Chapter 6

Kant's Theory of Justice

I

In the waning years of the century during which the utilitarian tradition took shape, Immanuel Kant produced a vigorous and critical response to that tradition that has remained a fertile alternative source of ideas about justice for over two hundred years. Like the advocates of a utility-based conception of justice, Kant wholeheartedly embraced the assumption that all human beings are of equal worth. On other fundamental points, however, he parted ways with Hume and his successors. Most importantly, Kant emphatically rejected the assumption that the promotion of human enjoyment or happiness can ever serve as a foundation for sound ideas about justice. For Kant, the essential truth about human beings – the truth that is relevant to considerations of justice – is that they are free, rational, and responsible agents. The proto- and early utilitarians did not deny that human beings are (at least potentially) free and rational creatures. However, such attributes did not constitute the basis of these philosophers' ideas about justice. For Kant, in contrast, the postulate that human beings are (potentially) free, rational, and responsible is the foundation of all sound ideas about justice and about morality as a whole.

An example Kant offers in his well-known essay on "Theory and practice" is emblematic of his differences with those who base their

A Brief History of Justice, First Edition. David Johnston.

ideas about justice on the concept of utility. Imagine that a person has been made the trustee of a large estate, the owner of which is deceased and the heirs to which are both ignorant of its existence and independently wealthy in their own right, while also being immensely wasteful and uncharitable. Suppose the trustee and his family of a wife and children are in dire financial straits and that the wealth contained in the estate would be sufficient to relieve them of their distress. Finally, assume that the trustee would be able, if he chose to do so, to appropriate the estate for his family's use without the possibility of his appropriation ever being discovered by the heirs or anyone else. It is clear in this scenario that the trustee would be able to increase the aggregate happiness of the concerned parties, taking into account all the heirs as well as all the members of his own family, by withholding the estate from the heirs and appropriating it for the relief of his family. He would be able to enhance the happiness of his family's members greatly, without diminishing that of the heirs by even the slightest measure. Yet Kant suggests that this act of appropriation would be wrong. The trustee has a duty to distribute the estate in accordance with the will of its deceased owner and would violate that duty by directing the estate to anyone other than the intended heirs. (Notice that Kant's reasoning would lead to the same conclusion if the impoverished persons, whose misery might be relieved if they were to receive some share of the estate, were strangers to the trustee.) Despite the tug some might feel to divert the resources in question from their intended beneficiaries in order to relieve human misery, Kant argues that the trustee's duty to distribute those resources in the manner their owner intended should trump the temptation to divert them for the promotion of happiness. This view has been summarized pithily in the observation that, for Kant, the *right* is (ethically or morally) prior to the *good*.

Kant's fame as one of the great modern philosophers was established once and for all with the publication of his *Critique of Pure Reason* in 1781. His major writings in moral and political philosophy came later, beginning with the *Grounding for the Metaphysics of Morals* in 1785 and culminating in his *Metaphysics of Morals* of 1797. During

the nearly two decades he devoted to these writings, Kant elaborated and sharpened his arguments. I shall base my discussion primarily on his writings from the 1790s, including his essays on "Theory and practice" (1793) and "Perpetual peace" (1795), as well as on the *Metaphysics of Morals*.

Kant repeatedly invokes two arguments in rebuttal of the notion that utility can serve as an appropriate basis for reasoning about morality and justice. The first is that any conclusions we might reach by reasoning from the ground of utility would be uncertain. This is the central point of his trusteeship example. Kant argues that the trustee who chooses to decide how to dispose of the estate on the basis of utilitarian consequences would be compelled to estimate the consequences of every possible disposition (for example seizing the estate all at once, using it up gradually, or distributing it to the heirs in the hope that by doing so he could enhance his reputation and ultimately benefit financially from that enhancement) – an exercise that is sure to be inconclusive, leaving the trustee without clear moral guidance. In contrast, he argues, the trustee who chooses to do what (Kant believes) duty requires need have no doubt about the rightful course of action. Even a child of eight or nine, he suggests, can understand how to act in accordance with duty.

Second, Kant also argues that a sound theory of morality cannot be based on happiness, because the causes of happiness vary from person to person, so that only the individual affected is well situated to decide how best to pursue his or her happiness (43 [215]). People must learn from experience what brings them joy, and each person's experience is distinctive to that person. Hence no general (or at least no universal) conclusions about morality can be reached on the basis of happiness – and, in Kant's view, the precepts of morality must by nature be universal, commanding every person in the same way and taking no account of inclinations that vary from person to person. Moreover, Kant argues that it is right for each human being to be allowed to pursue happiness in his own way and wrong to attempt to impose on human beings any particular conception of happiness. He appears to

suppose that it is characteristic of the utilitarian approach to attempt to impose happiness in this way.

