Introduction

A life devoted to the pursuit of philosophical discovery may be inwardly as full of drama and event – of obstacle and overcoming, battle and victory, challenge and conquest – as that of any general, politician, or explorer, and yet be outwardly so quiet and routine as to defy biographical narration. Immanuel Kant was born in 1724 in Königsberg, East Prussia, to a Pietist family of modest means.[1] Encouraged by his mother and the family pastor to pursue the career marked out by his intellectual gifts, Kant attended the University of Königsberg, and then worked for a time as a private tutor in the homes of various families in the neighborhood, while pursuing his researches in natural science. Later he got a position as a *Privatdozent*, an unsalaried lecturer who is paid by student fees, at the University. There Kant lectured on logic, metaphysics, ethics, geography, anthropology, mathematics, the foundations of natural science, and physics. In 1770, he finally obtained a regular professorship, the Chair of Logic and Metaphysics, at Königsberg. Because of limited means and variable health, Kant never married or travelled. He remained in the Königsberg area, a quiet, hardworking scholar and teacher, until his death in 1804.

But some time in the 1770s – we do not know exactly when – Kant began to work out ideas that were destined to challenge our conception of reason's relationship – and so of our own relationship – to the world around us. Kant himself compared his system to that of Copernicus, which explained the ordering of the heavens by turning them inside out, that is, by removing the earth – the human world – from the center, and making it revolve around the sun instead. Kant's own revolution also turns the world inside out, but in a very different way, for it places humanity back in the center. For Kant argued that the rational order which the metaphysician looks for in the world is neither something that we discover through experience, nor something that our reason assures us must be there. Instead, it is something which we human beings impose upon the world, in part through the construction of our knowledge, but also, in a different way, through our actions.

The implications for moral philosophy, first presented in the

[1] Pietism was a religious movement which emphasized inner religious experience, self-examination, and morally good works. Its emphasis on the importance of morality is often thought to have been a strong influence on Kant.

Groundwork of the Metaphysics of Morals, are profound. The *Groundwork* is an acknowledged philosophical classic, an introduction to one of the most influential accounts of our moral nature which the tradition has ever produced. Some of its central themes – that every human being is an end in himself or herself, not to be used as a mere means by others; that respect for one's own humanity finds its fullest expression in respect for that of others; and that morality is freedom, and evil a form of enslavement – have become not only well-established themes in moral philosophy, but also part of our moral culture.

But the *Groundwork* owes its popularity to its power, not to its accessibility. For like all of Kant's works, it is a difficult book. It is couched in the technical vocabulary which Kant developed for the presentation of his ideas. It presents us with a single, continuous argument, each of whose steps is itself an argument, which runs the length of the book. But the particular arguments which make up the whole are sufficiently difficult in themselves that their contribution to the larger argument is easy to lose sight of. The main aim of this Introduction will be to provide a kind of road map through the book, by showing how the material presented in each of the main sections contributes to the argument as a whole.

Section II

Although the argument of Section I proceeded from our ordinary ideas about morality, and involved the consideration of examples, it is not therefore an empirical argument. The examples do not serve as a kind of data from which conclusions about moral motivation are inductively drawn. Instead, the argument is based on our rational appraisal of the people in the examples, taking the facts about their motivation as given: if these people act from respect for law, as the examples stipulate, then their actions have moral worth. Whether anyone has ever actually acted from respect for law is a question about which moral philosophy must remain silent. So demonstrating that the categorical imperative governs

[7] Both here and later on in the discussion of the Formula of Universal Law, Kant makes it clear that he thinks the lawlike character of a maxim is a matter of its *form* rather than its *matter*. What does this mean? The distinction between form and matter is an inheritance of Aristotelian metaphysics, in which the matter of a thing is the materials or parts of which it is constructed, while the form is the arrangement of those parts that enables the object to serve its characteristic function. For instance if the function of a house is to serve as a shelter, we would say that the matter of the house is the walls and the roof, and the form is the way those parts are arranged so as to keep the weather out and the objects within protected. The parts of a maxim are usually the act which is done and the end for the sake of which it is done. We can show that the lawlike character of the maxim is a matter of the way the parts are arranged, the form, by considering a triple of maxims like this:

1 I will keep my weapon, because I want it for myself.
2 I will keep your weapon, because I want it for myself.
3 I will keep your weapon, because you have gone mad and may hurt someone.

