

3: Kant on Animality, Rationality, and the Moral Constitution of the Self

Ancient and Medieval Ethics of Nature

The preceding chapter provides a phenomenology or description of ordinary moral experience. Philosophical theory should attempt to reflect on this experience and produce a coherent conception of morality, a theory of morality. We can measure the validity of proposed ethical theories by their adequacy in reflecting the nature of actual moral experience. In this perspective, a major discrepancy immediately looms when we consider the dominant moral theories. In arguing that morality is not based, directly at least, on the desire for happiness, Kant implicitly criticizes the dominant “eudaimonistic” (*eudaimonia* is Greek for happiness) ethical theories of his time. These theories, whether based on reason or feeling, conflict with the phenomenology of morality that has just been set forth. They disregard the central importance of the experience of duty.

For the “natural law” theory of Aristotle and Aquinas moral experience is the expression of inner natural drives, expressed in feelings and desires and confirmed by reason, arising out the individual’s human nature. In this classical version of eudaimonistic ethics, the natural drives that motivate individuals are thought to be inherently social, and so the individual is regarded as a social or political being by nature. Following one’s inclinations and striving to fulfill the natural drive for happiness is therefore innately in accord with the requirements of life in state-organized society. Since to do the right thing, to do what morality requires, means fundamentally to do what you truly *want* to do, there is here no place here for a conception of morality as centering on the experience of duty.

There is however an apparent substitute for duty. Feelings can and often do mislead us when they are focused on short-range objectives rather than one’s long-range goals. In the eudaimonistic tradition, ethical reason is needed to prevent or correct such distortions of the basic natural tendencies. Reason can discover in the drives or inclinations the fundamental norms or rules of life that result in happiness for the individual and in harmony for society as a whole. Such rationality, which is ultimately in harmony with basic natural inclinations, can further promote the appropriate feelings and check any wayward, discordant, or short-term desires that conflict with true human happiness. In this checking of the discordant or irrational short-term desires, there is an approximation of duty. *I ought to* resist my short-term, destructive desires with their illusory conception of happiness for the sake of my realizing authentic desires, as revealed by a fully rational understanding of human nature. Here is a kind of duty to do what *really* makes one happy as opposed to what only *seems* to offer happiness. Essentially, however, we do not sacrifice happiness for the sake of duty, but only aim at doing what makes us “truly” happy. Morality is therefore ultimately a matter of rational self-interest for individuals who allegedly fulfill their natures, and so achieve happiness, when they cooperate with others in a rational arrangement of society.

In this perspective, where morality is based on tendencies inherent in human nature, the role of morality is not one of correcting or contradicting the existing order of things in the name of the moral “ought,” but of adhering to *what is* at the level of the nature or essence of the human being. A social order comes into being through the development or unfolding of a natural process, just as a fully grown plant or animal expresses its inherent nature. Aristotle recognized that not all societies are at the same degree of maturation. Some are more primitive or immature than others, like the barbarian societies that follow kinship rules. But his own Greek society, where city-states like Athens are governed by wise laws, has reached the peak of natural maturity for human beings. And so it is possible to discern the requirements of nature simply by observing the general state of the existing, mature society. A truly rational society, Aristotle therefore argues, is, like the society of his time, a slave society. The life of leisure that permits the development of the highest human abilities presupposes servitude on the part of many people. But this is only in accord with the diverse natures of human beings. Some people are naturally slavish, and so they fulfill their natures, and achieve happiness, by living under slavery. For the slave, happiness consists in following the orders of his master. Women too are subordinate members of society. Happiness for a woman involves fulfilling her nature by submitting to her husband, and finding her place in the home. She truly *wants to be* excluded from public life, which is the natural domain for the free male members of society.

In the previous chapter, we saw that Kant recognizes an indirect duty to one’s happiness in certain circumstances where short-term pleasures conflict with long-term happiness. For the most part however, our natural drive for happiness tends to conflict with morality. Following the dominant trend in modern ethics, Kant does not regard the drive for happiness as the basis of a harmonious social and political life. The individualism of modern society is expressed in the widespread understanding that the desire for happiness individuates and separates individuals, and so could not possibly be the basis for social harmony. No universal and necessary law can be derived from the individual’s idea of happiness, according to Kant, for this idea is ultimately quite personal, as well as being vague and elusive.

The Political Ethics of the Social Contract

For the social contract theories and utilitarian theories of modern times, the subject of rational ethical investigation is a naturally asocial, egotistical individual. While for Aristotle the individual is a “*zoon politicon*” or political animal, for Hobbes in the seventeenth century the State is not a natural feature of human existence but the artificial creation of individuals seeking the best means for realizing their individualistic, self-interested goals. Implicit in Hobbes’ social contract theory is the conception of society as the product of separate and independent individuals, rather than as a naturally existing organic whole. This individual exists first of all outside of society, in a “state of nature.” By contrast to ancient and medieval natural law theory with its conception of an inherently socializing nature, in this ethical scenario nature seems more straightforwardly natural. Nature and the natural inclinations that express nature, we often suppose, are wild and unpredictable, dangerous and impulsive drives that need to be reined in if life in society is to be possible. The pleasure-driven “id,” says Freud, must be restricted by the “super-ego” of societal demands. If our natural inclinations are polymorphously perverse, some non-natural checking or controlling agent is required to rein in and restrict our natural impulses,

and this is the function of morality. Morality is no longer in harmony with basic natural inclinations but tends to take on the more forbidding face of a hard duty that opposes natural desires.

What is the basis or source of such duty? In his apostrophe to duty cited at the end of the previous chapter, Kant raises this question: “what origin is worthy of you, and where is the root of your noble descent which proudly rejects all kinship with the inclinations and from which to be descended is the indispensable condition of the only worth which men alone can give themselves?” Kant raises this question because he has already excluded natural inclinations as the basis of such guidance. For the eudaimonistic ethical theories before Kant, it is however nature itself that provides the basis for limiting nature. That is, it is more enlightened, more long-range view of what nature requires that must provide guidance for a less enlightened, more short-range understanding of what makes one happy. Modern ethical theories however suppose an individualistic, egotistical nature. For the modern ethicist following the path of Hobbes, it is necessary to derive rules of a more harmonious social life *on the basis of one’s* individualistic, essentially asocial nature. If society is thought to be made up of naturally asocial or egotistical individuals, morality consists in following an enlightened understanding of individualistic self-interest that requires smoothing over the rough edges of nature, but not in contradicting it. This modern society is clearly less naturally harmonious than that of ancient Greek or Medieval Christian thinkers. But it is still one in which nature is regarded as sufficiently tame and reasonable that rules of social life can be contained within or founded upon it. Again, morality does not consist in following duty against desire, but in a more educated, well-thought-out understanding of how to be as happy as possible, as against the pressures of unenlightened, short-term desires.

For the social contract ethics of Hobbes each individual innately seeks the maximum amount of happiness for himself alone. Natural freedom—i.e., freedom in the state of nature that precedes the existence of rational political society—consists in the right to do whatever it takes, including murder and treason, to achieve the goal of personal happiness. Natural tendencies not only are largely different for different individuals, but even when they are the same, they produce conflicts, as hungry and vulnerable people will fight to possess limited resources. The result of unrestricted natural liberty to pursue one’s desires is foreseeable and comprehensible: life in such conditions, Hobbes concludes, is “solitary, poor, nasty, brutish and short.”¹ The rational individual eventually realizes through bitter experience that it is in his interest to accept limits to such natural freedom. Enlightened morality consists in recognizing that artificial social peace, not the natural state of war of all against all, is the necessary condition for the maximum realization of happiness for each individual. Enlightened morality recognizes that each person should follow the Golden Rule, and treat other individuals as he would want to be treated himself. For this is the best way to realize one’s *own* desires. This and other moral rules follow from enlightened self-interest.

But when one person follows such rules, while others do not, the ultimate ground of morality, which is the individual’s own survival, is violated. And since each individual suspects the other of a willingness to violate the moral rules for the sake short-term benefits, no rational person should follow the rules of morality simply because they are rational. Only a fool would help another person who was out to kill him. Morality by itself therefore turns out to be impotent. We

thus find ourselves in the paradoxical situation in which it is immoral to follow the requirements of morality. Morality by itself is powerless. This moral dilemma leads to the moral demand for a social contract: the establishment of an external power to enforce the moral rules—the power of the state. The social contract establishes what Hobbes calls the Leviathan—his Biblical name for the terrible power of the State that is necessary to enforce fundamental moral rules with the threat of punishment, and ultimately death for serious offenders. But then we see, in the light of our previous phenomenology, that true morality is ultimately abandoned for legalism. Instead of following the rules of morality for their own sake, because they are inherently what we ought to be doing, they are largely followed out of fear of punishment.

Thus far from being the expression of innate tendencies on the part of a naturally social species as Aristotle argues, the political state is a force against nature, an engine of violence which demands that individuals abandon their natural state of freedom for the sake of more limited, though also more beneficial, civil liberties. The social contract that establishes such a state consists primarily in rules that limit the natural liberty to use force and fraud against others to achieve one's goals. The more restricted civil liberties should repress natural liberty as little as possible, so that the individual remains free within the limits of the law to pursue his or her own conception of happiness. Here again there is a sense of pseudo-duty, for individuals must be willing to suppress their natural liberty for the sake of a limited conception of individual liberty that is compatible with the liberty of all. There is then an implicit conflict between state-empowered morality and the spontaneous tendencies of our natural desires. The individual's goal nevertheless remains the pursuit of his or her own happiness. Should she be able to avoid the legal penalties for violating the social contract—should she, like Kant's immoralist, be able to violate the interests of others without getting caught—her "sacred duty to herself" would require that she break the law. In Plato's *Republic*, the story is told of the Ring of Gyges, a ring of power which enables an individual to become invisible and by this means to get away with murder and other crimes. Why should an individual follow the moral laws if he can violate them with impunity? Hobbes' only reply to the story of the Ring of Gyges is that this is less of a problem with the existence of a State than without one.

The contemporary philosopher John Rawls adds to the libertarian requirements of the social contract a concern with equality. A truly enlightened egotist would want to live in a society that maximizes social and economic equality as well as enforces civil liberties. Only those inequalities will be permitted in Rawls' state that serve the interests of the poorest members of society. So if there is only one horse in the village, it should go to the doctor who can ride the horse to the home of the sick person across town. The motive for such egalitarianism remains however the separate individual's own happiness. For each person implicitly thinks—if I were such an impoverished sick person, I would want the doctor to have the only horse.

To arrive at this perspective, Rawls proposes a thought experiment. Imagine that you are unaware of your actual desires and interests, finding yourself in an "original position" "behind a veil of ignorance." Ask yourself then what rules you would select from such a position, which is analogous to the Hobbesian state of nature. You would want as much liberty as possible consistent with the liberty of others, and as little inequality as possible consistent with maximum well-being. If the Ring of Gyges is updated in Tolkien's epic of a ring of a Ring of Power which renders the wearer invisible, Rawls supposes a counter-device on behalf of morality: a

philosophical cloak that produces the illusion of forgetfulness and powerlessness, under which an individual deprives himself of whatever powers he may in fact have. Cloaked in such powerlessness and ignorance, he would then naturally desire laws that would defend him from the powerful. The rules in any case are based on the self-interest of the individual in his imagined state of disempowerment. Such would be the rules of morality to be adopted by the state.

But the first rule in this theory is the original *duty* to forget one's real position in life, and to imagine oneself in a state of powerlessness. Why should one adopt this unnatural position in the first place? In Hobbes' theory, where the state of nature is the starting point, the practical consequences of narrow-minded egotism motivate the establishment or acceptance of the social contract state. Ultimately, each individual realizes, after a long and bitter historical experience, that no matter how powerful she may be in fact, even if she is a monarch, she is ultimately powerless against the combined forces of other people potentially arrayed against her. Rawls' veil of ignorance, behind which each individual must imagine herself in a state of powerlessness, implicitly presupposes this lesson of Hobbes' state of nature. What else would motivate the recourse by inherently egotistical individuals to such philosophical inventions?

Hobbes did not need to invent such a philosophical thought experiment, for he was writing at a time of civil war, and his readers were all too painfully aware of the mayhem that results when civil order breaks down. But even in conditions of relative civil order there are plenty of occasions for wishing that the state was more powerful than it is—as travelers on lonely highways where cut-throat robbers lurk readily understand. As for Rawl's condition of maximum equality, Hobbes in anticipation of Rawls' argument replies that the existing structures of inequality do essentially work on behalf of the interests of the poor. For where the rich are threatened with the loss of property rights, civil war with all its depredations is the inevitable outcome. It is therefore better for the poor to remain poor while living in a state of social peace than for them to threaten the privileges of the rich and risk inevitable retaliation. The poor should therefore unite with the middle classes in the formation and support of a modern state that recognizes equal civil freedoms, but not raise the dangerous prospect of moving toward greater equality of wealth and social status.

Utilitarian Calculation of Happiness: Oneself versus Seven Billion Others

Utilitarianism is an alternative form of ego-based ethics developed by Jeremy Bentham and perfected by John Stuart Mill. Utilitarianism shifts the ground of morality from the happiness of the separate individual—the basis of social-contract ethics—to the happiness of the totality of individuals. The utilitarian argues that morality consists in the promotion of the greatest amount of happiness for the greatest number of individuals. That which is morally right is that action which produces the greatest happiness, not just for me, but for the maximum number of persons. Act utilitarianism focuses on the morality of the particular action that produces the greatest happiness, while rule utilitarianism focuses on the general rule that does so. Arguing that “It is better to be a human being dissatisfied than a pig satisfied,” Mill argues for the superiority of the satisfaction of higher, more refined desires to simpler or cruder desires.

Again morality is based on happiness, but since this is the greatest amount of total happiness, or the happiness of the greatest number of people, there is here a genuine sense of duty. The

individual ought to be willing to set aside her own happiness if this is outweighed by the happiness of others. The source of duty being the happiness of the majority of individuals, the particular individual, weighing his goals in the balance, may have to sacrifice his own happiness in order to promote the happiness of others. Here there is no recompense in the form of a long-range or more authentic happiness for the individual, as in the Aristotelian ethics. Nor is there the supposition that the demanding laws of the state which restrict his freedom are nevertheless there for the individual's greater welfare, as the social contract theorist argues. He may have to sacrifice the satisfaction of his desires for the sake of a greater good from which he receives no benefit whatsoever. His morally motivated actions are therefore genuinely dutiful. The utilitarian theory therefore seems to accord better than the other eudaimonistic theories with what we discover through moral phenomenology—i.e., the central importance of the experience of duty.

The phenomenology of morality stresses the potential *contradiction* between moral duty and sensuous or intellectual desires. Morality must be able to contradict the impulses of nature and individual desires, as well as the most refined calculations of self-interest that aim ultimately at the satisfaction of desire. This is what we conclude through reflection on given moral experience and before proposing answers to the question regarding the source or foundation (or “groundwork”) of this moral experience. This potential contradiction between the impulses of nature and the demands of moral duty is implicit in the previous theories. For classical natural law theory, short-term or lower desires for allegedly spurious happiness must be overridden so that the individual can realize supposedly long-term or higher desires for allegedly authentic happiness. Here immediate desires are repressed for something else, which is said to be deeper or higher desires. The experience here is duty-like. There is, implicitly, a duty to overcome the immediate press of desire. The natural law theory locates the source of this duty, however, in another desire or set of desires, the desires that express genuine human nature, which accord with the long-term happiness of the individual and harmony for society. Thus while it is recognized that morality involves the potential restriction of desire or happiness, the ultimate source of this semblance of duty is regarded as one's “true” desire for happiness, which is rooted in human nature.