Neither argument is compelling. The first assumes that there can be no such thing as a genuine conflict of moral duties. For if such conflicts were possible, the precepts of duty (as Kant conceives it) would sometimes fail to yield unambiguous conclusions about a person's rightful course of action. In that case, the alleged advantage of Kant's doctrine of duty over utilitarian reasoning would evaporate, since the conclusions of the former might be as uncertain as those of the latter approach to moral reasoning. Yet Kant's assumption that no genuine conflict of moral duties can arise seems strained. To borrow one of his own examples: Suppose that a person who has been shipwrecked is clinging to a plank to keep from drowning. Another survivor, who, like the first, is exhausted and certain to drown unless he can find some support to keep himself afloat, grabs onto the plank. Unfortunately, the plank is capable of supporting only one of them. Kant argues that it would be wrong for the first survivor to push the second away from the plank in order to save his own life. His reasoning is that it is an "absolute duty" for me not to take the life of another person, who has done me no wrong, but only a "relative duty" for me to preserve my own life. In other words, I am obliged to preserve my own life only if I can do so without committing a crime (that is what makes this duty a relative duty), and (he seems to suppose) pushing the other victim away from the plank would be committing a crime. Yet it is not evident that Kant's conclusion about this example is correct. Why is my duty to save my own life not equal in force to my duty not to deprive another person of life, when only one of us can live? It seems more plausible to conclude that this case is one of a genuine conflict of moral duties. A central reason for Kant's conclusion to the contrary appears to be his determination that his doctrine of morality should foreclose all possibilities for moral ambiguity, even if the basis for that foreclosure in some cases is less than fully persuasive.

His second argument is problematic – in part because it conflates moral principles, which arguably (and certainly in Kant's view) should

be unequivocal, with policy prescriptions, which by nature (and for reasons he discusses) often cannot be; and in part because it is based on a misunderstanding of utilitarianism. As we have seen, the advocates of a utility-based conception of justice recognized that the causes of happiness vary from person to person. That recognition is the point of Bentham's notion of "idiosyncratical" values and is fundamental to the policies that Hume, Bentham, and many like-minded thinkers advocated. It follows from this recognition that the policy prescriptions that flow from a utility-based conception of justice cannot be fine-tuned to a pitch at which we can be certain that they will maximize aggregate utility. Even if no other obstacles to this ideal were to arise, the amount of detailed information that would be required to achieve it would be far too great to be practicably obtainable. The utilitarians' response to this problem was to support laws and policies that would enhance the opportunities and resources available to individuals, so that they could use those advantages to pursue happiness in their own, often idiosyncratic ways. This response also effectively deflects the force of the second horn of Kant's argument, which appears to be based on the supposition that utilitarian theories of law and policy prescribe the imposition of happiness of a particular kind. The moral principles advocated by utilitarian thinkers (at least by Bentham, who for this reason is largely regarded as the first fully systematic representative of utilitarianism) were unequivocal, just as Kant believed they should be, even if the policy prescriptions that flowed from these principles were not always so. And these moral principles allowed a great deal of room for individuals to pursue happiness in their own ways.

II

Whatever the weaknesses of Kant's criticisms of utility-based ideas about justice, the real interest of his work lies in his alternative to those ideas. Kant argues that the proper basis of morality and

justice is freedom rather than happiness. In order to understand his conception of freedom and the implications for justice that he believes flow from it, we must first peer briefly into the peculiar world of Kant's metaphysics.

In the *Critique of Pure Reason*, Kant had argued that the scope of human knowledge is inevitably limited as a result of the ways in which human beings are (and are not) capable of knowing. Although it is unnecessary and would be unwise to attempt to summarize the argument of this famously esoteric work here, it is important for our purposes to note that, in the course of that argument, Kant develops a distinction between two fundamentally different ways of knowing. The first is the kind of knowing we can have of objects, or possible objects, as they appear, or could appear, to us. We can call the kind of knowledge we gain from this way of knowing *phenomenal knowledge*. (Phenomenal knowledge is roughly the same thing as empirical knowledge: the kind of knowledge we acquire through observations and experience of the world.) Kant argues that all our phenomenal knowledge is shaped a priori by certain universal (and hence inescapable) attributes, which he calls "categories." For example, whenever we conceive of anything in the world (i.e. in the universe), we conceive of it as being situated in space and as having spatial properties. (Even a point that we imagine to occupy no space has the spatial properties of occupying no space and of being located at a particular position in space.) In his view, everything that appears to us does so necessarily in a spatial way. Similarly, everything that appears to us necessarily does so in some relation to time: it has some temporal property. Kant also argues that all phenomenal knowledge – all our knowledge of things as they appear to us – is informed by certain categories through which we associate and disassociate things with one another. For example, everything of which we conceive appears to us to be related to other things by way of causality. Even when we do not know what the causal connections between things are, as we often do not, we think of things as causally connected with other things and we cannot really conceive of them otherwise.