Maxims 1 and 3 are maxims of good actions, while maxim 2 is of a bad action. Yet maxims 1 and 2 have the same purpose, and maxims 2 and 3 involve the same act. So the lawlike character of the maxim rests neither in the purpose, nor in the act, which are the parts or matter of the maxim. Instead it rests in the way those parts are combined – the form of the maxim. In a good maxim, the parts are so combined that the maxim can serve as a law: everyone could act on it.

our wills is not a matter of showing that we actually act on it. Instead, it is a matter of showing that we act on it insofar as we are rational. A comparison will help here. Showing that the principle of non-contradiction governs our beliefs is not a matter of showing that no one ever in fact holds contradictory beliefs, for people surely do. Nor is it a matter of showing that people are sometimes moved, say, to give up cherished beliefs when they realize those beliefs will embroil them in contradiction. Instead, it is a matter of showing that insofar as they are rational, that is what they do. Kant's project in Section II therefore is to "present distinctly the faculty of practical reason, from its general rules of determination to the point where the concept of duty arises from it" (AK 4: 412). In other words, in Section II Kant lays out a theory of practical reason, in which the moral law appears as one of the principles of practical reason.

It is a law of nature, very roughly speaking, that what goes up must come down. Toss this book into the air, and it will obey that law. But it will not, when it reaches its highest point, say to itself "I ought to go back down now, for gravity requires it." As rational beings, however, we do in this way reflect on, and sometimes even announce to ourselves, the principles on which we act. In Kant's words, we act not merely in accordance with laws, but in accordance with our representations or conceptions of laws (AK 4: 412).

Yet we human beings are not perfectly rational, since our desires, fears, and weaknesses may tempt us to act in irrational ways. This opens up the possibility of a gap between the principles upon which we actually act – our maxims or subjective principles – and the objective laws of practical reason. For this reason, we conceive the objective laws of practical reason as imperatives, telling us what we *ought* to do. The theory of practical reason is therefore a theory of imperatives.

Imperatives may be either hypothetical or categorical. A hypothetical imperative tells you that if you will something, you ought also to will something else: for example, if you will to be healthy, then you ought to exercise. That is an imperative of skill, telling you how to achieve some particular end. Kant believes that there are also hypothetical imperatives of prudence, suggesting what we must do given that we all will to be happy. A categorical imperative, by contrast, simply tells us what we ought to do, not on condition that we will something else, but unconditionally.

Kant asks how all these imperatives are "possible" (AK 4: 417), that is, how we can establish that they are legitimate requirements of reason, binding on the rational will. He thinks that in the case of hypothetical imperatives the answer is easy. A hypothetical imperative is based on the principle that whoever wills an end, insofar as he is rational, also wills the means to that end. This principle is analytic, since *willing* an end, as

opposed to merely wanting it or wishing for it or thinking it would be nice if it were so, is setting yourself to bring it about, to cause it. And setting yourself to cause something just is setting yourself to use the means to it. Since willing the means is conceptually contained in willing the end, if you will an end and yet fail to will the means to that end, you are guilty of a kind of practical contradiction.

Since a categorical imperative is unconditional, however, there is no condition given, like the prior willing of an end, which we can simply analyze to derive the "ought" statement. The categorical imperative must therefore be *synthetic*, so morality depends on the possibility of establishing a *synthetic a priori* practical principle.

The Formula of Universal Law

Kant does not, however, move immediately to that task; in fact, he will not be in a position to take that up until Section III. Section II is, like Section I, an analysis. Kant is still working towards uncovering *what* we have to prove *in order to* establish that moral requirements really bind our wills. The first step is to analyze the very idea of a categorical imperative in order to see what it "contains." Kant says:

> when I think of a categorical imperative I know at once what it contains. For since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. (AK 4: 420–1)

This is the sort of thing that makes even practiced readers of Kant gnash their teeth. A rough translation might go like this: the categorical imperative is a law, to which our maxims must conform. But the reason they must do so cannot be that there is some *further* condition they must meet, or some *other* law to which they must conform. For instance, suppose someone proposed that one must keep one's promises because it is the will of God that one should do so – the law would then "contain the condition" that our maxims should conform to the will of God. This would yield only a conditional requirement to keep one's promises – if you would obey the will of God, then you must keep your promises – whereas the categorical imperative must give us an *unconditional* requirement. Since there can be no such condition, all that remains is that the categorical imperative should tell us that our maxims themselves must be laws – that is, that they must be universal, that being the characteristic of laws.