Similarly, Rawls' contemporary social contract theory proposes that we be willing to give up the possibility of satisfying desires that we may have as a result of birth and circumstance and put ourselves in “the original position” (or imagine ourselves in an original state of nature) in order to discover general rules of social life that would conceivably be to one's own personal advantage. Here there is implicitly a duty to take up the impersonal original position in the first place, rather than simply to satisfy one's actual desires and maximize one's interests as dictated by one's possibly privileged position in society. Again, however, the opposition of duty and desire is said to be founded ultimately on the satisfaction of desires—the desires one would have in the imagined original position. Returning to actual life from this idealized state of mind, the individual recognizes certain duties—i.e., to recognize the equal freedom of others, or to sacrifice one's actual interests for the sake of rules that benefit the poor by redistributing income. But the ultimate ground of such duty is nevertheless the satisfaction of one's own desires and interests in the original position behind the veil of ignorance.

In these examples the adoption of a duty to go against one's desires is founded on desire itself, on the deeper or more lasting or more rationally derived (in the case of the veil of

ignorance) desires that require suppression of certain actual desires. The utilitarian moralist however more consistently acknowledges the contradiction between duty and desire. A utilitarian must be willing to sacrifice her own happiness for that of the others, for the desires of the greatest number. This is the *duty* of the moral person, since that greater happiness of the greatest number may conflict with her own personal happiness.

There is however an implicit inconsistency between this perspective and that proposed by our moral phenomenology. The inconsistency comes in with the utilitarian notion that one's duty must be derived from what is demanded by the satisfaction of the totality of desires. This is of course a painful duty when the desires of others outweigh one's own. It follows that I should be willing to give up my own piggish desires if in doing so I can contribute to satisfying the greater amount of piggish desires of others. If my group likes to insult the weakest member, then, I must contribute to the hilarity, and even submit to it if I am the butt of amusement. To possibly avoid this perspective, John Stuart Mill argues for a scale of happiness that rises to include the satisfaction of "higher" desires. It is better to be a Socrates whose material wants are unsatisfied, but has the satisfaction of his higher ideals, than a pig wallowing in the trough. But surely if a lofty desire weighs more on the scale of authentic happiness than a piggish one, a great number of piggish desires must ultimately outweigh a lofty one. It would certainly be the height of arrogance for Socrates to suppose, as weighed on a utilitarian scale of pleasures produced, that his solitary and unfulfilled pursuit of truth counts more than all the complacent opinions of the masses combined.

It would seem moreover that utilitarian ethics would ultimately deplete the duty-bound individual who adopts its standards. For each individual there are the desires of seven billion others to weigh in the balance. If most of these others are needy, while I am well-supplied, my duty requires that I should devote myself to the satisfaction of the needs of others. Altruism therefore becomes the predominant characteristic of utilitarian ethics. However, if the rule is that the greatest happiness of the greatest number provides the basis of duty, simple statistical logic determines that for most people most of the time their utilitarian duty will come down to satisfying *their own* desires. Thus the apparent altruism of utilitarianism might just as well turn out to be a rampant egotism of the average citizen.

Despite the fact that the above theories implicitly express a contradiction between desire and a rationally-based consciousness of duty, each fails to develop this understanding consistently and radically. In each case the realm of desire that morality must in one way or other subdue is reestablished as the very source of morality itself. After being suppressed in one form or another through a simulacrum of duty, happiness is nevertheless reintroduced as the ultimate foundation for moral norms. But to provide grounds for plausible moral rules, the desires that are to be satisfied must be reinterpreted, implicitly moralized desires, desires established in a certain framework as long-term, or as the most egalitarian or libertarian, or as those of the greatest number. Happiness is rationally reinterpreted as the long-term happiness of a social individual in natural law theory, abstracted as the happiness of the impersonal and ignorant ego in contemporary Rawlsian social contract theory, or recalculated as the many happinesses of the majority of people in utilitarian theory. Such "rationalistic" ethical theories are therefore rational only in a limited sense. In each case reason turns out eventually to be *subordinated* to the spontaneously emerging desires and ambitions of the individual—whether regarded as

essentially social or as naturally egotistical. Such “instrumental rationality” is typical of the calculating individual in modern economic life.

Reason or Feeling?

For *homo economicus*, reason is a tool of desire. But then how can such rationalistic moral theories, aiming in one form or another at producing happiness, pretend to be rational? Criticizing rationalistic ethics, Hume declares that reason is only the slave of the passions. It does not establish rules that govern our lives, but rationalizes the drives and passions that really move us. But morality, it must be admitted, really does move us. Morality must therefore be a matter of feeling or passion, not of reason. Hume is here essentially responding to the failure of previous rationalistic theories, especially that of Hobbes, to provide convincing arguments for their positions. For Hobbes admits that the rational morality that he proposes is essentially impotent to move people without the threat of punishment, with its motive of fear. Let us therefore put aside these ethical theories with their rationalistic orientations. Morality must be a particular kind feeling, Hume argues. It is a more disinterested sort of feeling, but a feeling nevertheless. Human beings may be egoists or rational calculators some of the time, but not always. We should recognize the variety of kinds of passions that move us, some of which incline us to help others without any selfish interest involved. We are benevolent individuals in some circumstances and self-interested ones in others.

In his theory of the passions, Hume essentially reverts to the position of Aristotle that the tendencies of nature are the basis of morality, and that these tendencies produce a natural inclination to live in political society. For Hume, not only are there selfish individualistic passions as well as benevolent moral ones, but there are overriding political passions uniting large numbers of peoples or whole nations in the project of establishing legal rules of justice. Such political passions are founded on the common interest of large numbers of people, and ultimately require subordination of the selfish and even the moral ones. If there is a law case involving inheritance in which the evidence favors a miserly rich person over a more needy and deserving poor one, moral sentiments of benevolence must take a back seat: the common law tradition demands that the rich person get the money. On this basis of the evolving traditions of law and society, Hume decisively rejects the social contract theories that would create a rational basis for the laws of the state. Thus he writes a history of England in place of a fanciful account of a state of nature being replaced at some point by a reason-based social contract. Society evolves its own standards over time in pursuit of common norms that bind members together for the sake of their common interests. There is no place here for a rational moral theory that criticizes the existing society. The common law tradition of accumulating case histories receives its philosophical justification in Hume’s empiricism, contrasting with the rationalist “civil law” theories that are defended by enlightenment philosophers on the European continent. The limited constitutional monarchy of England, with its inequalities of wealth and political power, and its slave trade, is thus justified, for the moment at least, by a natural process of historical evolution.

Kant’s phenomenological examination of the naturally philanthropic individual directly replies to Hume’s conception of morality. Natural feelings of benevolence, however amiable and laudable, have nothing directly to do with morality. In the first place, such feelings are fickle, changeable, unreliable. The individual who goes by her feelings may ignore you or turn on you

as easily as she may try to help you, depending on which way the winds of feeling are blowing at the moment. How can morality be built on such a foundation? Moreover, the forces of the feelings or passions are heteronomous powers that determine one's actions independently of any responsible choice. But morality presupposes freedom from deterministic forces of whatever kind.

Feeling or reason? If these are the only possible sources of morality, it seems that neither can satisfy the demands of moral phenomenology. We have just seen Kant's reply to Hume's theory of feeling as the source of morality. Let us look once more at the role of reason in the rationalistic theories founded on the pursuit of happiness. In all these theories, reason is subordinated to the realization of desire. The result of such subordination is that reason ultimately becomes powerless before desire, as Hume points out. Perhaps, Kant argues in the spirit of the Copernican Revolution in philosophy, the relation between reason and sensuously desiring individuality should be reversed. Suppose we take reason first, as something that can be operative prior to sensuous desires, as something to which desires themselves ought to be subordinated. This is the relationship suggested by a phenomenology of moral experience, in which desires and interests are found to be subordinated to duty. A *fully* rational ethics must look to reason in itself—practical and not theoretical reason, of course—to see whether it can reveal or supply the source of morality as well as provide justification for moral judgments. We should therefore criticize *impure* practical reason, reason subordinated to the promotion of desires, and attempt to turn our gaze on reason itself in its purity. This is the orientation of a “critique of (impure) practical reason.”

Animality and Humanity

To understand the intrinsic requirements of pure practical reason, let us consider the root difference between human beings and other animals. Non-human animals pursue highly *individualized* goals, as determined by a combination of internal biologically-determined instincts and external sensory stimuli. A hungry animal reacts directly to the presence of an *individual* piece of meat, and, if it is hungry and nothing checks its impulse, it straightaway sets about to eat. The animal's instincts determine for it the fact that it will notice and respond to a certain *kind* of food, say meat rather grass. It is moved by its inclinations, feelings, or desires in one way or another depending on its species orientation. But this species characteristic of the object of its desires is nothing *for* the animal itself. The general species of the food is not abstracted out of the individual instance and considered by itself. What the animal consciously desires is not a *kind* of food, but this *individual* piece of meat that is presented to its direct sensory perceptions as these evoke within it feelings of pleasure or pain, movements of desire or indifference, depending on its species-based instincts.

Human beings, by contrast, always implicitly pursue *general* goals. Eating, for instance, must first take the form of an idea before it can take the form of a practical activity. The individual holds in consciousness an *ideal representation* of a possible action. This representation evokes subjective feelings of pleasure or pain, but these feelings do not directly move her to act. The representation also has its own objective characteristics that are independent of these feelings. It is possible to direct one's attention to the representation itself, to change it, or substitute a different one. The feelings of pain and pleasure do not operate immediately or directly, but await

the decision of the agent whose impulses can remain in a state of suspension while the course of action is being deliberated.

The representation or idea of a possible action is expressed and objectified in language and language always involves general terms. The term “meat” explicitly designates a general type of food. The term designates the species abstracted from the individual instances to which it is applied. Thanks to language, the generalized object, the idea, attains an embodied, objectified form of existence before consciousness. The action that results in eating the piece of meat is the realization of an objectified ideal representation of the possible action. This idea is first held up in consciousness before it is realized in action. People create an ideal model in language or in other symbols of a possible activity and then let their actions be guided by that model. The same general model can be more or less repeated for many individual instances of actual eating and for different individuals. The general type of the activity is separated from its individual embodiment. This capacity to represent possible actions apart from existing individual circumstances is what we call reason or rationality. Thanks to the *abstractness* of reason, human beings can anticipate actions in the future, even the far distant future. Thanks to the *power* of reason human beings are liberated from immediate animal responsiveness to the desires and fears evoked by individual circumstances. Thanks to this *power of the generalized idea*, they can check the impulsiveness of immediate sensory feelings and desires when these seem to contradict idealized intentions. In this perspective, it hardly seems that reason is a slave of the passions. It is a power over those impulses of life that are all-powerful for the animal. But then on what basis does the individual finally act? If the role of reason is simply to determine “how great, how long-lasting, how easily obtained, and how often repeated” is a certain feeling of pleasure that is evoked by an idea of a certain kind of action, then we find ourselves with Kant’s amoralist to be slaves of our passions. Is then the power of reason to consist simply in a momentary subjugation of the desires for the sake of an ultimate capitulation?

Such a power brings with it its own distinctive feelings, for it inhibits natural inclinations, chokes them in their natural course, creates bottlenecks of action, frustrates the course of nature from following its inner direction. To the extent that the individual identifies with this natural flow, rationality impinges on desire as if it were an alien will. Such dissonance can be assuaged by enchaining reason to desire itself, by making reason the reliable slave of the passions. Depending on one’s orientation, such a conflict between reason and passion can give rise to misogyny, or feelings of hatred in the face of the requirements or pretensions of rationality. But it can also give rise to a *feeling of reverence*—the feeling that characterizes the moral person in the face of duty. For Duty, Kant says, “only holds forth a law which of itself finds entrance into the mind and yet gains reluctant reverence (though not always obedience)—a law before which all inclinations are mute even though they secretly work against it.”² Thus it is not necessary to choose between rationality and feeling, but we should recognize that pure practical reason produces within us its own distinctive feelings. These are distinctly moral feelings, not the feelings that arise out of nature and environment.

First and Second Signal Systems

For certain animal species, genetically determined desires are conditioned by individual circumstances. The circumstances can be manipulated consciously by humans. Dogs can learn to respond to patterns and commands deliberately established by humans. After the association of the sound of a bell and the presentation of food, Pavlov's dog eventually salivates at the sound of the bell by itself. The bell "signals" the coming of the meat. It might be doubted, however, whether the bell is a signal in the way language for humans represents some general type of action. For the dog, the bell is a concrete individual event that has been associated with the presence of meat and so, by a kind of contagion, it also comes to stimulate digestive processes. Once hunger is satisfied, ringing a bell fails to evoke any distinctive responses. However the idea of food expressed in the symbols of language remains a powerful force for the human being independently of whether she is presently hungry. Thus humans placidly plan the next day's eating agenda after a full meal. Consequently, linguistically-based signals of events constitute a "second signal system," one that is liberated from the reactions of individual circumstances and desires. This second signal system occupies the distinctive plane of human reason. Rather than externally arising in the environment as an object of perception, the second-order signaling is a product of the human agent herself.

The outline of a possible action is present to the human agent in advance through a verbal scheme, a second-order representation. Such linguistic representation is the means for creating an ideal schema to guide the individual act. Hunter-gatherers who portray the animals of the hunt on the deeply recessed walls of caves are engaged in a special kind of activity with spiritual-magical significance. The representation of game animals in cave drawings gives the hunters a feeling of power over the object of the hunt. But the same is true of representation in words as well, for the name has the magical power to summon up the presence of an absent reality, evoking a mysterious presence-in-absence of the represented object. The species characteristics of the object represented in language are experienced as its spiritual presence, its presence-in-absence, and so should not be invoked lightly. The power of such magic rules not only the objects of pursuit, but the pursuers as well. Thus between the human being and the object of desire is the power of an idea. While the animal is directly related to its food in perception, the relation between the human being and his or her food is indirect or mediated by imaginative, artistic, or verbal representation. The animal's sensory organs appear to put it in direct contact with the object of its (biologically oriented) desires. Another object in perception may be associated with the natural object and evoke similar feelings, as in Pavlov's experiments. So the lion returns to the water-hole where it previously found its prey. The very sight of the water hole signals for it the desired object. But for human agents, there must also be, beyond the direct or associated sensory stimuli in the environment, a self-produced ideal representation of the object. This representation itself evokes desires, feelings of pleasure or pain, in advance of the activity.

Before he acts, Kant's amoralist consults the feelings of pleasure and pain evoked ahead of time by the representation of a possible action. He acts solely on the basis of the kind and degree of feeling evoked by the ideal representation. But it is also possible to consider the representation in its own right, rather than merely as a means for stimulating feelings. It can be regarded as a thing in its own right, possibly as a source of beauty, awe, or reverence, feelings quite different from those related to ordinary sensuous experience.

Thanks to language, the hunters sitting around their campfire can revisit the events of the

day's hunt. They re-present to themselves what is now past, yet remains present in thought and words. So they are able to consider the contingencies of the preceding day and recognize patterns that can serve them for the future. Instead of vanishing into a past that no longer is, the past remains present through its representation in speech, so that the thinkers, the intercommunicating hunting party or the larger band, can consciously, deliberately compare and contrast different outcomes and so more effectively improve on their hunting technique over time. The predators and prey of the animal world, meanwhile, largely repeat the actions governed by directly perceived objects and associated signals.

As a result of language, early *homo sapiens* accumulates a wide variety of past experiences in the form of verbal representations, and then confronts these representations belonging to an ideal world or what Karl Popper calls a “third world”—i.e., not the external world of physical objects, nor the merely subjective desires and feelings of individuals—as objective entities. This feature of human thought enables the human beings to operate on their own thought as an objective phenomenon, and so to produce various forms of organization and reorganization—inventing new combinations of thought and choosing freely between different possibilities for reasons of efficiency, of logic, of poetry, or perhaps of legality, convention, and precedence, rather than solely because of the sensuous feelings they evoke. The capacity to rearrange or reorganize inner experience through its objectification in words and concepts gives rise to the possibility of new patterns of action operating *a priori* in producing new experiences. Animals, on the contrary, are largely stuck with their past. Because they are determined by *a posteriori* experiences arising in the field of perception, they are governed by the past in the form of instinct and previously conditioned responses. Unable to objectify their experiences in the form of linguistic representations, they cannot free themselves from determination by the past.³

An idea is not merely a picture of an individual thing. An idea generalizes; it represents a type or a class of things; it extracts or constructs the thing's general features or its underlying essence. Language, which expresses the idea, is inescapably general. Consequently, it is not *this* action only, but always a certain *kind* of action that the human being consciously considers. Even the word “this,” which serves to indicate the individual object, is a general term capable of indicating any and every individual object. The capacity for creating such ideal intermediaries between the desiring individual and the environment is what Kant calls practical reason. This idealizing source of action, or practical reason, sets human beings apart from all other existents in nature and gives rise to the possibility of human autonomy.