The second way of knowing proposed by Kant is the knowing of things as they are in themselves, that is, as they might be known if they could be shorn of their phenomenal attributes. Kant calls this way of knowing noumenal, after the Greek noun *nous*, which designated intelligence, intellect, or mind, and the related abstract name of action *noesis*, which Plato uses to label the highest and truest form of knowledge.

Plato gives his readers the clear impression that *noesis* is accessible to human beings, though only to those few who possess a philosophical nature and in whom this nature is cultivated to the highest point. For Kant, in contrast, knowledge of things as they are in themselves – *noumenal knowledge* – is inaccessible to humans. We can imagine that this kind of knowledge is possible for some kind of being, but we cannot obtain it and cannot even know (in any strict sense of knowing) that it is in principle accessible to any kind of intelligence at all. The notions of space and time, as well as categories like that of causality, are inherent in the way human beings are able to know. We are not capable of transcending the limitations imposed by those notions and categories; or, at least, we are not capable of obtaining knowledge that transcends them.

According to Kant, human beings have an intense practical interest in reasoning about three things about which we are incapable of obtaining phenomenal (or empirical) knowledge. Those objects are freedom of the will, immortality of the soul, and the existence of God. The first of these objects is central to Kant's theory of justice.

In Kant's view, we can neither prove nor have certain knowledge that human beings possess free will. However, we can show that morality makes sense only if human beings are free. On this basis, we can reasonably *postulate* that human beings are free. And, on the basis of this postulate, we can reason extensively about the content of morality and justice. Through this route we can use reason to discover *laws of freedom*, as Kant calls them, laws that prescribe to us what ought to happen and what our duties are, as contrasted with *laws of nature*, which merely help us to explain what actually does

happen in the world. This line of reasoning leads to the "doctrine of duties," in which

> Man can and should be represented in terms of the property of his capacity for freedom, which is wholly supersensible, and so too merely in terms of his *humanity*, his personality independent of physical attributes (*homo noumenon*), as distinguished from the same subject represented as affected by physical attributes, *man* (*homo phaenomenon*). (65 [239])

In other words, we must reason as if we knew that human beings, as they are in themselves, beyond the reach of our phenomenal knowledge, are free agents. And, although I shall not reproduce his line of argument here, for Kant it also follows that we must reason on the assumption that human beings are rational agents. Freedom and rationality are the attributes that form the basis of all moral reasoning; for, in the absence of these attributes, moral reasoning makes no sense.

Kant's theory of justice, then, is based on the same dualism that underpins his entire metaphysics. It is worth noting that this dualism between *homo phaenomenon* and *homo noumenon* is very much like the dualism between body and soul that has played a central role in Christian thinking since its earliest years. The body is the visible self; the soul is the invisible self in which the true personhood of men and women resides. In Christian thought, the soul is by far the more important partner in this pair. Similarly, in Kant's thought, *homo noumenon* plays by far the larger role. The non-physical ("supersensible") attributes of *homo noumenon* are the basis of Kant's theory of justice. Kant was a Protestant Christian, so it should be no surprise that some of the assumptions, concepts, and distinctions that underpin his theory of justice can be found within the large family of Christian thought.

As we have seen, the key postulate on which Kant bases his theory of morality, of which his theory of justice is a part, is that man, viewed as *homo noumenon*, is free. Because the subject of this statement is man viewed as *homo noumenon*, the statement is not an empirical one. In other words, it is not a statement about some attribute of man that

we can discover, prove, or disprove through observation. It is rather a statement about the (postulated) essential nature of man. It is also a normative statement, a statement about what ought to be. To say that man viewed as *homo noumenon* is free is, in part, to say that man ought to be free, that man is entitled, or has a right, to be free.

Kant's conception of freedom lies at the heart of his theory of justice. In order to understand that theory, therefore, it is important to grasp that Kant does not endorse the commonplace notion of freedom as lack of constraints on one's actions. Instead, he defines freedom as subjection to no other laws than those which a person gives to himself, either alone or along with others (50 [223]). For him, to be free is not to lack constraints on one's actions, but to be independent of the constraints imposed by the arbitrary wills of others (63 [237–38]).

Moreover, because this idea of freedom is based on man viewed as *homo noumenon* – on the postulated, but not strictly knowable, essential nature of man – the empirical differences that distinguish one person from another have no bearing on the implications of this idea for justice or for the rights of persons. Aristotle seems to have based his assumption that human beings are categorically unequal to one another by nature on the observation that, as an empirical matter, people differ dramatically in capabilities. Hobbes, Hume, and Adam Smith based their claims about the equality of human beings on the assertion that people in fact are roughly equal in capabilities, at least if we set aside the impact of society and differences in education. From a Kantian point of view, observations about people's capabilities or other empirical attributes are irrelevant to matters of rights and justice. From this point of view, each person possesses absolute worth, and does so in equal measure with all other human beings.