There is a simpler way to make this point. What could make it true

that we must keep our promises because it is the will of God? That would be true only if it were true that we must indeed obey the will of God, that is, if "obey the will of God" were itself a *categorical* imperative. Conditional requirements give rise to a regress; if there are unconditional requirements, we must at some point arrive at principles on which we are required to act, not because we are commanded to do so by some yet higher law, but because they are laws in themselves. The categorical imperative, in the most general sense, tells us to act on *those* principles, principles which are themselves laws. Kant continues:

> There is, therefore, only a single categorical imperative and it is this: *act only in accordance with that maxim through which you can at the same time will that it become a universal law.* (AK 4: 421)

Kant next shows us how this principle serves to identify our duties, by showing us that there are maxims which it rules out – maxims which we could not possibly will to become universal laws. He suggests that the way to test whether you can will your maxim as a universal law is by performing a kind of thought experiment, namely, asking whether you could will your maxim to be a law of nature in a world of which you yourself were going to be a part. He illustrates this with four examples, the clearest of which is the second.

A person in financial difficulties is considering "borrowing" money on the strength of a false promise. He needs money, and knows he will get it only if he says to another person, "I promise you I will pay you back next week." He also knows perfectly well that he will not be able to repay the money by then. His question is whether he can will that the maxim of making a false promise in order to get some money should become a law of nature. Although Kant does not do this, it helps to set out the test in a series of steps.

The first step is to formulate the maxim. In most cases, the person is considering doing a certain action for a certain end, so the basic form of the maxim is "I will do Action-A in order to achieve Purpose-P." Suppose then that your maxim is:

> I will make a false promise in order to get some ready cash.

Next we formulate the corresponding "law of nature." It would be:

> Everyone who needs some ready cash makes a false promise.

At least where duties to others are concerned, Kant's test may be regarded as a formalization of the familiar moral challenge: "What if everybody did that?" In order to answer this question, you are to imagine a world where everybody does indeed do that. We might call this the "World of the Universalized Maxim." At this point it is important to notice that Kant says the categorical imperative tells you to act on a

maxim which you can *at the same time* will to be a universal law: he means at the same time as you will the maxim itself. So you are to imagine that you are in the World of the Universalized Maxim, trying to act on your maxim. For instance, you imagine that you are attempting to secure some ready cash by means of a false promise in a world where everyone who needs a little ready cash tries to secure it by means of a false promise. Now, finally, you are to ask whether you could will this state of affairs, in particular, whether any contradiction arises when you try to do so. Kant says, in the example at hand, that it does,

> For, the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses. (AK 4: 422)

Why is this a contradiction? This question has attracted an enormous amount of philosophical attention and many interpretations have been proposed. The views that have been suggested may be divided into three broad categories.

Proponents of a logical contradiction interpretation think Kant means there is a straightforward logical contradiction in the proposed law of nature. One might argue, for instance, that universalization of the maxim of false promising would undercut the very practice of making and accepting promises, thus making promises impossible and the maxim literally inconceivable.[8]

Kant's use of teleological language in some of the examples has suggested to proponents of the teleological contradiction interpretation that the contradiction emerges only when the maxim is conceived as a possible teleological law of nature. False promising violates the "natural purpose" of promising, which is to create trust and cooperation, so that a universal law of false promising could not serve as part of a teleological system of natural laws.

According to proponents of the practical contradiction interpretation, the maxim's efficacy in achieving its purpose would be undercut by its universalization. In willing its universalization, therefore, the agent would be guilty of the same sort of practical contradiction involved in the violation of a hypothetical imperative. In fact, the maxim in the example is derived from a hypothetical imperative – "if you need some ready cash, you ought to make a false promise" – which in turn is derived from a "law of nature" or "causal law" – namely that false promising is a cause of, and so a means to, the possession of ready cash.