The natural world is governed by “heteronomous” laws (from the Greek, *heteros*, meaning other, and *nomos*, meaning law). Such laws reflect the fact that the movements of individual things, plants, and animals are governed by the movements of other things, plants, and animals, and so their individual movements and activities are responses to external conditions and internal impulses or drives. Through the welter and confusion of individual interactions, such dependencies on external conditions tend to constitute lawful systems of relations which operate in a regular or general way. The animal however is unaware of these regularities as such that permeate its existence. It does not consider them for their own sake, being content simply with the individualities immediately encountered through the discriminations inherent in its species make-up, plus the associated contingencies of reinforcement arising haphazardly or regularly out of the events of their lives—unless conscious human manipulation intervenes. In sum, the animal

“lives” its species life, but does not make that life an object for self-conscious, reflective consideration.

Human beings, however, are not immersed in their individual, sensuously perceivable circumstances. Thanks to the higher-level order of ideal representation, which we call reason, human behavior is only weakly governed by instincts. Instead of being unconsciously guided by instincts to act in ways that are appropriate to our species needs, we humans must consciously represent to ourselves the regular or general connections between things. Consequently, whereas an animal simply hungers for this individual prey, a human being can act to satisfy hunger in general. General *systems* of food-gathering and food-producing are consciously created. This is a result of the recognition that hunger is a regular feature of human life. It is the consequence of the general idea that one wants to be able to eat *whenever* one is hungry. It is not enough for humans that one’s present, individually felt hunger pangs are appeased. We must be assured of their general or regular appeasement, and, however full our bellies are for the moment, we suffer great anxiety when we are uncertain of this.

In satisfying our need for food, therefore, we are acting in accord with laws, as in the case of other animals. But there is this fundamental difference: we ideally represent such laws to ourselves, and in the light of this representation we govern our relation to the law. Consequently, instead of being simply *subject to the law*, as are other animals, *we give the law to ourselves*. The law of nature is independent of our wills, but the system of activity that we create in responding to this law exists only through the purposeful activity of human beings. *Giving the law to oneself, rather than being subject to laws arising from other things, is the essence of human autonomy* (from the Greek, *autos*, or self, and *nomos*, or law—i.e., self-determined law). Kant argues that genuine autonomy or freedom is the essential characteristic of morality.

Morality presupposes a negative relation to natural laws, in the sense that human beings cannot be morally responsible for their own actions if these are the direct results of deterministic laws. This negative relation to the laws of nature is implicit in the fact that the human agent represents a law of action to himself, treating the law as an object that is other than or *not* himself. But such a negative relation to the law is not sufficient to satisfy the demands of morality. By itself, free will produces only lawless anarchy. Morality, however, involves a sense of duty, a consciousness of what we believe we *must* or *ought to* do, not what we merely choose to do. It is not about the making of arbitrary choices. There must therefore be more to human freedom than the ability to escape from determinism. As well as having a *negative freedom from* something, we must have a *positive freedom for* something. In the awareness that we give the law to ourselves, we see this positive side of freedom. Such positive freedom presupposes negative freedom—i.e., freedom from animal-like determination by heteronomous laws. Thanks to the fact that we interpose an idea between ourselves and the object of our desires, we free ourselves from direct determination by those desires. But what are we free to do? In its positive aspect, the true freedom of human beings consists in the capacity to realize the laws of life as formulated by our own consciousness. Regarded abstractly, we may be following the same natural law of hunger that operates on the animals, but we do so indirectly, as determined by ourselves, in forms of our own devising.

We are sensuous creatures, too, and not simply rational ones. Like other animals, we are subject to biological and environmental conditioning and respond directly to the presence of

appetizing food. Human beings therefore have a dual nature—animal and rational. Consequently, we can allow our sensuous, animal side to overwhelm our reasoning side. We can let our actions be governed by the individual, immediate side of our circumstances, and then reason is silenced. Blinded by the pull of immediate needs, we can kill the goose that lays the golden eggs, or eat up the grain that should be planted for spring. Less drastically, reason, the reflective consideration of the action in advance, may not be completely silenced, but only subordinated to the seductions of sensuous individuality. The amoralist only considers “how great, how long-lasting, how easily obtained, and how often repeated” is the agreeable feeling evoked by the representation of the action. Then, it is sensuous, individual gratification that dominates the use of reason. Reason is then active in the service of its sensuous master. It is, as Hume says, a slave to passion. It is my individual desires in their individuality which set the goals to which I apply my reason. We can use reason as an instrument to improve our capacity to satisfy the demands of sense, of inclination, of feeling and sensuous desire. This is what I do when, rather than act on impulse or immediate desires, I calculate which actions will produce the most long-term gratification, which actions will best satisfy my desires *in general*, and not merely in this individual instance. I don’t eat the goose, because its golden eggs will supply me with limitless food. Seed is set aside for the spring sowing, but even more is set aside to lend at high rates of interest to my improvident neighbors so that eventually I may be able to eat grain without sowing at all. I cleverly recognize that my neighbor is a gold-laying goose in human form.

But in these examples, since these sensuous desires are the result of underlying, unconscious natural causes—such as the biologically-based need to eat—the implicit autonomy of human reason remains subordinated to the heteronomy of nature. Rational calculation of what would best satisfy natural desires remains enmeshed in the domain of nature and unfreedom. Hobbes defines freedom as the ability to realize one’s desires without outside interference. But any animal that is able to catch its prey, or flee from its predator, is “free” in this sense. Such freedom is compatible with thorough-going determinism, since one’s desires are the result of external causes in nature, circumstance, and education (the contingencies of past experience, deliberate training, or education). Paradoxically, however, this human unfreedom is not the same as the unfreedom of the animal, since implicitly or indirectly it is freely willed. Our willing subjection to the mechanism of desire makes of it a pseudo-mechanism, not a real determination by outside causes. But to understand this point it is necessary to recognize that there is an alternative course of action in which reason is not the slave of passion, but its master.

Hypothetical Imperatives

Human beings are “rational animals.” The term itself suggests the primacy of the animal side of the human duality, since “animal” is in the grammatical position of noun, while “rational,” in the adjectival form, is subordinate. As *rational* animals, humans do not rush headlong in the direction of satisfying their sensuous desires, but evaluate the various means of realizing some goal. We are hungry, say, and someone else possesses food. As sensuous individuals, we may naturally be inclined simply to take the food possessed by the other person. However as rational animals we check our immediate impulse. Perhaps, we think, the other person has weapons to defend his food. Perhaps there’s a policeman nearby. It might be more *prudent* for me to find my own food, or grow it, or, at a higher level of human complexity, even lend it to others on interest

and ultimately take possession of their food in that legally-sanctioned way of the higher civilizations.

Acting rationally, we not only develop *ideas* of action—i.e., symbolic mental models of possible actions that can be reused in other circumstances. At a higher level of generality we develop *ideal modes* of acting—i.e., general strategies of action. My general strategy for achieving my goals is expressed in a “maxim.” The uncertainties of the future may give rise to a focus on the present moment, and so I adopt the ideal strategy expressed in the maxim of the Roman poet Horace: *Carpe Diem*, Seize the day! On the other hand, I may devote myself more to the accumulation of means of enjoyment in the future, following the maxim of the American inventor and philosopher Benjamin Franklin: “Time is money.” The maxim of the action here is to pursue my long-term advantage—including, perhaps, that of my family as well—rather than focusing on the immediate pleasures of the day. In either of these opposing strategies or ideals of action, it is still my sensuous, individual gratification which dominates the use of my reason. Reason here is “instrumental reason,” reason subordinated to goals that stem from natural desires and inclinations.

“Seize the day” or “Time is money” are commands or “imperatives” of (instrumental) reason. Reason tells us through such imperatives how we *ought to* act under certain circumstances. Such commands or imperatives of action are nevertheless subordinate to an overriding command that does not come from reason. This is the command of our natures to satisfy our desires as best as we can. For all its power to suspend the promptings of desire, reason merely prolongs and complicates the means and methods for yielding to its sway. Reason is in this way essentially a slave to passion, as Hume says—an instrument in the pursuit of “happiness.” Kant calls such imperatives of reason that depend on our ways of pursuing happiness, *hypothetical* imperatives. Reason here does not command absolutely, but only relatively. On his birthday, for example, the exhausted business man yields to the fatigues of nature and decides to forget for the time being the imperative of future delights, “Time is money!” For a change, he gives himself the command of the pleasurable present, “Seize the day!” Hypothetical imperatives which aim at achieving happiness are not rocket science. They are only approximations or guesses in that most problematic of projects which is the pursuit of happiness. Hence such imperatives are “hypothetical,” not absolute or “categorical.” They constitute hypotheses for testing rather than verified laws of action.

We inevitably desire happiness, and so this law of nature may seem to command us categorically. But what exactly do we understand by happiness? This goal of nature is different for different individuals. How do we achieve our own idea of happiness—by saving for the future or by living for the moment? Because of both the vagueness of the goal and the uncertainty of the means for attaining it the commands or maxims of happiness are not absolute or categorical. The end of happiness is necessary but the means for obtaining it are optional. It is therefore the source of many varied and conflicting hypothetical imperatives. No definite line of action is set out, whether for human beings in general or for any human individual in particular. What Kant calls the “pragmatic maxims” of happiness command imperfectly. To be happy we should probably exercise regularly, in some way or other. But how, or how often, is debatable and variable, and for some people there seems to be no necessity whatsoever. We know we want to be happy, but we don’t know what this means concretely, or how precisely to achieve this

goal. Happiness, Kant says, is a regulative Idea. We move toward it as a practical horizon by trying to maximize the realization of our desires. The ideal of happiness in the practical sphere is like the ideal of complete scientific knowledge for theoretical reason. It represents a totality of satisfactions of desires, which is not anything definite and recedes as we approach it through the attainment of particular satisfactions.

There are however many imperatives that seem in isolation to command absolutely. For the computer to work, it *absolutely must* be plugged in. On deeper consideration we see that such strict imperatives are so only on the condition that we have already adopted certain goals which are themselves based on desires, and ultimately on the desire for happiness. Thus the weakness already pointed out regarding the maxims of happiness infects even these strict, “technical” imperatives. Technical imperatives command us absolutely to act in certain ways, but only *if* we choose to achieve certain goals. This “if” makes technical imperatives similarly hypothetical. If you want a better-paid, more interesting white collar job, then you *must* have a higher education. And if you want to have a higher education, you *must* have a certain amount of money, have a certain level of grades, be able to read and write, etc. If there is no public transportation and you live far away, you must have a car to get to work. Nothing however demands absolutely that we pursue the larger goals—e.g., working for such-and-such a company whose offices are across town. We are free to decide on certain goals, motivated ultimately by the general desire for happiness. We hope that our secondary, optional, goals will bring us nearer to this primary, necessary, one. Once these secondary goals are decided upon, the means to achieve them may be more or less necessary, and reason then issues its technical imperatives, or imperatives of skill.

The technical imperatives arise out of the laws and regularities discovered by experience and science, or theoretical reason. But it is not theoretical reason itself that motivates us to act. The fact that computers must be plugged in rests on scientific laws and is confirmed by ordinary practical experience. But its necessity is not what moves us to act. It is the choice of a goal, motivated by desires and interests, that moves us to consider the best way to achieve this goal. If practical reason is subordinated to our desires, theoretical reason comes into play only because it is here subordinated to practical reason. This is quite different from the primary motives of *theoretical* reason *per se*—the challenge of conflicting opinions, the tantalizing pull of a problem to be solved, the desire for truth. The scientist also has *practical* concerns, such as the need for an income or the prestige of a high position in society. But to make headway in the science itself she needs to enter into its inner theoretical requirements and respond to their logic.

In the various chains of ends and means that constitutes practical life, Kant therefore distinguishes two types of hypothetical imperatives: those pertaining directly to the goal of happiness, which he calls pragmatic imperatives, and those that command as means for achieving such goals, which he calls technical imperatives or imperatives of skill. Technical imperatives seem to have a kind of strict necessity, and so to command absolutely, but as they depend on optional goals, their necessity, like that of the pragmatic imperatives, is only hypothetical. Thus in each type of hypothetical imperative, technical and pragmatic, there is a mixture of necessity and arbitrariness. Since happiness is the over-all goal, both pragmatic and technical imperatives constitute reason-based means toward an end that is not itself rational. In this entire sphere of practical action, encompassing the myriad pursuits of modern civilization with all its vaunted science and technology, reason itself is never what really moves us. It always

remains a slave of the passions.

The Means/Ends Relationship

It might seem that in the case of technical imperatives that while the ultimate goal is governed by desires, the actions aimed at performing the means, the intermediate goals, are motivated by reason alone. And so it might seem that it is in the realization of the means that we can see the operation of some kind of pure practical reason. The operation of practical reason in realizing the means would therefore not be subordinated to desire, but must dominate the feelings and desires that might interfere with my realization of the means. So for example, if I am hungry and want to eat I must have a certain amount of money to buy my meal. But on the way to the restaurant I see a nice tie in a shop window, and feel like buying it. However, as a rational being who has willed a certain end, I resist this interfering desire and persist in the course of my action, holding tight to my wallet. Otherwise I would be acting irrationally. Out of respect for my status as a rational human being, I constrain myself to resist the desire to buy the nice tie so that I may continue to pursue my original decision to have lunch. Thus even in ordinary actions involving the realization of our desires we seem to see the operation of pure practical reason, dominating our desires rather than being subordinate to them.⁴

The above understanding seems to be suggested in the following passage, where Kant writes:⁵

Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power. This proposition is, as regards the volition, analytic; for in the volition of an object as my effect, my causality as acting cause, that is, the use of means, is already thought, and the imperative extracts the concept of actions necessary to this end merely from the concept of a volition of this end (synthetic propositions no doubt belong to determining the means themselves to a purpose intended, but they do not have to do with the ground for actualizing the act of will but for actualizing the object.)

This passage might be interpreted⁶ as arguing that as the relation between end and means is analytic, it follows that this implies the operation of an *a priori* rational norm. So if my end is to treat a certain disease, Z, I must (ought to, on pain of failure of rationality) will the means of applying a certain medicine, Y. The pure rationality or logic of the means-end relationship therefore constrains my will in the event of contrary desires arising to distract me from my original goal. Thus while I desire to help the patient, my application of the means to do so is not itself the result of this desire. It is a result of pure practical reason constraining me to act on the basis of the *a priori* analytical proposition, if I want the end I must will the means. Failure to apply the medicine in the event of a contrary desire (e.g., to watch a TV program that comes on at this time), is therefore a failure of rationality itself. And so rationality operates as a distinct motivating factor in the application of the means to ends, though not to pursuing the ends themselves.

Thus it might be argued against the Humean empiricist conception of action that even in ordinary cases of actions in which the ends are motivated by the irrational forces of desires, feelings, or interests, pure practical reason must be called upon to move us to apply the means

required to realize one's goals. And so anyone who doubts that morality is motivated by pure reason—those who think that we are *only* moved by our desires—fail to understand how people are motivated by pure rationality in implementing the means to an end in cases where we are moved by ordinary desires and interests in pursuing the end itself. Clearly, it seems, we do not desire the means themselves. I do not enjoy paying money; I enjoy eating. But as I recognize the necessity of paying money in order to eat, pure practical reason motivates us to pursue the means to the end. So it is in relation to the means that I am moved by a purely rational consideration: if I have willed a certain end I must will the means to get to it, however little I may like to do so. Thus even in clearly non-moral cases, it seems necessary for pure practical reason to create its own distinctive form of motivation independent of antecedent desires.