III

Kant builds his entire moral theory, including his theory of justice, on the foundation provided by the postulate that man viewed as *homo*

noumenon is free. He argues that, by reasoning from that postulate, we can arrive at a single supreme principle of morality, which he calls the "categorical imperative" (CI). The categorical imperative is an *imperative* in the sense that it is a command, in effect an order that describes for people what they can and cannot (should and should not) do. It is *categorical* in the sense that it applies to every person (indeed, to every creature who is free and rational), whatever aspirations, intentions, or objectives that person might have.

We can contrast the categorical imperative with (what Kant calls) "hypothetical imperatives." An imperative, or command, is hypothetical if its applicability to persons is contingent on the particular aspirations, intentions, or objectives they happen to adopt. (A more apt name might be "conditional imperative.") If I acquire the aspiration to become a virtuoso violinist, then it is a hypothetical imperative for me to pursue the means necessary to achieve this aspiration by taking lessons, practicing, and the like. I am under no moral obligation to adopt or to realize this aspiration and many other people do not share it. So this particular hypothetical imperative does not apply to them and applies to me only insofar as I maintain my aspiration. In contrast, categorical imperatives apply with equal force to every person. And they apply with greater force than hypothetical imperatives. Should any conflict arise that would make it impossible for me to obey both a hypothetical imperative and the categorical imperative in a given instance, the categorical imperative takes priority.

Although Kant formulated the categorical imperative in several different ways, he argues that, in reality, one and only one categorical imperative exists. That imperative (in one of its formulations) is: "Act only according to that maxim whereby you can at the same time will that it should become a universal law." According to Kant, the whole theory of justice is derived from this single command.

It has often been asserted that Kant's categorical imperative is a version of the "Golden Rule," which is commonly formulated in the statement: "Do unto others as you would have them do unto you." Certainly the similarities between the two are considerable. Both

statements are categorical, in other words unconditional, commands. Both are reflexive, in the sense that both demand that the person to whom they are addressed put himself or herself in the place of another and consider whether a contemplated action would be acceptable to him- or herself in that hypothetical circumstance. Both statements are intended to apply universally to all human beings.

Nevertheless, Kant's statement differs from the Golden Rule as we find it in the Gospel according to Matthew. The Golden Rule is intended to apply to discrete actions. It demands from its addressee to consider how he or she would want others to act toward him or her. Although the categorical imperative also applies to discrete actions, it does so through the intermediary of maxims (maxims are principles or rules of action individuals adopt in pursuit of the objectives, purposes, and projects they happen to choose), and it asks us to judge maxims not on the basis of what we would want to happen to us, but by considering whether we could will that our maxims should become universal laws. The process of reflection Kant demands that we undergo resembles the process commanded in the Golden Rule, but it is more complex, more abstract, and more generalized.

The differences between Kant's formulation of the categorical imperative and the statement of the Golden Rule in the Gospel according to Matthew are significant in the context of Kant's theory of justice. We can gain an inkling of these differences by considering another well-known passage from the Sermon on the Mount – the same speech in which Jesus offers his statement of the Golden Rule. Shortly before he articulates that rule, Jesus observes:

> You have learned that they were told, "An eye for an eye, and a tooth for a tooth." But what I tell you is this: Do not set yourself against the man who wrongs you. If someone slaps you on the right cheek, turn and offer him your left. If a man wants to sue you for your shirt, let him have your coat as well. If a man in authority makes you go one mile, go with him two.

A central message of the Sermon on the Mount, including the Golden Rule, is that it is unjust to do harm to others. This message is strikingly similar to Socrates' argument in the *Republic* that it can never be just to do harm to others, an argument he deploys in order to undermine the several versions of the idea of justice as reciprocity with which he is presented in the opening arguments of Plato's work. The notion that reciprocity is fundamental to justice is as alien to the Sermon on the Mount – and to the message of the Gospels as a whole – as it was to Plato in the *Republic*. Yet, as we shall see, the concept of reciprocity plays an integral and vital role in Kant's theory of justice.

In the formulation cited above, the type of object to which the categorical imperative is intended to apply is ambiguous. The most obvious type of object of the CI's commands is an action. But the CI commands that a person determine whether an action is permissible or not by reflecting on the maxim behind that action, so that, indirectly at least, the CI seems to apply to maxims (principles, rules of action, or types of action) as well. In Kant's moral theory as a whole, the CI is in fact the basis for commands – for moral laws – that apply to both types of objects. The distinction between these two types (actions on the one hand and maxims on the other) is fundamental to a division between the two major branches of his moral theory. As applied to *maxims* (and to the ends or objectives at which those maxims aim), moral laws are called *ethical* laws. Ethical laws constitute prescriptions about the range of intentions and objectives we can rightfully adopt. As applied to *actions*, moral laws are called *juridical* laws (42 [214]). Juridical laws place limits on people's conduct, not on their intentions or objectives. Kant's moral theory as a whole encompasses both types of laws. His theory of justice, however, is concerned solely with juridical laws and with the external actions that can be controlled through them.