[8] For the notion of a practice and the logical dependence of actions falling under the practice on the existence of the practice itself, see John Rawls, "Two Concepts of Rules," *Philosophical Review* 64 (January 1955), 3–32.

In the World of the Universalized Maxim, however, this law no longer obtains. So in willing the World of the Universalized Maxim the agent undercuts the causal law behind the hypothetical imperative from which his own maxim is derived, making his method of getting the money ineffective. Language supporting all of these interpretations can be found in Kant's texts, and different interpretations fit different examples better. The problem of finding a single account of the contradiction test that produces the right answers in all cases is one on which Kantians are still at work.

The question is complicated by the fact that Kant himself thinks contradictions may arise in two different ways (AK 4: 421, 424). In some cases, he says, the maxim cannot even be thought as a universal law of nature: the contradiction is in the very conception of the universalized maxim as a law. The example we have been considering is of that kind: there could not *be* a law that everyone who needs money should make false promises, so the maxim fails what is often called "the contradiction in conception test." Maxims which fail this test are in violation of strict or perfect duties, particular actions or omissions we owe to particular people, such as the duty to keep a promise, tell the truth, or respect someone's rights. But there are also maxims which we can conceive as universal laws, but which it would still be contradictory to *will* as laws: these maxims fail what is often called "the contradiction in the will test." They violate wide or imperfect duties, such as the duty to help others when they are in need, or to make worthwhile use of your talents.[9] Here again, there is disagreement about exactly what the contradiction is. Kant suggests that "all sorts of possible purposes" (AK 4: 423) would have to go unfulfilled in a world in which we had neglected our abilities and in which we could not count on the help

[9] In the *Groundwork*, Kant lines up the distinction between the contradiction in conception test and the contradiction in the will test with the traditional distinction between perfect and imperfect duties (described above) at AK 4: 421, and with a less familiar distinction between strict or narrow duties and wide duties at AK 4: 424. This parallel might be taken to suggest that these are just two sets of names for the same distinction, or at any rate that they coincide. But in the later *Metaphysics of Morals* Kant describes a category of duties which are characterized as perfect duties and yet which, because they are duties of virtue and all of those are wide, must be wide (AK 6: 421ff.). Kant explains the distinction between narrow and wide obligation in the *Metaphysics of Morals* at AK 6: 390–4. We have a duty of narrow obligation when we are required to perform a particular action, while we have a duty of wide obligation when we are required to adopt a certain general maxim (e.g. to promote the happiness of others) but have leeway as to how to carry the duty out. This explanation leaves the difference between the two distinctions unclear, and Kant never directly addresses the question how the two distinctions are related. If Kant's considered view is that these two distinctions do not coincide, we are left uncertain whether the contradiction in conception test is best understood as a test for perfect duties, or as a test for strict duties. These rather intricate issues about the categorization of duties matter to the reader of the *Groundwork* because one of the duties Kant uses as an example here – the duty not to commit suicide in order to avoid misery – is one of those apparently identified in the later work as a perfect duty of wide obligation. This should perhaps make us cautious about this example.

of others when we are in need. Since rationality commits us to willing the means to our ends, we must will a world in which these most general means – our own abilities and the help of others – would be available to us.

These examples are offered simply as a few illustrations to show how the categorical imperative works to establish the moral status of our actions. Generally, if a maxim passes the categorical imperative test, the action is permissible; if it fails, the action is forbidden, and, in that case, the opposite action or omission is required. The maxims in the examples fail the test, showing, for instance, that making a false promise is forbidden, and that helping others when they are in need is required. For a more complete account of what Kant thinks morality requires of us, however, the reader must look to the *Metaphysics of Morals*.

The thought experiment we have just considered shows us *how* to determine whether a maxim can be willed as a universal law, not *why* we should will only maxims that can be universal laws. Kant is not claiming that it is irrational to perform immoral actions because it actually embroils us in contradictions. The contradictions emerge only when we attempt to universalize our maxims, and the question why we should do that remains to be answered. It is to this question Kant turns next.