However in the above passage Kant writes that the necessity of willing the means is “as regards the volition, analytic; for in the volition of an object as my effect, my causality as acting cause, that is, the use of means, is already thought.” What is *analytic* here is not the relation of means to end in itself, but the *willing* of the means to the end. The relation of the means to the end in itself is a different matter. Whether a certain means will produce the desired end is a *synthetic* proposition, based on scientific experience. I.e., we learn from scientifically-informed experience that a certain medicine will help in certain cases. Theoretical reason here enters the picture in determining the relation of means to end as an objective process. But this does not imply that in applying the means the individual is, contrary to Hume, motivated by pure or theoretical reason. What Kant says is analytic is not the relation of means to end regarded as an objective process for theoretical reason, but the *volition* of applying the means to achieve the end. In other words, it is the *willing* of the end that contains already the *motivation* for the actions involved in implementing the means. And so if willing the end is based on a desire, as is the case in hypothetical imperatives, the same desire motivates the actions involved in applying the means to this end. The desire governs *both* means and end as far as the motive of the action is concerned. I do not insert an additional purely rational motive regarding the means between my desire and its realization. The rational choice of the best means for the desired end is informed by synthetic, scientific or empirical reason, but this is not the motive for the practical action of applying the means. Here is what is analytic: if I really desire a certain goal, then I also at the same time desire to apply the means as well. There is no room here for a purely rational consciousness to intervene with its own independent, purely rational motivation. There is no hint here of the operation of pure practical reason. Thus Kant completely accepts the standard Hobbsean-Humean concept of action, with its portrayal of the dominant motivating rule of the passions, in relation to hypothetical imperatives.

Let us consider the example of making tea. I want to make tea, as it is my usual tea time, and so I recognize that I *must* boil some water. The effective motive for boiling water is not the pure logic of the means-end relationship, but the desire for tea, spilling over into the application of the means to achieve it. I do not first desire tea, and then constrain my wayward desires for other things while I rationally motivate myself to undertake the instrumental action of boiling the water. If the boring task of boiling the water is easily distractible, I do not motivate myself by the thought that I will fail to be a rational human being if, having decided to pursue a certain goal, I do not also carry out the means to that end. Rather, I motivate myself while boiling the water by the thought of how nice the tea will taste when I am finished. Suppose, however, that just as I am about to turn on the kettle, I remember that there is an important TV program that is due to start

just now. The TV is in the next room, and if I boil the water now and leave it to watch TV, there is a danger that I will forget about it and start a meltdown in the kitchen. I therefore do not boil the water, not because of a failure of rationality, but because of a stronger desire overriding my desire for tea at the moment. I decide that I would rather watch TV for now and so put off making tea until the end of the program. There is nothing irrational about changing my mind when it comes to the realization of my desires, for, as we have argued, the desire for happiness does not impose on us strict norms of action. It is a hypothetical, not a categorical imperative. It allows us to change our minds about what would really make us happy.

Kant's point is completely in line with the standard theory of action, espoused by Hobbes and Hume, that instrumental reason is motivated by desire or passion. He agrees here with Hume that no mere rational understanding of an analytical truth has the power to constrain us in the face of a contrary desire. If an individual has so little desire to help a person in need that the thought of a TV program would deter him from applying the necessary means, the rational thought that the end requires the appropriate means would hardly move him to continue to carry out his original plan. All the rationality of means ends relationships as far as the objective causal process is concerned is the result of synthetic empirical or scientific reasoning. But such a theoretical understanding of the relation of means to end is not what moves us to act. The relation of boiling water to making tea would remain a purely contemplative observation on our part were it not for the motivating power of a desire for drinking tea. As far as volition is concerned the connection is analytic: he who wills the end by this very willing wills also the means. The entire process is therefore subject to the ruling force of the sensuous desire for a particular end. No purely rational desire to produce the means because one is a rational person interferes with or supplements the realization of this desire, which remains dominant throughout. The will to perform the means is already contained in the decision to realize the desire for the end. The rationality of means end relationships is therefore entirely subordinate, as far as the practical motivation of the action is concerned, to the ruling desire for the end, and has no independent power to move me.

Pure rationality enters the picture *only* as a result of morality. Suppose a doctor wants to help his patient for whatever pragmatic reasons—to make money, to uphold a reputation, etc. Just as he is about to apply the required treatment, the thought of a fascinating TV program that begins at this moment occurs to him. And suppose that his desire to watch this program is greater than his desire to make money by performing the operation just now. He could easily postpone the operation for an hour, making up some excuse. *An urgent matter*, he could say, *requires his attention. Put the patient on hold*. No one would even think of challenging him. Is he going to thwart this greater desire to watch the TV program on the basis of such an anemic motive as the rationality of the means ends relationship—i.e., the idea that, having decided on an end it would be irrational of him not to perform the means? If decisions to realize desires are all that matters, there is nothing irrational about changing his mind about what he wants. He does not carry on with the treatment of his patient on pain of violating the rationality of the means-end relation, but because another motive intrudes: his *duty* as a doctor to his patient, superseding all purely personal desires. But this takes us beyond hypothetical imperatives to categorical ones.

The difference between the two kinds of imperatives was touched on in the previous chapter. While we have a universal and necessary desire for happiness, this is a “wavering Idea,” not a rationally binding law, contrary to a certain scholastic interpretation of Aristotle's eudaimonism

according to which the desire for happiness produces universal laws of behavior. Such may have seemed to be the case in a relatively static medieval society in which roles are determined mostly by birth, and in which the different classes of people are integrated into an organic whole and hierarchy. Thus the happiness of the serfs consists in working the land, while the happiness of their lords requires feasting periodically on the spoils of domination, but not before the obligatory services of the priests who must bless their meals. In such a world, a system of laws appears to arise that is based on the pursuit of happiness. However, for the emerging modern world as expressed in Hobbesian individualism, the pursuit of happiness is different for different individuals, and so there can be no universal and necessary social laws based on this idea. This does not prevent the desire for happiness from *affecting* everyone, and doing so necessarily. So we can say that everyone necessarily desires to be happy, although what this means differs from one person to the other, and is capable of changing from moment to moment for the individual herself.⁷

Consequently there is room in the pursuit of happiness for the exception to the rule, as in the case of the man suffering from gout. Yes we are moved to pursue happiness, but we don't always do what our idea of happiness requires of us. In this case the desire for happiness is contradicted by another, more potent desire for rich foods. This more focused desire for a small happiness in the short term at the cost of a larger happiness in the long term results in a refusal to do what the larger idea of happiness requires: give up certain cherished rich foods for the sake of a longer and healthier life. We should note that the failure of the idea of happiness to motivate in such cases is indeed a failure of the power of rationality itself—i.e., a failure of instrumental rationality. Suppose that the individual understands clearly that a sacrifice of certain eating habits is rationally (scientifically) required for achieving a longer and so, quantitatively speaking, a happier life. He ought therefore, *simply as a rational individual*, to behave accordingly. But this idea of the irrationality of his eating habits fails to move him in the presence of a more powerful desire. Here Hume's dictum appears verified: reason is simply the slave of passion. "Cold reason" has little power to move us in the heat of the passions evoked by a warm and savory kitchen. The philosopher who chides Kant's friend with gout on his failure to live up to his lofty status as a rational animal may evoke in him a momentary sense of embarrassment. But life is short and his dinner is sweet. If you want philosophy, our gourmand replies, take this: *Carpe diem!*

Thus Locke devotes considerable attention to the paradox that people generally are not moved to do that which they themselves believe will produce their own greater good. Facing down the fires of hell for all eternity, the truth of which he is firmly convinced, the drunkard orders up another drink. Something in addition to the irrationality of his eating habits, or the irrationality of drinking himself to death, is needed to supplement the weak influence of the person's ideal of happiness. Theoretical reason here, which recognizes that the vaguely desired end of happiness requires the means of a restricted diet, fails to move the individual who is motivated by his habitual pleasures. He must therefore turn to some other source of motivation than the vague force of the ideal of happiness, or the requirements of reason based on that pursuit. If he wills the end he must will the means, Kant says. But when the end is a vague ideal, it is willed vaguely. If the theoretical reason of scientific medicine prescribes a restricted diet, this theoretical necessity only has the motivating power of the end itself. Recognition of its rationality by itself is a matter of cold reason, and moves us as little as does the recognition,

considered by itself, that having tea requires boiling the water. To act on this necessity of the means it is necessary to desire the tea. In face of the competing and more compelling desire to watch a TV show, the rationality of this means-end relationship avails not at all. And so it is with the man with gout. He wants to be happy. I.e., he doesn't want to suffer from gout and all the encumbrances it brings with it. But he also wants his juicy chops, and the pain in his foot is not so bad right now. Kant here addresses an issue that Locke puts in the center of his theory of action: why is it that people who know what is really good for them so seldom act on the basis of that knowledge?

Happily, the gout-sufferer can find within himself another source of action that requires that he shift his attention from his personal pleasures to higher grounds. This higher source cannot be reduced to the idea of the irrationality of his behavior as a means for obtaining the end of happiness. If the pursuit of happiness, rationally understood as requiring a different diet, fails to move him, he can nevertheless find in himself a source of action whose imperatives bind categorically and not merely hypothetically, and contain a power that is capable of superseding all the forces of desire. This is the force of moral duty, with its categorically commanding imperative.

Reversing the Role of Reason

There is another kind of rational evaluation whose commands are categorical, not hypothetical. The commands of moral duty are absolute or categorical, reversing the dependence on sensuous desires that characterizes hypothetical imperatives. The Copernican Revolution in the theoretical sphere should therefore be extended to that of practical reason. In relation to the theoretical sphere, there is a reversal of the traditional relationship in which concepts are supposed to be derived from objects. Instead, we saw, objects derive their very objectivity—within our experience—from the *a priori* concepts or categories of the understanding. In the practical sphere, such an *a priori* form of giving order to experience is missing in that sphere of action ultimately governed by sensuous desires or the vague and weakly motivating ideal of happiness. Reason here can find no objective order, for its would-be imperatives dissolve in uncertain estimations and variable calculations. No objective duties can be forthcoming from such a perspective, as everything depends on the shifting and uncertain sands of the impulse for happiness.

The utilitarian, who believes that the satisfaction of desires is the basis of morality, directly acknowledges this variability in denying of that any action is intrinsically moral or immoral, regarded independently of its consequences. The social contract approach seems at first to command an absolute law of liberty, and perhaps also of equality. I should respect the freedom of the other person as equal to may own. But, argues Hobbes, I should do so only as a means for satisfying my own egotistical desire for happiness. However, since I cannot trust the other person to respect *my* freedom, I cannot respect hers. Hence there must be an external, heteronomous law of the state, or of religion, that can frighten both me and her into doing so. Because it only ultimately serves to promote the happiness of the individual, an alleged law of respecting the freedom of others cannot therefore command categorically in the conscience of the individual himself.

The natural law approach of Aristotle and the medieval philosophers attempts to derive lawfulness from the observation of nature, directly abstracting essence from appearance. Everyone naturally and necessarily desires to be happy, and so it would seem that such universality and necessity of human nature should provide binding laws of action. In the pre-Copernican framework of ancient and medieval ethics, ethical laws are regarded as abstractions from observable social and natural relations. The world as it is, presented to us in ordinary experience, reveals the way it ought to be. In his play *Galileo*, Berthold Brecht satirically epitomizes the “laws” issuing from such a perspective:⁸

When the Almighty made the universe
He made the earth and then he made the sun.
Then round the earth he bade the sun to turn—
That’s in the Bible, Genesis, Chapter One.
And from that time all beings here below
Were in obedient circles meant to go:
Around the popes the cardinals
Around the cardinals the bishops
Around the bishops the secretaries
Around the secretaries the aldermen
Around the aldermen the craftsmen
Around the craftsmen the servants
Around the servants the dogs, the chickens and the beggars. . . .

It was in the spirit of the Copernican revolution in science that early modern philosophers attempted to derive ethical laws from some other, more primordial ground than by contemplating the existing order of the society. Thus Hobbes argues that social order is an historical construction arising out of its fundamental elements—human individuals who are essentially free and practically equal to one another. All the hierarchical subordinations of human beings that we observe in experience must be explained as arising out of these fundamental elements, just as the movements of the planets must ultimately be explained by more fundamental, elementary movements. However, as we have seen, all modern attempts at deriving binding laws or imperatives of behavior from the admittedly natural desire of the individual for happiness ultimately fail. It would follow from this failure that the search for the fundamental basis of social order has therefore not been sufficiently deep and radical. The basic unit of society may not after all be the separate individual, naturally striving like all the animals for the satisfaction of natural desires. Such an approach can never give us laws for a distinctly human society.

Perhaps therefore we should experiment with reversing the relationship between reason and passion. Perhaps reason has its own goals, to which sensuous individuality ought to conform. In the spirit of the Copernican Revolution, let us see if we can reverse the traditional relation in which human reason has been put to the task of serving the natural inclinations of sensuous

individuality. Let us put what is distinctly human in us, namely our reason, ahead of our animality. Instead of considering ourselves to be “rational animals,” let us look on ourselves instead as sensuous intellects, desiring intelligences, or rational spirits using bodily instruments. Morality is precisely the demand for such a fundamental reversal in the structure of human motivation.

Bringing about such a reversal implies that reason should propose goals from out of its own inner substance, so to speak, instead of depending always on those goals that spring up out of our myriad individual desires. But how does reason command *a priori*? Doesn't the motive of action always spring from some kind of desire? Should we then try to set aside our sensuous desires and see whether there are some purely rational desires or goals? But even the most refined intellectual activities seem to be pursued for the sake of satisfying the desires that they inspire. Moreover, we have not traveled so far from our animal history as to be able to ignore the claims that animal-like sensuality place on us. We should never forget that, whatever else we are, we are also animals, creatures with desires and biological needs. We shouldn't pretend that we belong to some higher realm of pure, angelic intelligences. Therefore, for all these reasons, we *begin* with the fact that we have desires.

But we should also recognize that these are always implicitly *human* desires, the desires of a desiring intelligence. Our desires are inevitably formulated in language and take the form of an idea to be realized. Inevitably, the individual desire is transformed into an idea that is no longer focused on the purely individual level of the desire that stands behind it. Conscious awareness or pure practical reason is therefore an essential moment in the process of human action. Let us acknowledge that there are natural laws that operate in our lives through our impulses and longings, our desires and fears. But instead of allowing ourselves to be submerged by those laws or ultimately governed by them, let us face them consciously and act on them in the full light of conscious awareness.

Suppose we are hungry and the food we desire belongs to someone else. Unlike the animal that naturally reacts to its sensuous impulses, we must represent to ourselves the course of our action first of all in the form of an idea. We say to ourselves, in so many words, “It would be nice for me to eat food that belongs to her.” Such an action, in which the property of another is taken without her consent, is called stealing. Thus what for a stray cat would be a very simple, instinctive action, is for a human being a complex one with multiple levels of meaning involving purely individual aspects as well as complex social dimensions. On a further level of generality, this assertion of one's individual desire in the form of a general idea evokes or implies various possible maxims of action, usually recognized as moral ones, such as “It is never right to steal.” Such maxims do not assert the best possible means of realizing some desire or of becoming happy in general, but proclaim standards of action to which our desires should be subject. Such moral maxims are normally supposed, in the common sense understanding, to command absolutely. They announce categorical, and not merely hypothetical, imperatives. The question then must be addressed: are such categorical imperatives justifiable? And, what is the ground of their possibility?

The Birth of Morality

Two questions must be answered in connection with the existence of moral maxims or rules. The first question has to do with the legitimacy of the claim itself. How do we know that our claim that something is right or wrong is a valid one? In the second place we need to ask ourselves about the power of such claims to motivate us. What is the basis of the power of moral ideas to motivate individuals in the face of competing inclinations, desires, and interests? Both questions are about the source of morality. So we must ask ourselves, where does morality come from? Kant's notion of a Copernican revolution in philosophy provides a key to understanding this question. Historically, we begin with one orientation to morality, and then move on to another, more adequate one. We begin with a pre-Copernican conception in which morality appears to spring from our desires.