Kant believed that it is impossible to force people to adopt intentions. (It also appears to follow from his understanding of freedom that it would be wrongful to force people to adopt intentions, even if it were possible to do so.) For this reason, duties of benevolence, which can be

fulfilled only by a person who maintains intentions that are appropriate to those duties, are subject to ethical laws, not to juridical laws (i.e. laws of justice). However, it does not follow that intentions are irrelevant to Kant's theory of justice (as distinct from his ethical theory). According to that theory, public laws may *prohibit* people from acting with certain intentions, even if they may not *require* people to act with any prescribed intentions. For example, a prohibition on premeditated murder according to which that offence is distinguished from negligent homicide by the presence, in the offender, of the intent to commit the act would be consistent with Kant's theory of justice. Kant's theory of justice is designed to apply to actions and only to actions, but intentionality is integral to the description of some actions (such as premeditated murder), and the laws of justice may be directed at these kinds of actions as well as at those to which intentionality is not integral.

IV

Kant's theory of justice, then, is a theory of the moral laws or laws of freedom that place limits on people's external actions, limits that can be coercively enforced. The basis of this theory is the universal principle of right, which is derived from the categorical imperative and which Kant formulates as follows:

> Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law. (56 [230])

As Kant quickly suggests, a major point of this principle is to justify the use of coercion to prevent people from hindering the freedom of others.

It is widely believed that Kant thought that legal coercion can be justified *only* to secure freedom. Notice, however, that this is not what

his universal principle of right says. It is certainly true that, according to that principle, a coercive action (such as an action to enforce a coercive law) must be *compatible* with everyone's freedom in accordance with a universal law. Yet it does not follow that the only permissible *purpose* or objective of such a coercive action is to secure freedom. It is also important to bear in mind that Kant did not equate freedom with a lack, or even a minimization, of constraints on people's actions. In his view, a person is free if that person is subject only to laws he or she gives to him- or herself, either alone or together with others. Moreover, the relevant lawgiver in his understanding is *homo noumenon* – not *homo phaenomenon*. The person who gives laws to him- or herself is the person shorn of his or her physical attributes, including his or her individualized desires and inclinations, not the person who is laden with those attributes. A person's freedom is not curtailed if that person, as an empirical self – as *homo phaenomenon* – is subject to laws given to him by his "supersensible" self – by *homo noumenon* – nor if those laws are imposed by a decision in which others participate.

Kant draws upon the basic principle of his theory of justice – the universal principle of right – to reach conclusions about two types of subject matters. The first has to do with discrete relations among persons. His label for this subject matter as a whole is *private right*, within which he includes the now familiar subjects of property, transactions, and contracts. He also includes under this heading a discussion of rights over persons. ("Rights over persons" include the rights of a man over his wife, the rights of parents over their children, and the rights of a head of household over its servants. Although Kant believed that all human beings possess absolute worth in equal measure with all others, he also assumed that it is natural for some persons to occupy positions of superiority to others within the family.) The second subject, *public right*, has to do with the civil condition, that is, the state (or civil society, as it had been called in a usage that would soon become anachronistic).

These two subjects are inextricably intertwined. For example, Kant distinguishes between possession and property and argues that

property, which is one of the principal components of private right, is possible only in a civil condition in which claims to property rights can be recognized and enforced. Even though he discusses private right before going on to the subject of public right, for Kant there can be no private right outside the civil condition, that is, outside of a state with coercive powers through which the private rights of the citizens can be enforced.

For Kant, a just society is one whose members reciprocally respect each other's rights by refraining from violations of them. Like Hume and his successors, Kant mounted a vigorous defense of the right to private property. He recognized that people acquire property because they expect it to be useful to themselves. But Kant's argument for the rightfulness of private property does not rest on the claim that the institution of private property is useful. For him, that right is based on the freedom that is inherent in human beings as *homines noumena*. The postulate that is fundamental to moral theory – without which, in Kant's view, morality makes no sense – is that human beings possess free will. To say that human beings possess free will is to say that our decisions and actions are not inexorably and exclusively caused by empirical inclinations and desires. It is to say that we are by nature capable of subjecting those actions to our wills, in accordance with laws of freedom. And, just as we are capable of subjecting our actions to our wills, we are also capable of asserting our wills over things. The right of private property is justified by the capacity of human beings to assert their wills over things.

Kant argued that the rights of human beings are rooted in the original right to freedom, which belongs to every human being by virtue of his or her humanity. He also argued that all human beings are inherently equal, in the sense that all are entitled not to be bound by others any more than others can be bound by them (63 [237]). Kant denies, however, that this inherent equality entails a right to equality in possessions. He asserts that all subjects of a state are entitled to being treated as equals by the laws, so that none should receive special privileges or be subject to unfavorable discrimination in legal matters.

He also asserts that every member of a state should be within his rights to compete for whatever positions of privileged status a society may offer, and he explicitly criticizes the institution of hereditary aristocracy. But Kant firmly defends inequality in possessions, among which he includes physical and mental advantages and skills, as well as material possessions in the more usual sense.

