to anyone, but because there couldn't be a system of •universal legislation that included it as one of its principles, and •that is the kind of legislation that reason forces me to respect. I don't yet see what it is based on (a question that a philosopher may investigate), but I at least understand these two:

- It is something whose value far outweighs all the value of everything aimed at by *desire*,
- My duty consists in my having to act from pure respect for the practical law.

Every other motive must yield to duty, because it is the condition of a •will that is good in itself, and the value of •that surpasses everything.

And so in the common-sense understanding of morality we have worked our way through to its *principle*. Admittedly, common sense doesn't have the abstract thought of this principle as something universal, but it always has the principle in view and uses it as the standard for its judgments.

It would be easy to show how common sense, with this compass in its hand, knows very well how to distinguish •good from •bad, •consistent with duty from •inconsistent with duty. To do this it doesn't have to be taught anything new; it merely needs (Socrates-fashion) to have its attention drawn to the principle that it already has; and thus •we can see• that neither science nor philosophy is needed in order to know what one must do to be honest and good, and even to be wise and virtuous. That's something we might well have assumed in advance: that the knowledge of what every person is obliged to do (and thus also what everyone is obliged to know) is *everyone's* business, even the most common person's. We can't help admiring the way common sense's ability to make •practical judgment outstrips its ability to make •theoretical ones. In •theoretical judgments, if common sense ventures to go beyond the laws of experience and perceptions of the senses, it falls into sheer inconceivabilities and self-contradictions, or at least into a chaos of uncertainty, obscurity, and instability. On the other hand, it is just when common sense *excludes* •everything empirical—that is•, all action-drivers that bring in the senses—that its ability to make •practical judgments first shows itself to advantage. It may then start splitting hairs, quibbling with its own conscience or with other claims concerning what should be called right, or wanting to satisfy itself about the exact worth of certain actions; and the great

thing about these activities of common sense is that in them it has as good a chance of getting it right as any philosopher has—perhaps even a better chance, because the philosopher doesn't have any *principle* that common sense lacks and his judgment is easily confused by a mass of irrelevant considerations so that it easily goes astray. Here are two ways in which we could inter-relate common-sense morality and philosophy: **(1)** We could go along with common-sense moral judgments, and bring in philosophy—if at all—only so as to make the system of morals more complete and comprehensible and its rules more convenient for use, especially in disputation. **(2)** We could steer common sense away from its fortunate simplicity in practical matters, and lead it *through* philosophy into a new path of inquiry and instruction. From what I have said, isn't it clear that **(1)** is the wiser option to take?

Innocence is indeed a glorious thing, but it is very sad that it doesn't take care of itself, and is easily led astray. For this reason, even wisdom—which consists in •doing and allowing more than in •knowing—needs science [*Wissenschaft*], not as something to learn from but as something that will ensure that wisdom's precepts get into the mind and stay there. [Knowing translates *Wissen*, which is half the word translated as 'science', an overlap that Kant surely intended. The 'science' in question here is presumably metaphysics.] Without that help, they are not likely to 'stay there', and here is why. Against all commands of duty that a man's reason presents to him as deserving of so much respect, he feels in himself a powerful •counter-weight—namely, his needs and preferences, the complete satisfaction of which he lumps together as 'happiness'. Reason issues inexorable commands without promising the preferences anything •by way of recompense. It ignores and has no respect for the claims •that desire makes—claims that are so impetuous and yet so

plausible, and which refuse to give way to any command. This gives rise to a natural dialectic—an intellectual conflict or contradiction—in the form of a propensity to argue against the stern laws of duty and their validity, or at least to cast doubt on their purity and strictness, and, where possible, to make them more accordant with our wishes and desires. This undermines the very foundations of duty's laws and destroys their dignity—which is something that even *ordinary* practical reason can't, when it gets right down to it, call good.

In this way common sense is driven to go outside its own territory and to take a step into the field of practical *philosophy*. It doesn't do this because of any speculative (= 'theory-building') need, which is something that never occurs to it so long as it is satisfied to remain merely healthy reason. [Kant's phrase translated here as 'common sense' is *gemeine Menschvernunft*, which contains *Vernunft* = 'reason'. Putting its bits together it could be taken to mean 'general human reason', but 'common sense' is about right.] Rather, it is driven to philosophy in order to become •informed and clearly •directed regarding the source of its principle and how exactly it differs from the maxims based on needs and preferences. It does this so as to escape from the embarrassment of opposing claims, and to avoid risking the loss of all genuine moral principles through the ambiguity in which common sense is easily involved—the ambiguity between the moral and prudential readings of questions about what one ought to do. Thus when common-sense moral thought develops itself, a dialectic surreptitiously occurs that forces it to look to philosophy for help, and the very same thing happens in common-sense theoretical thinking. It is true of each kind of ordinary or common-sense thought: each can come to rest only in a complete critical examination of our reason.

## 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.

### 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.<sup>16</sup>

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

<sup>16</sup> 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.<sup>17</sup> 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

<sup>17</sup> 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.