Genealogically considered, morality and desire first exist in an undifferentiated unity. There is a natural psychological transition in our thinking from our having a certain desire for something to our belief that it is right to do that which satisfies the desire. "Good" or "evil" seem to be different terms for "pleasurable" or "painful," desirable or undesirable. We want to have something because we believe that it is good for us to have it. Spontaneous egocentrism suggests that what is pleasurable or *good for me* to do is also good or *right in itself*. I have an inclination to regard the world of my desires as reality, and to think that what seems good for me is really good in itself. A child egocentrically elides the language of desire and the language of morality.

But experience challenges this primitive egocentrism. The child learns that there are other selves in the world, with claims of their own that may contradict hers. She has to find a place in her worldview for other selves. Jean-Paul Sartre cites the poet Baudelaire's recollection of the nature of his childish universe. Mentally addressing his mother, Baudelaire says, "I was always living in you; you belonged to me alone. You were at once an idol and a friend." Sartre comments:⁹

It would be impossible to improve upon his description of the sacred nature of their union. The mother was an idol, the child *consecrated* by her affection for him. Far from feeling that his existence was vague, aimless, superfluous, he thought of himself as *son by divine right*. He was always living in her which meant that he had found a sanctuary. He himself was nothing and did not want to be anything but an emanation of the divinity, a little thought which was always present in her mind. It was precisely because he was completely absorbed in a being who appeared to be a necessary being, to exist as of right, that he was shielded from any feeling of disquiet, that he melted into the absolute and was *justified*.

Originally, in the egocentric world of the child, there is no distinction between what the child *wants* to do and what she believes to be *right* to do, what makes her "justified." There is a spontaneous inclination to identify what I want with what is right, or, perhaps, to identify what is supposedly right—e.g., what my mother demands—with what I want. The distinction between desire and right comes with later development, involving a break from early egocentrism and the recognition of a world with many different centers. Then my desires become recognized as *my* individual desires, the desires *I* have, as distinguished from the desires *others* might have. The world no longer appears to center on what I want, or, conversely, what I want no longer appears to melt into the absolute of what ought to be. The realm of individuality becomes psychologically distinguished from the realm of the universal; what is good for me is separated out from what is good in itself, what is good for others, or for people in general.

With the awakening of individuality through the contrast between subjectivity and objectivity, between the individual and the universal, the individual with his diffuse wishes and desires no longer feels automatically justified. The individual must justify his actions in some way or another. Biblically expressed, such a separation of individuality from all the rest of existence appears to involve a fall from the original state of grace that parallels the natural harmony of childhood. All was good in the beginning, says the *Book of Genesis*, but our first parents decided to contradict the inherent good by choosing what merely *seemed* good to themselves but was not *really* good.

Scientifically expressed, childhood egocentrism is prolonged in the geocentric perception of the universe. Just as the post-Copernican heliocentric understanding requires a revolution in relation to our spontaneous perception of reality, so it becomes necessary ethically for the individual to recognize that what merely seems good is not necessarily good in itself. The ancient ethics of Aristotle, operating within the geocentric conception of the world and its metaphysics, continues in a more sophisticated way to base moral right on natural inclination. What is good for the slave-owner is by nature what is good in itself and so slaves, because of their semi-humanity, must naturally desire to have their lives dictated by others. But post-Copernican ethics, reflecting the individualism of modern times, draws a sharp line between what one individual wants and what the other wants. What is good in itself can no longer be thought to arise essentially out of one's own natural feelings or inclinations. Social contract theory sees morality as providing a limiting framework within which individuals can safely fight over the pursuit of their different goals, their different conceptions of happiness. Utilitarianism adds up the individual preferences, identifying the morally good with the desires of the democratic majority.

More radically considered, the psychological evolution from childhood toward adult maturity parallels or recapitulates the evolution of nature in which rationality becomes distinguished from animality. The animal does what is naturally right for its species by instinct, just as the child follows its childish impulses without guilt and in the indulgent eyes of its parents. Morality begins when I perceive myself as distinct from my parents, distinct from others, and so having to take responsibility for what I as a separate individual do. What I want to do is then perceived as *not* what my parents want for me, as contrary to what others would like me to do. The individual becomes aware of his own individuality in the context of a psychological break, whether implicit or dramatically expressed, from parents and from others. Then what I want to do is no longer what is spontaneously right, what is simply done, what belongs to the eternal order of things. I am on my own and must decide for myself.

The Abyss of Freedom

Kant holds that the transcendental origin of moral demands, their basis in the nature of practical reason, consists in one's rational awareness that the general nature of any action is in potential conflict with one's individual desires. When I desire some particular food belonging to another person, I necessarily state the goal of attaining that food in general terms: *I would like to take food belonging to another person so that I can have it for myself*. But if I pay attention to my goal expressed in this way, I recognize that I am proposing, in general terms, to acquire property by destroying property. Moreover, I recognize that this is not only a self-contradictory law of

action, but one that I would not want applied to myself. The maxim, “It is wrong to steal,” is therefore not merely an introjection into individual consciousness of external social rules. We do acquire moral instruction solely from the example and teachings of others, but we also have an internal source for validating such notions. Moral maxims are distillations of the generality implicit in the ideas by which, as rational, purposefully acting beings, we inevitably guide our actions. Implicit in “It would be nice for me to eat food that belongs to her” is a general maxim, “It is good, right to steal.”

There is first of all an awareness of a desire for a possession belonging to another person. Then, on the basis of the desire, there is the formulation of a possible project of action—e.g., taking that possession from the other person. The desire itself belongs to the plane of sensuous individuality. However this plane is inevitably transcended in the ideal formulation of a project of action. At this ideal level an action is inevitably represented in general terms. On this ideal or idealized plane, it is no longer a question of a singular sensuous “me,” the individual who wants “this” object. There is at the same time, interrupting such sensuous solipsism, a person-in-general who desires an object belonging to another person-in-general. The concrete individual with a desire must confront and identify with an abstract entity, the human being that I am, with a definite kind of plan. Morality emerges with the strange sensation of an “alien will” intruding on the individual will. The sensuous desiring individual must realize his desires in relation to the hovering *spirit* or idea, the presence-in-absence of a future action—that of a person who intends to take away, by force, stealth, or deceit, the property of another person. The sensuous individual with a concrete desire must then reconcile himself with this ghostly spirit, this vampiric entity seeking to inhabit his flesh.

From the scientific point of view, this generality implicit in the formation of goals may be formulated in third person, objective language: people often have desires to take the property of others, desires on which they frequently act. The social scientist then attempts to explain the existence of such desires and the circumstances that favor acts of theft. From this standpoint of empirical fact, as Hume argues, there is no legitimate transition to duty or right, to moral legitimacy. There are simply the facts involving various kinds of desires and the examination of their possible causes. What we call morality too is a certain kind of desire, the relatively disinterested feeling on the part of some individuals, some of the time, to act in certain selfless or benevolent ways. Morality is simply another type of human behavior that must be taken into consideration. It too has its presumed causes, such as the requirement of society to protect its system of property.

But from the first-person standpoint of the agent the rational or general dimension of the action implicitly takes a distinctive first-person form, the form of right or duty, the form of morality. If “I” believe it is good for “me” to take the property of another person, I cannot help but formulate *the claim* that it is good for persons in general to take the property of other persons. For “I” and “me” are general terms, universals, although they are applied to a singular individual. As sensuous intelligences we are unable to speak the language of individuality *per se*. Following Kant to this extent, Søren Kierkegaard writes, “The ethical as such is the universal, and as the universal it applies to everyone, which can be expressed from another point of view by saying it applies at every instant.”¹⁰ Kierkegaard regards the religious sphere as transcending the ethical. Abraham’s call from God to sacrifice his son Isaac is incomprehensible from the ethical

point of view because it cannot be justified on the plane of universality. For Kierkegaard the religious command is addressed to the singular individual alone, though it is not addressed to the sensuous individual caught up in his desires. Kierkegaard recognizes three planes of action, that of sensuous desire with its external sources of attraction, that of moral universality, and that of religious singularity, which transcends and possibly contradicts the plane of morality. But as the ethical plane is also the plane of language, Abraham is unable to speak about the command of God that is proposed to his religious faith. Hence Kierkegaard writes: “Abraham cannot be mediated, and the same thing can be expressed by saying that he cannot talk. So soon as I talk, I express the universal, and if I do not do so, no one can understand me.”¹¹ Thus morality involves mediation, the intervention of the universal into the desires of the individual.

Essentially following Kant in his understanding of morality, while, like Kierkegaard, seeking to find something different or better to replace it, Nietzsche writes pejoratively of the origin of morality as “bad conscience” in the emergence of human consciousness itself: “I take bad conscience to be a deep-seated malady to which man succumbed under the pressure of the most profound transformation he ever underwent ...” Essentially reiterating Kant’s earlier inquiry into the origins of reason in nature, Nietzsche picturesquely and poignantly describes the transformation of human animals forced to rely on their feeble intelligence in order to survive.¹²

Just as happened in the case of those sea creatures who were forced to become land animals in order to survive, these semi-animals, happily adapted to the wilderness, to war, free roaming, and adventure, were forced to change their nature. Of a sudden they found all their instincts devalued, unhinged. They must walk on legs and carry themselves, where before the water had carried them: a terrible heaviness weighed on them. They felt inapt for the simplest manipulations, for in this new, unknown world they could no longer count on the guidance of their unconscious drives. They were forced to think, deduce, calculate, weigh cause and effect—unhappy people, reduced to their weakest, most fallible organ, their consciousness!

As Kant writes, the goal of happiness “could have been maintained far more surely by instinct than it ever can be by reason.”¹³ Dependent on reason to determine our ends, we human beings are in a sorry state if what we want to achieve by means of our reasoning powers is happiness. Despite the manifold efforts of ethicists in the Western philosophical tradition, unhappiness, not happiness, pervades the precarious terrain of reason. From the standpoint of animal happiness for the individual, the substitution of reason for instinct was an evolutionary disaster. The human species may triumph as a result of the tools of reason, but only to the detriment of the individuals whose happiness is sacrificed on the altar of the species. As soon as we left the realm of animality and entered the realm of consciousness we lost the ability to be full-fledged individuals, preoccupied with individual happiness under the sure guidance of animal instinct. If nature never does anything in vain, Kant argues, she must have given us our reason for some other purpose than as a means for attaining happiness. Otherwise Nietzsche is right and human existence—the existence of rational animals—is a natural catastrophe and a theoretical absurdity.

In his “Conjectural Beginning of Human History” Kant offers a philosophical examination of the mythic-religious origin of Nietzsche’s and Sartre’s conjectures about human beginnings in the *Book of Genesis*. According to *Genesis*, in the beginning all was well because human beings followed the commands of God in the Garden of Eden. How shall we interpret this notion from

the standpoint of philosophy? *Genesis* is here essentially describing animal instinct, says Kant. “In the beginning, [the first man] must have been guided by instinct alone, that voice of God which is obeyed by all the animals.”¹⁴ Instinct is the command of God, or nature’s way of directing the individual animal to the satisfaction of its desires and so to the attainment of its happiness.

Reason, however, implies quite different ends. The incipient rational animal breaks from the commands of instinct, for “reason has this peculiarity that, aided by the imagination, it can create artificial desires which are not only unsupported by natural instinct but actually contrary to it.”¹⁵ The newly awakened rational animal looks around and spies another species following its instincts to eat the fruit of a certain tree, and decides to try out the alternative food source. Violating the natural instinct or command of nature to eat only certain foods and not others, she finds pleasure in representing, in the idealized form of reason-based imagination, the action of eating that strange but appealing fruit. And so she takes the plunge, breaks for the first time from the commands of instinct, and tastes the forbidden fruit. However insignificant the immediate outcome of her action—perhaps she got a slight stomach ache—the overall, long-range result is a calamity for the human animal. For the first time she acts *against* the voice of nature. In other words, for the first time she makes a free choice. The eyes of the first human beings are opened, for they now know both good and evil—and not merely the good prescribed by instinct. The first human being discovers¹⁶

a power of choosing for himself a way of life, of not being bound without alternative to a single way, like the animals. Perhaps the discovery of this advantage created a moment of delight. But of necessity anxiety and alarm as to how he was to deal with this newly discovered power quickly followed; for man was a being who did not yet know either the secret properties or the remote effects of anything. He stood, as it were, at the brink of an abyss. Until that moment, instinct had directed him toward specific objects of desire. But from these there now opened up an infinity of such objects, and he did not yet know how to choose between them. On the other hand, it was impossible for him to return to the state of servitude (i.e., subjection to instinct) from the state of freedom, once he had tasted the latter.

The real fruit that is forbidden by nature, then, is freedom. Freedom is the expression of the power of reason to break from the forces of nature and to represent and deliberate upon alternative courses of action. Reason conveys a power, Kant stresses. It is a power of breaking from natural instinct, from the natural flow of desires. It sets the individual apart from nature. It opens up an unlimited horizon of possibility—i.e., the desire to be infinite, to be God, or God-like, which is the promise made to Eve by the cunning serpent. Without any sure knowledge of long-term effects, the individual is free to choose, and so must choose. Delight in such freedom is immediately countered by *anxiety* before freedom itself, before the infinity of possibility without any guidance from nature. Expressing Kant’s disconcerting thought of the *angoisse* evoked by freedom, Sartre says that the human being is “condemned to be free.”¹⁷ Emerging from nature’s womb, like the individual awakening from childhood, the reasoning being has mounted from the sure ground of nature to another level of existence, to a kind of aerial platform from which all of nature is potentially in view. The dizzying view is exhilarating but also terrifying. Having renounced the guidance of nature, the rational individual must take a step over an abyss. No wonder then that she seeks desperately to return to the state of animality for

guidance by attempting choose out of a multitude of possibilities the course of action that seems to offer the most pleasure. But it is now too late to turn back. The appeal of pleasure can no longer justify by itself. For the agent must confront her desires with the ghostly forms of their ideal representations in consciousness. A new and quite different power looms over the frustrated and insecure man who hates his favored brother. He hears within his mind a voice more powerful than thunder, commanding *Thou shalt not kill*.

In such an “original position,” the tragedy for animal existence of the birth of reason is hardly a recipe on the part of nature for improving on instinct. Reason is a power of representing other possibilities than those originally prescribed by nature. Holding in mind multiple possible actions, the rational individual stimulates ever new desires. Instead of being a means for satisfying desire, an improved method by comparison with nature’s more primitive means, reason in this role is a recipe for endless misery. How could reason be the means for satisfying desire if it ineluctably creates more desires than it fulfills? If happiness is the satisfaction of desire, then the rational individual, through the power of imagination implicit in rationality, inevitably makes himself unhappy by awakening more desires than he could possibly satisfy. The animal contentedly follows its narrow species-determined way to happiness as dictated by instinct. Thanks to reason, however, the rational animal stands before an abyss of unlimited possibility from which he must choose but one limited course of action. And so, in the inevitable loss of all the rest, he is subject to endless dissatisfaction.

If nature does nothing in vain, there must be a different purpose in nature in the emergence of reason than that of improving on the method of instinct for gaining happiness. There must be a providential plan, a secret evolutionary goal, an implicit truth, which is being enacted though this novel, seemingly unnatural situation into which reason plunges the formerly contented animal. It must be nature’s purpose to create a being who, instead of fitting into another limited environmental niche, creates his own nature, a second nature over and above that inherited from biology, a nature for which the human being alone is responsible. The purpose of reason is then not happiness by itself, but the emergence of a self-determining being who is capable of earning or deserving happiness. If the rational individual aims at happiness directly, she discovers that it is no longer possible to reach it, just as one will never get the horizon if that is one’s goal, for the very means used in its attainment pushes it ever further away. But if the goal is something different, if it is a goal that is implicit in reason itself, if reason is not the means but implicitly holds within itself its own ends, then its purpose is no longer a matter of chasing endless artificial desires. And so a *reasonable* happiness may ultimately be possible—a happiness based on moral worth, on whether or not one deserves to be happy.