The Formula of Humanity

We have now seen what the categorical imperative says. In order to show that we actually have unconditional requirements, and so that ethics is real, we have to show that this principle is one that necessarily governs our wills. This investigation is in part a motivational one. Although Kant denies that we can ever know for sure that someone has been morally motivated, the moral law cannot have authority over our wills unless it is *possible* for us to be motivated by it. But Kant warns us that we cannot appeal to any empirical and contingent sources of motivation when making this argument. As we saw earlier, the sense in which we are trying to show that the moral law governs our wills is not that it actually moves us, either always or sometimes, but that it moves us insofar as we are rational. So the argument must show that the moral law has authority for any rational being, and this means it must appeal only to the principles of pure rational psychology.

As rational beings, as Kant said before, we act in accordance with our representations or conceptions of laws. But what inspires us to formulate a maxim or a law ("what serves the will as the objective ground of its self-determination") is an end (AK 4: 427). Whenever we actually decide to take action, it is always with some end in view: either we regard the action as good in itself, or we are doing it as a means to some further end. If there are unconditional requirements, incumbent on all rational

beings, then there must be ends that are necessarily shared by all rational beings – objective ends. Are there any such ends?

The ends that we set before ourselves in our ordinary actions, Kant urges, do not have absolute but only relative value: "their mere relation to a specially constituted faculty of desire on the part of the subject gives them their worth" (AK 4: 427). The point here is that most objects of human endeavor get their value *from* the way in which they serve our needs, desires, and interests. Just as technology is valuable because it serves our needs, so pure science is valuable because human beings, as Aristotle says, desire to know; the visual arts and music are valuable because they arouse the human capacity for the disinterested enjoyment of sensory experience; literature and philosophy are valuable because they serve our thirst for self-understanding, and so forth. Although these other things are not mere means like technology, yet still their value is not absolute or intrinsic, but relative to *our* nature. Yet, since we are rational beings, and we do pursue these things, we must think that they really are important, that there is reason to pursue them, that they are good. If their value does not rest in themselves, but rather in the fact that they are important to us, then in pursuing them, we are in effect taking ourselves to be important. In that sense, Kant says, it is a "subjective principle of human actions" that we treat *ourselves* as ends (AK 4: 429).

This suggests that the objective end which we need in order to explain why the moral law has authority for us is "the human being, and in general every rational being." Accordingly, the categorical imperative can now be reformulated as a law instructing us to respect the value of this objective end:

> *So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.* (AK 4: 429)

Using the same examples he did before, Kant proceeds to demonstrate how this principle can serve as a moral guide. Being of absolute value, human beings should not sacrifice themselves or one another for merely relatively valuable ends. Since it is insofar as we are rational beings that we accord ourselves this absolute value, the formula enjoins us to respect ourselves and each other *as* rational beings. We should develop our rational capacities, and promote one another's chosen ends. Respecting someone as a rational being also means respecting her right to make her own decisions about her own life and actions. This leads to particularly strong injunctions against coercion and deception, since these involve attempts to take other people's decisions out of their own hands, to manipulate their wills for one's own ends. Someone who makes you a false promise in order to get some money, for instance,

wants you to decide to give him the money. He predicts that you will not decide to give him the money unless he says he will pay it back, and therefore he says he will pay it back, even though he cannot do so. His decision about what to say to you is entirely determined by what he thinks will *work* to get the result he wants. In that sense he treats your reason, your capacity for making decisions, as if it were merely an instrument for his own use. This is a violation of the respect he owes to you and your humanity.

This example brings out something important about Kant's conception of morality. What is wrong with the false promiser is not merely that he does not tell the truth. What is wrong with him is the *reason* that he does not tell the truth – because he thinks it will not get the result he wants – and the attitude towards you which that reason embodies. Even if he told you the truth, if it were *only* because he thought it would get the result he wanted, he would *still* be regarding you as a mere means. Instead, we must tell the truth so that others may exercise their own reason freely – and that means that, in telling them the truth, we are inviting them to reason together with us, to share in our deliberations. When we need the cooperation of others, we must also be prepared to give them a voice in the decision about what is to be done. This leads Kant to a vision of an ideal human community, in which people reason together about what to do. Because this is the community of people who regard themselves and one another as ends in themselves, Kant calls it the kingdom of ends.