Despite Kant's assumption that some human beings are suited, by virtue of their "physical" attributes (their attributes as *homines phaenomena*), to occupy positions of superiority over others for the purpose of decision-making within the family, and despite his defense of inequality (even "the utmost inequality") in possessions, the concept of reciprocity is central to his theory of private right. Indeed, the most fundamental theme of that theory is that just relations among persons who are equal by nature (which, for him, means all persons) are relations of balanced reciprocity in which the relevant point of reference is *homo noumenon*, not *homo phaenomenon*. Here is one representative passage from "Theory and practice":

> Man's *freedom* as a human being, as a principle for the constitution of a commonwealth, can be expressed in the following formula. No-one can compel me to be happy in accordance with his conception of the welfare of others, for each may seek his happiness in whatever way he sees fit, so long as he does not infringe upon the freedom of others to pursue a similar end which can be reconciled with the freedom of everyone else within a workable general law – i.e. *he must accord to others the same right as he enjoys himself.*

Kant's emphasis on reciprocity among persons considered as bearers of free will is in fact evident in his formulations of the categorical imperative and of the universal principle of right. When he highlights universality in those formulations, he is at the same time highlighting reciprocity among persons considered as possessors of the capacity for freedom.

Nowhere is Kant's emphasis on reciprocity plainer than in his theory of punishment, which he discusses under the heading of public right.

Here as elsewhere Kant draws attention to the differences between utilitarian reasoning and his own approach to justice. "Punishment by a court" (here he includes judgments reached against persons for civil wrongs, as well as punishments imposed for criminal violations), he asserts, "can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*." The clear target of criticism here is the fact that utilitarian reasoning about justice could lead to the punishment of an innocent person for the sake of the greater good. It might be argued that this concern is directed at a possibility that is merely hypothetical, since utilitarians have not generally defended the punishment of the innocent. But Kant is equally concerned that utilitarian reasoning can lead to punishment that is insufficiently severe to constitute a balanced response to criminal wrongdoing. Indeed he aims some of the most forceful statements in his entire political philosophy at this target:

> The principle of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaemonism [the view that the objective of producing happiness should be at the basis of the principles of morality] in order to discover something that releases the criminal from punishment or even reduces its amount [. . .]. For if justice goes, there is no longer any value in men's living on earth. (141 [331–32])

As a matter of justice, Kant believed that we should be as concerned about punishing the guilty too lightly as about inflicting undeserved punishment on an innocent person for the sake of an ostensible greater good.

What kind and quantity of punishment does justice demand? Kant's answer is unequivocal. The principle of punishment is

> [n]one other than the principle of equality (in the position of the needle on the scale of justice), to incline no more to one side than to the other. Accordingly, whatever undeserved evil you inflict upon another within

> the people, that you inflict upon yourself [. . .] only the *law of retribution* (*ius talionis*)[. . .] can specify definitely the quality and the quantity of punishment. (141 [332])

Kant places himself squarely in the camp of those who hold that the concept of reciprocity forms part of the bedrock on which the idea of justice is based. Moreover, it is clear from his arguments about punishment that he is a firm adherent to the notion of strictly balanced reciprocity. Although he does not insist that punishments inflicted on perpetrators of wrongs should always be identical in kind to the wrongs they have imposed on their victims, his theory of punishment is a very near relation to the biblical teaching of an eye for an eye, a tooth for a tooth.

Ancient writings widely endorse the notion that justice is a matter of balanced reciprocity among equals and of imbalanced reciprocity among unequals. Although Kant accepted inequalities of legitimate power as natural and was a fierce defender of inequalities in possessions (including physical and mental powers, as well as external goods), he adhered closely to the view that the relevant point of reference for thinking about justice is *homo noumenon*, not *homo phaenomenon*, the person conceived as a possessor of the capacity for freedom and as a bearer of rights rather than the individual laden with physical (including psychological) attributes. As possessors of the capacity for freedom, all human beings are equal. For Kant, then, the principle that should underpin punishment for all persons, regardless of rank, should be the principle of balanced reciprocity. Even if the punishment of people who are superior in rank cannot always be identical in kind to the punishment of those who are their inferiors, Kant argues that the punishment that is meted out to privileged persons should be equivalent in its effects to that imposed on ordinary persons (141 [332]). In his view, punishment that is too light – no matter for what reason, whether it be consideration of an ostensible greater good or regard for a person's social rank – is as serious an injustice as punishment that is too harsh (or imposed on an innocent person).