The Implicit Commands of Reason

This hidden purpose of nature can be deciphered in the existence of the general maxim that every action elicits. Every maxim implies that it is not one’s *individual* happiness that is primarily at stake—as is the case for the animal—but something general or universal. The moral language of duty and right is an expression of the burden that rational consciousness implicitly undertakes as we inevitably formulate our individual goals as the actions of *persons*, of human beings in general. The assumption of our status as individual *representatives of humanity* is the underlying basis of the moral maxim that inevitably accompanies the representation in ideal terms of our

desires and our strategies to satisfy them. Reason implicitly commands the individual to formulate her individual action as something that is right, i.e., good for people in general to do, and not just good for herself as an individual.

We implicitly propose to ourselves the idea that if an action seems good for me it should also be good for others, or at least permissible for me from the standpoint of others. This is because, as a conscious, rational being, I am aware that I am a *human* being, and that the other individual is a human being like myself. Such awareness is implicit in the fact that the pronouns used in the language in which I formulate my goal are general terms, terms I learned from others and that are capable of being used by others. To continue the example of desiring someone's food, implicit in the idea that stands between me and the satisfaction of my desire is the general proposition, "It is right, not just here and now, but in general, for one human being, should he/she so desire, to take food that belongs to another human being." And even more generally, "It is right (good, permissible) for one person to take the property of another." Or, more succinctly, *It is right to steal*. Implicit in every action is some kind of general idea, principle, or maxim that governs the action. As sensuous intellects or desiring intelligences, we are commanded—we command ourselves—to look at the goal of one's action as a *general kind* of goal, a goal mediated by a general idea. We must inevitably consider its ideal, general, or rational properties for their own sake. The sensuous individual seeking to satisfy desires hears from within himself a voice more powerful than thunder. It is his own voice!

Unlike the egocentric child, then, the awakening rational individual is aware that the action he is proposing is not *necessarily* right. He has become aware of others with different desires and so different maxims. He spontaneously tends to think that taking her food is right and good, but he recognizes that she has the opposite point of view. Moreover, when someone else takes his food, he finds himself formulating a quite different maxim in response to this other act of theft. Then he doesn't like it, and so implicitly formulates another maxim: *Thou shalt not steal!* Reason has its own law which commands the individual not to contradict himself. In theoretical reason, this law is empty of particular content. But in relation to practical reason, where the content is yet to be created, it imposes its own content on the individual, commanding him to remain faithful to himself, to create and maintain an identity over time. If I say one thing at one time, I should not say the opposite at another. In this way the law of logic is not merely analytic. From a practical point of view, it implies an *a priori* synthetic law: the maxim I hold now binds me in the future. In violating the law of logic, the individual violates his Self—not the ordinary empirical self that is a bundle of desires which change from moment to moment, for there are many such selves dispersed among the diversity of desires, but the Self of reason, the unifying *personality* that only exists on its own higher plane. Thus Kant argues that source of morality is not the desires of *animality*, and not even the rationality of *humanity*—i.e., reason subordinated to animality—but the *personality*, the autonomous agent who gives the law to herself in connection with all other such agents in a *kingdom of ends*.

Once the person becomes aware of such contradictions between his maxims, he must reconcile them. He must formulate a principle of moral justification. We spontaneously tend to formulate the maxim of the rightness of our desires, but we recognize that one's maxim must be validated or justified, for we also see that the other person naturally regards it as wrong—just as I would if I were in her shoes. Once we have uncovered the implied maxim, the next step is

therefore to evaluate it. The question that we have to ask ourselves is this: In full conscious awareness of what it is that I am intending to do, as I consider the idea I am realizing in terms of its inherent universality, can I will this maxim? Can I consciously accept the maxim that I would be affirming and realizing through this action? Can I, in other words, affirm my maxim as a law that binds me over time in any similar situation?

Take the property of another, when it suits me—this is what my sensuality, and often my self-interest, urge me to do. Yet unlike the animal, which reacts according to a balance of appetite and fear, I represent the goal of my desires in the form of a conscious principle, a maxim of action: “One may take another’s property when it suits one, especially when it doesn’t get one into trouble.” But suppose it is my turn to possess the desirable goods. Will I be consistent with the generalized, rational form of the maxim? Will I consistently maintain the principle that it is right for anyone to take the property of another when this is feasible? Faced with a threat to my own property I am tempted to adopt a different, contradictory maxim. “I don’t want this person to take my property.” Inherent in this desire is the general maxim, “It is wrong to take the property of another person.” Individual desires often change and there is nothing problematic about this. Change, multiplicity, even contradiction, are the laws of lawless desire. But as soon as the maxims implicit in these desires are formulated, the rational individual has a problem. For his own existence, his own Self, is at stake—his existence, his Self, as a conscious, rational being.

When our actions are dominated by the sensuous individuality of our desiring ego we find ourselves shifting abruptly between radically different, contradictory maxims. In one circumstance we are attracted to the maxim that denies the right of property, and yet in another circumstance we find the maxim of affirming property rights to be appealing. We oscillate between the two contradictory maxims depending on what we desire at the moment, or, more thoughtfully, on which maxim happens to be to our advantage. In fact, we actually contradict ourselves in one and the same breath. For do we not want to steal the property of another in order to have that property for ourselves? So, when we look at theft on a general level, we are really denying property (to the other person) to affirm property (for ourselves).

The thief contradicts the denial of property that is implicit in his action by simultaneously asserting that there should be property; there should be property so that there be something for him to steal and so that he may be able to enjoy the result of his theft. For one to be able to steal there must be property belonging to another. The thief therefore implicitly wants there to be property—both so that he can take it from others, and so that he can possess things for himself. And yet his action, taken in its generality, is destructive of that very property that he desires. Consequently the maxim of the thief, regarded in its fundamental essence as a universal idea, reduces to the clear contradiction: the right to property should exist and the right to property should not exist. Perhaps nothing expresses so pathetically the heteronomous nature of the desiring ego left to its own devices than such abrupt oscillation of principles, or such unprincipled duplicity and self-contradiction.

Laws of Reason

At this point, the natural law theorist might employ scholastic argumentation to convince us that respect for property is rooted in human nature, while the attraction of the opposing maxim is the

result of miseducation, bad influences, etc. Our deepest nature impels us to respect another's property, says the natural law theorist, and so this is what we "really desire." We only "think" we want to steal during times of temptation by short-term desires. If we proceed along this course of action we are going against our nature, our natural inclinations. The conflict between the denial and affirmation of property is therefore really a conflict between short-term desires and the fundamental inclinations of our human nature. The utilitarian or the social contract theorist, on the contrary, denies that human nature is so respectful of the rights of others. We are naturally egoists, and each of us is basically concerned only with his or her own property, not anyone else's. However, we should respect the property rights of others, says the social contract theorist, because without such a limiting rule the war of all against all would soon reduce each of us to misery. We should respect the property of others, says the utilitarian, for that is what will make most people (though not everyone) actually (and not, as the natural law theorist would say, "truly") happy.

In the novel, *Crime and Punishment*, Dostoevsky's tormented hero, Raskolnikov, decides that his acts of stealing and murder will make most people happy. He regards the old woman pawn broker he will steal from and murder as a social parasite, preying on the misfortunes of others. With her stolen wealth his own family will be rescued from terrible peril. The completion of his education will allow him to help the poor, contribute to humanity, etc. The social contract theorist might produce a different sort of rationalization. More good than harm might be achieved if the poor student in Dostoevski's novel murders the horrid pawn broker. But from the original position, from behind the veil of ignorance, a person would want to look at things from the pawn broker's perspective too. From the original position one would not know whether or not one would end up being a pawn broker oneself. Her enterprise, after all, accords with economic laws of supply and demand. She does not use force but respects the liberty of her customers. Moreover, such individuals, it could be argued, perform a needed service to the poor. Without them, to whom would the desperate person in need of ready cash turn? Consequently, the liberty to steal the property of others should be denied because it infringes on the equal right of others to own property as well as destroying a form of social inequality from which the poor in fact benefit. This second line of thought may seem dubious to Raskolnikov, should he become interested in social contract theory. Don't pawn brokers only deepen the poverty from which they fatten themselves? In that case, murdering the old woman would be a step toward abolishing an unjustifiable social institution. But then Raskolnikov, a law student, reasoning from the original position, might not want to justify taking the law into his own hands. Suppose someone else were to use the same line of reasoning to conclude with Shakespeare's Dick, considering what he and his bar mates would do when they get charge of things, that, "The first thing we do, let's kill all the lawyers."¹⁸

By contrast to such reflections based on the precarious grounds of satisfying desires and promoting happiness, Kant asks us to consider the proposed action from a qualitatively different standpoint. We should ask ourselves not how best to increase the personal advantages I (or others) might derive from the action, but rather how I should relate to the laws that are actually implicit in my own reason, i.e., be true to my own conscious awareness of what it is that I am doing. Whether the institution of property is itself justified is not what is here at issue. We will see when we turn to Kant's theory of history that Kant has his own answer to this question. Here we are only following the inner logic of the action as it arises in the mind of the thief. He wants

property in the very process of destroying property. Unlike the revolutionary who challenges the institution of property, this person has no such idea. And yet his action implies a kind of devolutionary, if not revolutionary, logic.

The maxims which the thief formulates at different times, and even at the same time, violate the basic law of reason that forbids contradicting oneself. It is irrational or illogical both to affirm and deny the same proposition, without making any significant distinction between them. It is a contradiction for someone to affirm that property should be respected and at the same time to deny that property should be respected. It's even worse to promote the preservation of laws of property, which the thief inevitably does, as a condition for destroying those same laws. The two contradictory maxims of the thief, or the self-contradictory character of his one maxim, are a violation of the basic law of reason against contradicting oneself. The act, which is the realization of such a contradictory idea, is therefore inherently *unintelligible*. To become aware of this fact is the essence of moral consciousness, which is nothing other than reflection on the intrinsic nature one's activity. As a rational, implicitly self-conscious person, the thief must reject this contradictoriness. It is the intrinsic requirement of reason, and not any consequences that may stem from the action in relation to the satisfaction of desires, that compels him to admit that something here is fundamentally wrong.

Who (or What) Am I?

It may not at first be evident that the maxim of the thief is necessarily self-contradictory. The appearance of a contradiction can be removed if there is a significant difference between the two seemingly contradictory statements. For example, it may seem like a contradiction for someone to support the Biblical command, "Thou shalt not kill," and at the same time support at least some wars. But when the matter is looked into more closely, what is really being forbidden is not killing in general, but killing innocent people. There is no contradiction between forbidding the killing of innocent people, on the one hand, and allowing the killing of aggressors in self-defense. What is presented here is not an exception to the rule, but a more specifically formulated rule. This rule is, do not kill except when the person is an unjust aggressor, or a heinous criminal. The exceptions here do not negate the rule, but specify it more exactly. But this is no different from any scientific law which must be formulated in terms of its conditions: "Water always boils at 100 degrees Celsius" requires the more precise specification of its conditions—*at sea level*. This can be put in terms of "exceptions": "Water always boils at 100 degrees Celsius—except when above or below sea level."

It seems as if the thief makes a similar formulation that specifies more exactly what the rule forbids. The thief asserts, "*I* can steal, if it serves my own interests, but *others* shouldn't steal from me." Or, more generally, "Property should be respected, except when it is the property of people other than me." It may not seem at first that this is a self-contradictory or irrational notion. In the two parts of the supposed contradiction there is a clear difference. Property is affirmed when it is *my* property, and it is denied when it is *others'* property. This does not seem to be the same as the simple affirmation and denial of property without any qualification.

When I say that "My property should be respected" while "The property of others need not be respected" I am supposing that there is an important difference between "me" and "others."

No doubt every individual feels that there is a fundamental difference between oneself and others. But the very fact that this is a feeling that *everyone* has means that this is a difference that is the same for everyone. It is a peculiar kind of difference which, when looked as a general statement, applies to everyone, universally. It is not a *specific* difference, the kind of difference that differentiates one *type* of thing from another. If I believe that the United States has a right to govern itself, but Canada does not, this is not, on the surface, a contradiction. But then I have to ask what is the difference between the U.S. and Canada which allows us to deny that right to Canada. If the U.S. has a right to govern itself *because it is a nation*, and because all nations have a right to govern themselves, then there is a contradiction in denying that right to Canada, which is also a nation. If I believe that the U.S. has the right to govern itself merely because that is the country in which *I* live, then someone will naturally reply, “Yes, but someone in Canada might feel the same way about Canada. What makes *you* so special?” Singling yourself out as a special case, for no other reason than that you are yourself, does not establish the kind of specific difference that would remove the appearance of contradiction. At least, this is how we think when we think rationally.

Similarly, the maxim of the thief who denies property to others while affirming it for himself is one that everyone might tend to *feel*, since we all have a strong *sense* of our own individuality. Saying that, however, is not to point out a *rationally* distinctive difference between myself and others but to recognize a characteristic we all have in common. *We all* tend to regard ourselves egocentrically, like the young Baudelaire, as “*son by divine right*... an emanation of the divinity.” But insofar as I refer to myself simply as “I,” there is no *rational* difference between “me” and others, who also refer to themselves, in the same way, as “I” or “me.”

There are many differences between myself and others. E.g., my mother and father were different from someone else’s mother and father. But in the maxim of the thief it is not because of his mother and father that he claims a privileged right to property—he will admit that there is nothing special about them—but simply because he is himself. It may be because of my mother and father that I can claim the right to possess this particular house, which I have inherited from them. But this right supposes a general rule: individuals have the right to inherit property from their parents. This, however, is a rule that applies to the children of other parents, as well as to me. But the maxim of the thief that we are now considering supposes the right to steal simply because the thief is himself. But this is no *specific* difference that distinguishes him from others. It is not a rationally distinguishable difference. “I have a right to steal because I am I,” when regarded rationally, i.e., at the level of generality or universality, is the same as saying, “Every person has the right to steal.” What the individual *means* to say cannot in fact be said—the immediate sensuous certainty of one’s animal existence. But by virtue of their rationality humans inevitably go beyond the immediacy of their animal existence, representing themselves to themselves in general terms, through ideal models of selfhood.

The term “I,” which reflects the experience of personal existence, has a twofold meaning. In saying “I,” I may *mean* to focus exclusively on my own individuality, my sensuous uniqueness, my individual existence as something separate from that of others. But the term “I” is a general term that gives unity to all my experience as mine. Kant calls this “I” the transcendental unity of apperception. Like the categories of cause and effect, it is a fundamental condition of experiencing that one recognize the experience as *one’s own* experience. This

condition of experiencing is universal—an *a priori* synthetic condition of human experience. As I perceive the house before me, or the boat moving down the river, it is always “I” that am the center of the experience. A great variety of experiences, past, present, and future, are unified as *my* experiences. Kant here follows Leibniz. According to Leibniz, animals have perception, but only humans are capable of that self-awareness in perceiving other things that he calls “apperception.” Leibniz is continuing the insight of Descartes into the fundamental starting-point of all human experience: “I think,” in the sense of self-awareness. In practical experience, “I” am not only the center of experiences coming to me from the world, but the origin of practical actions that I initiate when I propose to do something. I think of myself as the original source of my actions. Who or what then am “I” that I must think of myself as having such God-like creative power to initiate a chain of causes by my own *fiat*?

The critique of pure reason, of course, forbids directly turning such a necessary condition of experience into a spiritualistic metaphysical reality. The passage from “I think” to “I, the thinking substance, exist” is fraught with the subterfuges of transcendental subreption. But awareness of such problems also prevents the opposite materialist metaphysics from taking hold, which declares consciousness to be nothing but a product of the forces of nature. The transcendental unity of apperception of the first *Critique* gives rise to the noumenal idea of a free, self-determining subject of action, which is the underlying supposition or belief or postulate of moral consciousness. Within the framework of rational faith, practical reason provides new experiential content, new forms of evidence, to enrich this necessary idea of theoretical consciousness.

Implicitly, then, I am aware of the transcendental character of this “I.” “I” is a fundamental idea, not just a sensuous feeling. An explicit recognition of this common characteristic of all human beings to regard themselves as the sources of their own actions can itself be a profoundly significant moral experience. As a thinking, rational being I implicitly recognize that each of us is an “I.” I consequently recognize that in permitting theft in my own case, “I” am “legislating” for others as well. If it is right for me to steal simply because I am myself, then it is equally right for someone else, for all other “Is” or “selves.” Unless I can discover some specific reason why *I* should be allowed to steal, I cannot consciously, logically, or rationally deny this right to anyone else. Perhaps I am stealing *because my children are starving*. That is not the same as the maxim of the thief considered above. It implies a universalizable rule: People have the right to steal from individuals with a surplus of goods to save the lives of their children. Or, people have a right to property, except when this right conflicts with the survival of needy people. Such a maxim contains no contradiction if the person who steals to preserve life regards himself as also subject to it should a similar situation arise and someone else urgently needs food that he has in surplus.

It is important to stress that what is at issue here for the distinctly moral consideration is not the goal to be achieved, but the nature of the action in itself regarded in its universality. Of course I want my children to live. Actions generally begin with considerations of utility, with goals one wants to achieve that involve the satisfaction of desires and interests, whether for oneself or for those one is concerned with. My desire to feed myself and my children arises out of my difficult situation. Stealing the bread of life just as readily comes to mind as the means for achieving this natural goal. But to attain this goal, I could kill the children of my neighbor and

reach my goal by this means. I could steal the food of other poor people, condemning them to starvation, so that *my* children can live. However, by taking food that someone else has in abundance, by robbing the rich to give to the poor, I am engaged in an altogether different enterprise. In addition to the goals I want to achieve, I must consider the law that is implicit in my action, a law that exists over and above the ends I wish to accomplish. I ask myself whether the law of my action, the law implicit in the concept of my action, is a law I can live with in all its implications for myself or for others in the future. Can I bind myself to live according to that law when the tables are turned and I find myself in comfortable circumstances facing the individual in desperation who takes my bread for *his* children?

However, the simple maxim of the thief that we are considering—I should be able to steal because I am *me!*—cannot be justified before the tribunal of the thief’s own consciousness, should he fully awaken to the rule that he is in fact proclaiming and attempting to realize. As soon as the intention that he is putting into reality is affirmed it logically contradicts and pragmatically destroys itself. If I affirm stealing as legitimate for me, simply because I am myself, then I am at the same time giving others the right to steal from me. Since, on the other hand, I at the same time do not intend to give others the right to steal from me, because I want a certain property for myself, it follows that I am contradicting *myself*. The thief is ultimately acting against the deepest part of himself, his fundamental humanity, the rational consciousness that raises him above his animal existence.

The Constitution of the Self

Consequently, the maxim of the simple thief has to be rejected as irrational. The thief contradicts himself, not as this individual sensuous being, but as an intelligent or conscious being who implicitly recognizes the universality of the “I” in each of us. We should think seriously about this fact of *self*-contradiction. The author of the law that we are affirming is none other than one’s Self. The law that one breaks, in trying to make an exception for oneself, is *one’s own* law. It is not a heteronomous law laid down by society, or an external and arbitrary command of God. The laws of morality are autonomous laws, laws that individuals give to themselves. As a merely sensuous individual, of course, there is no question of one’s being a self-conscious agent. On this level of sensuous individual existence, one follows one’s desires as they are blown in different directions by the winds of circumstance. But as rational or thinking beings we turn back from the whirlwind of desire that impels us in various directions toward a center of selfhood, and question ourselves. Following Kant in this respect, Hannah Arendt writes of moral thinking as “the two-in-one of the soundless dialogue.” Such thinking, she writes, “actualizes the difference within our identity as given in consciousness and thereby results in conscience as its by-product.”¹⁹ It is this dialog with oneself that is characteristic of apperception—not just perception of what is outside oneself, but reflective perception of oneself as a necessary condition of such outwardly directed perception. The moral dialogue is what constitutes the self as a relation to itself.

If from a theoretical point of view the transcendental unity of apperception stamps my experiences as mine, moral conscience is the self checking with itself to see whether it *practically* preserves its identity in all its actions. Morality is the expression of an ideal identity as the source of a practical law to which personal empirical existence, with its vagueness, multiformities, and inconsistencies, should be subject. Paradoxically, the individual who puts his

self-interest above everything else and is supposedly self-centered, really has no self, no center. The ego-consciousness that identifies with sensuous existence is always looking to the outside world to determine what to be or do under present circumstances. Its pretence to be a center of existence turns out to be an illusion. No doubt there are great villains in literature and life who succeed in subjecting circumstances to their self-aggrandizement. The greatest scandal for morality consists in the apparently successful lives of scoundrels who are never brought to justice. To solve this problem the great ethical religions have promised justice in another life. But moralists concerned with the here and now might find some satisfaction in wondering, in relation to such individuals, whether anything persists that deserves the term self or person, other than an individual mortal body, a physical location around which wealth is accumulated. Peering into the souls of criminals, Kant finds that there is self-punishment in the always accusing voice of conscience, in the discord of divided consciousness, in the self-destruction of the self:²⁰

There is no one, not even the most hardened scoundrel—provided only he is accustomed to use reason in other ways—who, when presented with examples of honesty in purpose, of faithfulness to good maxims, of sympathy, and of kindness towards all (even when these are bound up with great sacrifices of advantage and comfort), does not wish that he too might be a man of like spirit.

We can see now that it is through such reflective thinking-with-oneself that a person establishes, or at least envisages, real personal unity—that stable core of self that is missing in the example of the amoralist. The moral individual is the principled individual. Such principled existence requires moments of reflection, times when one addresses oneself in an aloneness that is still social, because one is at the same time a relationship of two-in-one. At this level, as the postulated substance of practical life, the rational self is an intrinsically social, an intrinsically universal being. For the self that he contemplates is a representative of humanity. Without such silent communing-with-oneself I can be one person in one situation, and someone quite different in some other situation. But the division in the self is worse than this, for I can never really separate myself off into different selves for different occasions. The unity of self, rooted in self-conscious thinking or awareness (the “I” that Descartes establishes as the starting point of philosophical science) is always implicit in one’s particular actions. When I affirm different contradictory maxims I am not merely following different principles at different times. For the self is not a momentary entity. The self is a principle of temporal unity, a unity that is beyond time and so the basis for connecting the different moments of time into the coherence of the true self. Thinking, rational-moral thinking, calls us back to ourselves, to consider the nature of what we are doing *as persons*. Morality is the centripetal force that counteracts the centrifugal forces produced by our sensuous desires and circumstances. It is the force of attraction on the plane of conscious life, paralleling the force of gravity in material existence. In drawing ourselves to ourselves, it also draws together all authentic selves. The community of self with self is an expression of a community between selves which is implicit in the universality of reason and the very source of that universality. In my individual action, therefore, I implicitly legislate both for and with all of humanity.

To be the law-giver in morality, by listening to the inner voice of conscience, is ultimately the imperative of my own conscious awareness. This is what Kant means by pure practical reason. When I act immorally, I do not violate some law laid down for me by someone else, by

an alien law-giver. The voice of the moral command seems alien only to that state of consciousness that is immersed in sensuous desires. Because this contrast is an inevitable feature of human duality, morality takes the form of duty. But when I recognize myself as a rational agent I recognize that this voice is truly my own. In acting inconsistently I violate my own law. I contradict *myself*. I contradict myself as a rational or conscious being, and not as a sensuous individual. In fact, it is only as a rational being that I have a unified self to contradict. Sensuous individuality, or ego, which I often want to mean when I say “I,” is not self-governing. Its laws are “other-determined,” heteronomous, not “self-determined,” not autonomous. In acting morally, however, I act according to a law I give myself. In respecting the right of property simply *because* to steal is to contradict *myself*, I act autonomously. As rational beings, therefore, we are inherently self-ruling, truly free law-givers.

But if morality directs us to a self, to a center and source at the heart of the multiplicity of our actions, it simultaneously directs us to other persons as other selves. The self that we discover in our moral reflection—as we turn back from the centrifugal repulsions occasioned by the world of competing, threatening egos—is not the individualistic “self” or ego dominated by desires. It is the self that we recognize as equally central to the human dignity of other persons. The ego is *my* self, in separation from the other egos. But the transcendental “I” of free action affirms the simultaneous dignity of all other self-aware centers of action. This affirmation of equal dignity is inherent in the nature of consciousness with its universality of “I.” The moral center of gravity that makes us truly individual persons, self-determining centers of action, at the same time links us to others on this same plane of noumenal existence. The laws we give ourselves are not willful, arbitrary, singular laws, if the term “law” can be given to arbitrary decrees. True laws are not merely “laws for me.” The reasoning that obliges us to condemn theft because it involves self-contradiction is valid for any person with the same maxim. Even the thief recognizes the impossibility of making his action a universal law. As long as we suppose that he acts consciously, that he understands what he is doing, that he wants there to be property even while he does that which is within his power to destroy property, we must suppose that he understands that what he is doing is wrong, which means that what he is doing is self-contradictory, destructive of the object that he desires as well as of himself as a human being. Certainly not every thief will pay equal attention to the voice of moral thought, the voice of conscience. The voice of morality grows ever more faint to the distracted mind that fears to look within itself. And yet, in despising those suckers who are born every minute, but who nevertheless make his life possible, he implicitly acknowledges his own parasitical nature. He depends on those he pretends to scorn. He depends on the law and must will its existence, as a necessary condition for his own lawlessness. And so, when presented with examples of honesty in purpose, of faithfulness to good maxims, of sympathy, and of kindness towards all, somewhere deep in the well-springs of his murky existence from which his humanity flows, he inevitably wishes that he were of like spirit.

Willing the Maxim as a Universal Law

Instead of letting our actions be guided by our own feelings and desires, or by what we believe will satisfy the desires of others, let us acknowledge the law that is dictated by our own reason. This law of pure practical reason, this principle of morality, can be formulated as follows: “I

ought never to act except in such a way *that I can also will that my maxim should become a universal law.*"²¹

In his example of the person who is tempted to make a false promise, Kant first points to the difference between acting out of fear of harmful consequences and acting out of principle, i.e., acting out of "the concept of the action." If I look to anticipated consequences of keeping, or failing to keep, my promises, I am judging my action by effects which are determined by a host of causes—my own individual circumstances, needs, and desires, as well as the actions of others, motivated by various and often contradictory goals. In such a consequentialist or utilitarian frame of mind, "I have first of all to look around elsewhere," letting myself be guided by circumstances that are independent of my will. With such a framework of evaluation, however well-intentioned I may be, I inevitably become subject to heteronomous laws arising out of this welter of diverse and conflicting goals, laws that are not determined by myself but by some alien "other." Theoretically, I must be open to the possibility that making false-promises might produce more good than harm, and therefore could be justified by its consequences. But what can it mean to say that it might be a good thing to make false promises, to follow a law of making promises without the intention of keeping them? Whatever might be the good that theoretically could be achieved by such a law, it must be recognized that such a would-be law, looked at in itself, makes no sense at all. For promises one does not intend to keep are not promises at all. No law is therefore possible on this basis. What is possible, is that while admitting the existence of such a law, an individual decides to violate it. But then he must recognize that his action depends on the existence of a law that he plays no part in maintaining. There is no true, positive freedom here, however negatively "free" I may be from outside interference. For the law-breaker depends on others to maintain the law in order to accomplish his goals. On the other hand, when I act out of principle, from "the concept of the action," I am taking responsibility for the law of my action, rather than abdicating that responsibility to others.

The maxim of the promise-breaker might be formulated as, "Keep promises when it is convenient or prudent to do so, but when keeping a promise is inconvenient or against my interests I will break my promise." But this maxim is self-contradictory. The concept of promise-breaking at one and the same time both presupposes the existence of promising as a basic social practice, and denies that existence. Of course, I try to limit this denial to myself alone. I want promises to be kept most of the time, only not when it is inconvenient to me. Hence, I don't want my maxim to be universalized—I don't want everyone to follow my principle, since I want others to keep *their* promises. So I *do not* affirm this exception I make for myself as a general principle, as a law for everyone. If this principle I hold for myself were to become a universal law, no one would believe me when I made a false promise. For my stratagem to pay off, I must will the existence of promise-keeping even as I in fact destroy that social practice by my particular action. I do so first of all ideally and in a thorough-going way, and then I do so materially to the extent of my possibility. The fact that my individual material action seems only very minimally to damage the practice of promise-keeping is an illusion. Striking at the very root of this practice in the idea itself is like spreading a virus that could become a plague. And yet I don't want there to be such a plague, and hope that others will resist the contagion. The fact that it is (still) practically possible to break promises is due to the fact that other people keep theirs. My action is parasitical on the very practice that I both affirm and simultaneously implicitly destroy. All social practices have their origins in the ideas of human agents and so when I violate

this idea I am striking at the heart of human existence. Unlike those laws of nature that exist independently of anyone's will, unlike the laws of animal existence that are governed by instinct, human beings also act according to a certain kind of distinctly human law that exists only because of their choices.

Therefore, I can't, as Kant puts it, will this maxim as a universal law. Related to this inability is the fact that I can't announce my maxim publicly. I can't say to the other person, "I will keep this promise, unless it becomes inconvenient for me to do so." Such a statement contradicts the very "concept of the action." No promising at all is possible with such publicity given to my promise-breaking maxim. Earlier we noted that no individual could reasonably announce to his employer that he would regard the pursuit of his own interests as his "sacred duty." Here we see a general application of this moral absurdity in the fact that no one can *say* that he would keep promises only when it pleased her or worked to her advantage. Morality is therefore in accord with public language. Immorality supposes a contradiction between our public speech and our private thoughts. Here again we contradict ourselves by being obliged to announce publicly our adherence to norms implicit in the very concept of promises, while privately rejecting those norms. Thus, as Kierkegaard points out in the case of Abraham on his road to the sacrifice of his beloved son Isaac, the moral lawbreaker can only keep silent.

False Positives and False Negatives

Kant's conception is sometimes criticized as leaving room for "false positives," maxims that seemingly can be universalized but that are obviously immoral, as well as allowing for "false negatives," maxims that seem to be ruled out by this procedure but that are obviously permissible. Suppose I have a gullible friend, Joe, and know that I can easily borrow money from Joe without having to repay it—as long as I promise that I will. Joe is so good-natured and forgetful that I can always tap him for ready cash. However, there is nothing contradictory in universalizing the maxim, "let everyone borrow from Joe Smith when they need money, without intending to repay Joe."²² Here is a so-called false positive: I seem to be able to universalize the maxim, and yet it is obviously immoral.

The film *Rain Man* offers a graphic example of such a situation. Charlie Babbit, played by Tom Cruise, discovers that he has a brother, Raymond, played by Dustin Hoffman, who is an "autistic savant." Charlie learns that his father, unjustly as Charlie sees it, left three million dollars to Raymond, whose name Charlie as a small child once translated as "rain man" By claiming Raymond as his brother, and taking him out of the institution in which he finds him, Charlie hopes to take for himself the money left for the care of his brother. Consider, then, making a maxim that describes such a unique and unusual situation and formulating it as a universal law: whenever someone has an autistic brother, who doesn't even understand the meaning of money, he will take the money by making all sorts of false promises. But clearly there is a more general maxim that underlies this specific one. Susanna, Charlie's girlfriend, puts this aptly: "You use me, you use Raymond, you use everybody." Kant formulates the underlying principle this way: never use humanity, in your own person or in another, as a means only, but always as an end in itself. Susanna says it more succinctly: don't use people. Making false promises is only one way of using people. But people should not be used simply as means to one's own ends, while their ends are ignored, manipulated, steamrolled or bulldozed over.²³ Why

so? When *I* use others as means to *my* own ends, I am proposing a law for all “I’s.” This is not only a law I cannot accept for myself; it is a law that is inherently incomprehensible when properly understood. It simultaneously affirms that persons act on the basis of ends that they propose for themselves and, by regarding (other) persons as means only, denies that essential feature of humanity.