Kant reserves some of his most severe criticism for opponents of capital punishment, singling out Cesare Beccaria by name. Beccaria, he argues, was "moved by overly compassionate feelings of an affected humanity" to indulge in arguments that are nothing but "sophistry and juristic trickery" (143 [334–35]). A person suffers punishment because "he has willed a *punishable action*" (143 [335]). Beccaria's error is to fail to distinguish between *homo noumenon*, the self of pure reason who legislates in accordance with the universal principle of right, and *homo phaenomenon*, the self with physical attributes, including impulses and inclinations, which often lead to transgressions of laws and of the rights of others. Capital punishment is the penalty willed by the rational self in accordance with strict retributive justice on any person who wills and carries out the act of murder. While he wavers occasionally on points of detail, Kant never wavers in his resolution that the principle on which punishment should be based is the principle of balanced reciprocity.

V

By discussing the topic of punishment, we have already dipped into the domain Kant calls public right. Kant was emphatic that punishment can be imposed justly only by a public body, namely the state, even though most of the violations for which punishment is imposed occur in discrete relations among persons (which are the focus of private right). Here as in many other areas of his theory of justice, private right in his view is functionally dependent on public right.

Kant believed, in other words, that it is possible to maintain justice in relations among persons only by entering into the civil condition – that is, by joining with others in a commonwealth (or state). He maintained that human beings are under an absolute duty to enter into that condition:

> a union as an end in itself which they all *ought to share* and which is thus an absolute and primary duty in all external relationships whatsoever among human beings [. . .] is only found in a society in so far as it constitutes a civil state, i.e. a commonwealth.

So vital, in fact, is the civil condition to justice that anyone who is inclined to avoid membership in a commonwealth and remain in a "natural," pre-political state can rightly be compelled or forced to join the commonwealth.

Kant's main argument for this conclusion is that only by constituting a collective (or general) will backed by great coercive power can people be assured that others will respect their rights. Prior to the creation of such a power, each person possesses the right to do what seems right and good to that person. However, in this pre-political condition, each person is also exposed to the possibility of being constrained by the arbitrary will of others. So the first thing any set of people must do is to join with others to create a state that possesses sufficient power to enforce the rights of its citizens. The only possible kind of just society is a just state.

Kant also maintains that resistance of any kind to the legislative authority of a state, under any circumstance, is absolutely contrary to justice (130–131 [319–320]). As a matter of justice, the subjects of a state owe absolute obedience to its sovereign. His reasoning is that there can be no justice without a state, and that any rebellion, sedition, or resistance to a state constitutes a threat to its very existence, and hence a threat to justice. Writing in a period when the turmoil unleashed by the French Revolution had not yet settled, he does not seem to have considered the possibility that a political regime may be constituted in such a way that it can actually be strengthened by some forms of resistance to public laws and policies rather than being threatened by them.

These claims about the domain of public right – that people can and should justly be forced to join a state if they are reluctant to do so, and that the subjects of a state owe absolute obedience to its

ruler – are striking in a political philosophy that is based on the ideas of freedom and of the absolute worth of each person. Here as elsewhere, it is important to bear in mind Kant's distinctive conception of freedom as subjection to no laws other than those that the individual gives him- or herself, either alone or along with others. Freedom is *not* "a warrant to do whatever one wishes unless it means doing injustice to others." Although Kant believed strongly that each person possesses absolute worth in equal measure with all others, his conception of freedom is far more social in character than many readers of his work have realized.

Kant asks us to think of the civil condition or state as the product of an "original contract" agreed upon by those who become members of that state. For him, this contract is an "idea of reason" rather than an empirical or historical fact – much as the idea of *homo noumenon* is an idea of reason – but one that, in his view, is of great practical import. His writings on public right show that, while a large part of the point of the civil condition is to protect the rights of individuals against each other, the terms of the agreement that underpins that condition are much more expansive than this formula might seem to convey.

Kant is firmly opposed to a paternalistic state, by which he means a state that treats its subjects as if they were children incapable of discerning what is useful and what is harmful to them. A state of this kind, however benevolent its intentions, is in his view "the greatest conceivable despotism," for it denies its citizens the fundamental human right to seek happiness in whatever ways they see fit, as long as in doing so they do not infringe on the freedom of others to do likewise. He also appears to be opposed to the redistribution of wealth for the purpose of achieving equality in possessions, since he argues that the equal treatment to which all persons are entitled under the laws of a state is entirely consistent with great inequality in possessions. These observations have led some readers to conclude that Kant was an advocate of a minimal state, one that should do little more than provide for a common defense and enforce personal rights, property rights, and contracts.

The truth is that Kant advocated a state that is much more robust than the minimal state. He did so on the basis of the idea of an original contract, which is the central idea of his theory of justice in the domain of public right. While it may be unjust for a state to redistribute wealth among its citizens for the purpose of equalizing their possessions, the state is not merely allowed, but required as a matter of justice to redistribute wealth whenever that action is necessary for the purpose of meeting needs (136 [326]). Kant is clear that the state can justly fulfill its duty to provide for those members of society who are unable to maintain themselves by transferring possessions from the wealthy to the poor. He also maintains that, as a matter of justice, the state should effect this transfer of wealth through coercively imposed taxation, explicitly ruling out the possibility that the needs of the poor can justly be met through a program of voluntary contributions. Far from supporting a minimal state, Kant argued that a just state is one that ensures that the needs of all its members, including those who are unable to provide for themselves, are met through coercively imposed measures requiring the wealthy to contribute a portion of their possessions to meet the needs of others.

Kant's reasoning is direct. The wealthy owe their very existence, and a fortiori their wealth, to the state in the sense that without it they would be unable to live, let alone prosper. They are therefore obliged, in return for these benefits, to contribute as necessary to the well-being of their fellow citizens (136 [326]). Their obligation to help support the poor is based on a principle of reciprocity.

Kant's conclusion is implicit in the idea of the original contract. If the theme of his theory of private right is that just relations among persons are relations of balanced reciprocity in which the relevant point of reference is *homo noumenon*, the theme of his theory of public right is that the idea of the original contract – a hypothetical agreement in which the members of a state assume obligations toward their fellow citizens in return for the assurance that their own rights and needs will be secured – is the principle behind all public rights, the test of whether public laws and policies are just or unjust.

For Kant, the idea of the original contract is the vehicle for determining whether or not laws and policies are just, much as the categorical imperative is the test for ascertaining whether or not individuals' maxims and discrete actions are rightful. If a law is such that a whole people could not have agreed to it in an original contract, then that law is unjust. If, on the other hand, a law is such that it could have been the object of such an agreement – an agreement to which an entire people might have given its assent – then it is at least arguably just (and Kant believes that it is the people's duty to consider such a law to be just even if they disapprove of it). A set of laws that would allow some members of a state to be deprived of the means required in order to meet their needs is a set of laws from which at least some people would have withheld their consent in an original contract. Such a set of laws would therefore be unjust. More generally, *any* law or policy that could not have commanded the assent of the entire people in an original contract is unjust. Kant did not believe that resistance to the ruler of a state could ever be just, even if that resistance is designed to oppose unjust laws. But he *did* believe that laws are sometimes unjust and that the idea of the original contract supplies an intellectually rigorous test for determining whether or not they are just.

VI

In contrast to the utilitarian writers, Kant placed the concept of reciprocity squarely at the center of his theory of justice. In the domain of private right, which is concerned with discrete relations among persons, he endorsed the notion of balanced reciprocity among equals, and, since he considered all persons, regarded as *homines noumena*, to be equal, he viewed balanced reciprocity as *the* basis of just relations among private persons. In the domain of public right, Kant's view is more difficult to categorize. It is clear that the concept of reciprocity plays a large role here as well, in the sense that all those

who enjoy the benefits of the civil condition must in return assume obligations toward their fellow citizens – obligations that, under some circumstances, would compel them to give up some of their wealth to supply the needs of others. Yet it is difficult to characterize precisely the notion of reciprocity Kant deploys here. The concept of *balanced* reciprocity in a well-defined sense seems not to apply.

Kant had opened up new territory for thinking about justice. While Adam Smith had attained a vision of society as a systemic whole, with properties and products that are best accounted for by the whole rather than by its parts taken severally (this is the thrust of his conception of the division of labor and its consequences), he did not frame a distinctive conception of justice around this vision. Kant did precisely that. By reworking the already familiar notion of an original contract into a test for the legislation of a commonwealth viewed systemically, he had stepped into a new world, perhaps no more knowingly than Christopher Columbus in his first voyage to the Americas. Although the phrase "social justice" had not yet been coined, Kant had hit upon features that have remained essential to the concept of social justice for the past two centuries.

Kant bequeathed to posterity a vision of a just society that was self-consciously at odds with the vision defended by the theorists of utility. He offers a window onto that vision in this passage from the *Metaphysics of Morals*:

> By the well-being of a state must not be understood the *welfare* of its citizens and their *happiness*; for happiness can perhaps come to them more easily and as they would like it to [come] in a state of nature (as Rousseau asserts) or even under a despotic government. By the well-being of a state is understood, instead, that condition in which its constitution conforms most fully to principles of Right; it is that condition which reason, *by a categorical imperative*, makes it obligatory for us to strive after. (129 [318])

A just society, for Kant, is one whose principal aim and tendency is to maintain social relations of mutual respect and reciprocity among free

and equal citizens, not to enhance their welfare, conceived in material terms. His conception of this society was flawed, to be sure. Viewing human beings in highly abstract terms, as *homines noumena*, Kant underestimated the extent to which the quality of relations among persons is necessarily a hostage to their relative circumstances. Further, his conception of private right rooted in strictly balanced reciprocity stands in some tension with his conception of public right, based as it is on the idea of an original contract, which cannot be reduced to the terms of balanced reciprocity. Yet Kant had staked out the territory, if not the terminology, of social justice, and he had done so with a vision that transcended the class struggles that were destined to turn that territory into a battlefield throughout the nineteenth century and beyond.