The seemingly unique case of Raymond is not really that unique when it comes to fundamentals. He is a vulnerable person and so can be manipulated by others with particular ease. But for that very reason, the manipulation is all the more despicable. The fact that only one person fits someone’s idea of when it is possible to make a false promise safely does not make that case any less an instance of a general rule. We need only ask, *why* do you choose Joe to borrow money from? And the answer is, because he is a safe bet, because he is a vulnerable individual. It may be possible to manipulate moral rules by making them seem so specific as to be harmless at the level of generality, but if we could look into the heart of the manipulator we would see the game he is playing, perhaps with himself as well as before the court of public opinion. Attempting to avoid the universalizability implications of his action, the individual pleads: “I only did it this once.” And prays: “Dear God, let me get away with it just this once, and I’ll never do it again.” These are clearly excuses in which a person tries to extract himself from the general rule.²⁴

The case of Joe Smith, when universalized, is the case of a human being. The “I” of the maxim is also a human being. So when universalized the seemingly specific Joe Smith maxim becomes, “Human beings may make false promises, promises they do not intend to keep, when they can safely do so.” The fact that this sensuous individual, this *I*, can only realize the maxim only in relation to that sensuous individual, *Joe Smith*, is irrelevant on the plane of universality that is implicit in the individual situation.

The argument that Kant’s universalizability principle allows for “false negatives” is based on such maxims as, “I will play tennis on Sunday mornings when most people are in church, and so the courts are generally free.” When universalized, this allegedly becomes “Everyone should play tennis on Sunday mornings ...” which is not universalizable, for it defeats the purpose of the maxim. But does that make playing tennis on Sunday mornings immoral? Kant’s universalizability rule seems to turn a clearly harmless and permissible activity into something immoral. We often act on the basis of our perception of exceptional circumstances, situations in which everyone does not follow our rule. We buy stocks when everyone else is selling, in hopes of making the financial kill. Is that really immoral, or just playing the rules of game better than everyone else? What about the collector who buys model trains, but never sells them? If everyone did the same, there would be no trains to buy.²⁵

But it is necessary to ask, *why* the individual has a rule of playing tennis on Sunday mornings. The specific maxim, to play tennis on Sunday mornings, is the result of reflecting on an underlying goal or objective that must be brought forward to grasp the underlying maxim. There is a deeper rule, underlying this specific application, the rule of prudence: people should do what they perceive to be in their own self-interest. Here we are dealing with those hypothetical imperatives that pertain to the pursuit of happiness. The specific rules are pragmatically based rules of thumb for implementing a rule of prudence. The rules for attaining happiness are diverse. They are different for different individuals. They are not absolute rules

guaranteed to produce their desired goal of a happy life. Hence, Kant argues, the laws of morality cannot be founded on alleged universal rules of happiness—contrary to the Aristotelians.

Such rules need to be nestled within a higher set of rules: everyone has the right to pursue happiness, or what they perceive to be in their self-interest, *as long as* they do not violate the equal right of others to do the same. The limited formulation, do whatever is in your self-interest, period, is clearly false and is not universalizable. It leads to the Hobbesian war of all against all. It cannot be willed as a law for oneself, because that means willing that others trample on one's own freedom of action. The true universalization of the rule thus has this limitation involving respecting the rights of others to pursue happiness as they see it, such as by going to church on Sunday morning. There is nothing wrong with exceptional behavior, as each person seeks to excel in some way or another. But not by violating the rules of the game itself—not, for example, by insider trading on the stock market.

Tearing the Fabric of Society

It is important to understand how promise-keeping, when done as a matter of duty rather than prudence, is an expression of autonomy—of living according to laws we give ourselves. There are of course regularities or laws governing social life as well as natural life. Stable social relations over time are essential to human existence. People must do certain things in order to maintain their own existence. In an advanced civilization, a complex network of social activities, involving an extensive division of labor, must be maintained for individuals to satisfy even the simple biological need for food. Relatively stable relations result from innumerable acts of human will. If Adam Smith is right, the only psychological motive needed to understand social laws is that of personal satisfaction or self-interest. Each seeks to adapt one's actions to the combined actions of others in ways advantageous to oneself. General patterns of social life unconsciously emerge out of this adaptive self-interest on the part of everyone, as an unintended by-product of purely personal goals. People do not however act *in order to* maintain such necessary patterns of social life.

Every workday morning I go to my place of work, and because others also do the same the job gets done. But I do this for a wage, to support myself and my family, not for the sake of the work itself. I don't work for the sake of the work itself, but rather adapt myself to external social conditions in order to benefit myself as best I can. Those external conditions are the result of the activities of other individuals, similarly adapting themselves to their given social conditions, of which my own activity is a part. The complex social world that results is thus the unconscious byproduct of the self-interested activities of innumerable individuals. There is nothing specifically human about this. Plants and animals too adapt themselves to their environment. This is an environment made up in large part of the activities of other plants and animals. As a result of the behavior of all the plants and animals adapting to one another, the diverse system of the biological world is in fact created. Although each implicitly contributes something to the creation of its environment, each plant and animal directly seeks only its own survival in an environment that confronts it as something external and to which it must adapt.

While human beings employ reason as a means to achieving their personal ends, their

adaptive behavior continues to be subject to heteronomous laws, even when these laws are in fact the result of their own combined acts of will. They normally do not intend such laws or directly will them, for they directly intend only their own survival and well-being. It is the social scientist who, like Adam Smith, discovers the laws that operate as an “invisible hand” in guiding the actions of innumerable individuals in accord with the common welfare. Unlike other animals, however, human beings can recognize the laws governing their behavior. We know that without the implicit cooperation of other human beings we would perish. Normally, however, such awareness is contemplative, rather than practical. It is an observation we can make, but we don’t employ this knowledge in a practical way. If our knowledge of the larger whole becomes practical it is employed only to further our own individual aims. For example, a person who knows a lot about the stock market may buy stocks when most other people are selling. In acting in this way he is implicitly contributing to the laws of the market. He doesn’t however act for the purpose of maintaining laws of the stock market itself, but merely for his own well-being.

In morally motivated activity, however, simple respect for laws becomes the main motive of activity. For example, through promises we bind our relations to others over time, transcending the impulses of the moment and the contingencies of circumstances. The principle of promising is that my word is good in the future. I bind myself in the future, and refuse to allow my passions and the enticements and threats of variable circumstances to change my will. The whole vast web of society can be regarded as the outcome of innumerable promises that are generally being kept. For the most part, however, such promise-keeping results from “prudence,” from the advantage I get from keeping my promises weighed against the disadvantages I suffer from breaking them. Society operates in various ways to penalize promise-breakers and so to maintain itself in existence to the extent that it depends on the social practice of promising and keeping promises. Frequently, promises are broken, but as long as such acts are relatively minimal the social fabric is only slightly tattered.

Each of us recognizes that without the keeping of promises, social life would be impossible. But I recognize too that there will be no great collapse of society if I break my particular promise in this particular situation. I am willing to do this because I believe that there will be no bad consequence to me for breaking my promise. Least of all do I think that society itself will collapse. What I am doing when I break a promise is allowing others to preserve the basic fabric of society, while I momentarily or minimally fray it. In making an exception for myself from the general rule, while depending on others to preserve the functioning of that rule, I am clearly acting heteronomously. I do not take the rule as something whose realization is *my own* responsibility, but I depend on *others* to maintain this law that is fundamental to my existence.

In traditional Christian ethical-religious language, this is described as acting from the standpoint of the “natural man.” The “natural man,” the sensuous individual or “rational animal,” feels the constraint of duty as an outside force, as a command to which one is loathe to submit. But this natural individual is in fact only a slave to desires and impulses, to the pressures of interpersonal life, to the adaptations required by the market, or to the limits to action and threats of punishment imposed by the state or envisioned by religion. When, however, I act with a view to respecting the law itself, particularly if it is not to my advantage to do so, I am acting autonomously. I am taking this law of social existence—that promises be kept—and consciously making the realization of this law the determining principle of my action. I am thereby taking

personal responsibility for society itself. I behave as if society depended on me. It is precisely in this capacity of the rational-moral will to set aside desires, inclinations, and natural fellow-feeling, and to respect and follow its own law, that its liberating power—its capacity to create the authentically free human personality—lies. It is precisely through the inner commands of morality that we discover in ourselves a capacity to act independently, with real self-determination. Acting morally, acting out of respect for the principle of the action itself, we act according to the law which we create ourselves.

There is something majestic or regal in this idea. In promising, I give my word, and bind myself by it. An Arabian prince in Walter Farley's *Black Stallion* promises the foal of "the Black" to the American boy who befriended the prince's horse when the two were shipwrecked on an isolated island. At the start of the sequel, the boy begins to doubt whether the promise will be kept. When the promised colt at last arrives, he is ashamed of his former uncertainty. The word of a prince, he realizes, binds absolutely, even over himself. As ruler of the kingdom of morality, each person ought to exhibit such dignified royalty.

ENDNOTES

¹ Hobbes, *Leviathan*, Part I, Chapter. XIV; GB, Vol. 23, 86.

² *Critique of Practical Reason*, *op. cit.*, 90.

³ See Z. M. Orudjev, *The Nature of Man and the Sense of History* (Moscow: Librokom Publishing House, 2009), 133-159. In Russian.

⁴ As Allen Wood puts this argument, "Some reject the idea of categorical imperatives ... because a categorical imperative is supposed to bind us irrespective of any desire for an end, and they think that actions are always produced by a desire for something plus a belief on how to get it. The desire supplies the motivation for the action, while the belief determines which actions are rational to perform once the end has been sent.... The threat to rationality in pursuing an end is not that our desires will not be informed by the right beliefs about how to achieve the end, but rather that when the time comes to perform the necessary action, our desires may no longer conform to the norms of conduct we established in setting the end. The function of instrumental reason is not to inform desire regarding means but to constrain the will to hold to its rational plan to pursue an end, perhaps even in the face of distracting or contrary desires that tempt the will to abandon the plan." Thus it is not the desire that moves us to act in such cases of hypothetical imperatives, but the "desire from which we act is a function of instrumental reason only if it is produced by pure reason, in accordance with an *a priori* principle ..." such as, whoever wills the end must will the means. *Kant's Ethical Thought*, 64-5.

⁵ *Groundwork* 4:417; 28 (Cambridge edition).

⁶ As does Allen Wood, *ibid.*

⁷ Allen Wood finds fault with Kant's conception that the desire for happiness is universal and necessary. If he means that everyone has desires to satisfy, this is true but uninteresting. If he means happiness in the collective, rather than distributive sense, as a sum of all desires, then it is simply false that we all pursue such a sum. Often we put our individual desires ahead of our happiness in this sense, as does the gout sufferer who chooses to eat rich food rather than sacrifice this specific pleasure for the dubious and remote value of a promised happiness. To take happiness as an end by natural necessity, moreover, contradicts Kant's argument that all our choices are free. Wood concludes that "it belongs to the essence of rationality that a rational being is bound to form the idea of its happiness and make that happiness an end." Wood, *op. cit.*, 66. This analysis of the desire for happiness as either consisting of satisfying particular desires or as a totality of desires misses the middle ground that Kant intends by calling happiness an ideal. We are moved by the ideal of happiness, but not in the same way we are moved by instances of the ideal, and so

despite its real power to move us the ideal of happiness can fail to motivate in the face of a strongly competing desire.

⁸ Berthold Brecht, *Galileo*, trans. Charles Laughton, in *Seven Plays by Berthold Brecht*, ed. Eric Bentley (New York: Grove Press, Inc., 1961), 379–81.

⁹ Jean-Paul Sartre, *Baudelaire* (New York: New Directions Paperback, 1950), pp. 16-17.

10. Søren Kierkegaard, *Fear and Trembling and The Sickness Unto Death*, Walter Lowrie, trans. (New York: Doubleday & Company, 1954). 64.

11. *Ibid.*, 70.

¹² Friedrich Nietzsche, *The Birth of Tragedy & The Genealogy of Morals* (New York: Doubleday Anchor Books, 1956), 217.

¹³ Kant, *Groundwork of the Metaphysics of Morals* (New York: Harper Torchbooks, 1956), 62-63 [359].

¹⁴ Immanuel Kant, “Conjectural Beginning of Human History,” in *Kant On History*, ed. by Lewis White Beck (New York: Macmillan Publishing Co., 1986), 55; 111 of Prussian Academy edition, volume 7/8.

¹⁵ *Ibid.*, 56; 111.

¹⁶ *Ibid.*, 56; 112.

¹⁷ Jean-Paul Sartre, *Being and Nothingness*, trans. by Hazel Barnes (New York, Washington Square Press, 1972), 567.

¹⁸ William Shakespeare, King Henry VI, Pt. II, IV, ii, 86.

¹⁹ Hannah Arendt, “Thinking and Moral Considerations,” *Social Research* 38 (1971): 445-46; cited from Ronald Beiner’s “Interpretive Essay” in Hannah Arendt, *Lectures on Kant’s Political Philosophy* (Chicago: The University of Chicago Press, 1982), 110.

²⁰ Kant, *Groundwork*, 4: 454; 122 (Paton edition).

²¹ Kant, *Groundwork*, 4: 402; 70.

²² Regarding such an example, Allen Wood argues that Kant’s universalizability formulas fail for this reason. He concludes that there must be some way to determine what the moral laws are ahead of time, so as to be able to produce the right kind of maxims on this basis. See Wood, *op. cit.*, 102-105.

²³ Allen Wood is happier with this “second formulation”, the formula of humanity, than with the first formula of universal law. But the second rests on the first, for it rests on recognizing in each individual a universal truth, that person’s humanity.

²⁴ Allen Wood writes regarding the universalizability principle or “formula of universal law” (FUL), “if FUL is really supposed to supply us with a rigorous test of the morality of maxims, it ought to work not only for the maxims people commonly act on, but also, one would hope, even for strange maxims that would be adopted only under very unusual circumstances or by a very weird agent.” Wood, *op. cit.*, 103. But the purpose of Kant’s procedure is to guide the person of good will from within, not to provide an ironclad rule that would cover all conceivable cases that can be imagined. Kant might say: “You know what your real maxim is.” But then self-deception or what Sartre calls bad faith, must also be recognized. Wood notes that “Kant is very much aware of our tendency to deceive ourselves in such matters.” (53)

Arguing against Onora O’Neill’s position that the appropriate maxims consist of “those underlying principles or intentions by which we guide and control our more specific intentions” (O’Neill, *Constructions of Reason*, 84), Wood argues that this doesn’t tell us what part of the underlying maxim, or what level, is the appropriate one. Thus, the more fundamental underlying maxim for the case that Kant considers might be “borrowing money from a friend when in financial distress,” omitting the crucial, but less fundamental feature “without intending to repay it.” See Wood, *op. cit.*, 104. When O’Neill refers to the maxims by which “we guide” our actions, she is referring to the first

person perspective of the individual's own conscience. Charlie Babbit may be able to rationalize his action before a judge in a law court, but deep down he knows what his real intention is. He is not just "borrowing from a brother," as he might put it in a court of law. Moral maxims operate in the inner court of human intentions and conscience, and are not meant to serve as justifying principles in a court of law—as the arguments put forward by Wood implicitly suggest.

²⁵ Wood, *op. cit.*, 105-7, presents these examples, as showing the failure of the universalizability test. He cites Christine Korsgaard (in *Creating the Kingdom of Ends*, 92) as making similar objections to the formula of universal law, and then as arguing that the only kind of exception that Kant rejects is "free riding," where an individual makes himself an exception to a recognized moral law. Wood replies that this supposes that we know already what the universal moral law is, something that the universalizability formulas do not establish. Moreover, he says, free riding on public transportation is actually a good thing, ecologically sound and socially progressive. (Wood, *op. cit.*, 360.) We will return to this question of exceptions in chapter eight.