Economists have developed a very powerful model of human decision-making, often personified in the metaphor of “economic man” or *homo economicus*. While certainly not without its share of criticism, from both mainstream and heterodox economists as well as other social scientists and philosophers, this model has proven extremely useful in helping us explain countless aspects and examples of human behavior, from common business decisions to government policy-making, and from choosing a life of crime to selecting a mate. Recent developments in behavioral economics, which question some of the core assumptions of the standard economic model of decision-making, have illuminated many standard deficiencies in rational decision-making, resulting in an even richer conception of human choice.

But despite the success of these models of choice in explaining countless types of behavior, they have struggled to explain behavior motivated by ethical concerns. The simplest approach is simply to introduce a “taste for morality” into the standard set of preferences or utility function; in such a model, ethical actors are simply indulging one preference out of many. Either this approach extends the usual assumption of the self-interested agent to include other-regarding preferences, or it simply subsumes such preferences within the very concept of self-interest, which stretches
and thereby weakens the term beyond all recognition. Other approaches to modeling ethical behavior involve interdependent preferences, wherein one person has preferences over another’s well-being, and as a result takes that other person’s interests into account in her own decision-making, or various conceptions of reciprocity, wherein altruistic behavior is strategically undertaken (consciously or not) to maximize long-term self-interest.¹

What ties together these various approaches to modeling ethical behavior is that they all assume that the utility-maximizing agent seeks to achieve the best outcome (as measured by her preferences) out of all the possible outcomes over which she has influence (as determined by her exogenously imposed constraints). If we translate this into the language of moral philosophy, this implies that the agent is a hedonist or egoist, if narrowly self-interested; an altruist, if she takes into account the welfare of select others (such as family or friends); or a broader utilitarian, if she takes into consideration the well-being of all persons (and perhaps other species as well). In all of these cases, the agent can be described as a consequentialist, since she determines the moral worth of actions according to the goodness of their outcomes. This is the implicit assumption made in most economic models of individual choice, which has been extended to study behavior in the context of the law and the family, for example, as well as more traditional market transactions.

But this pervasive reliance on consequentialism in economics comes at a price. Attempts to force a utility-maximizing explanation on “sacrificing” behavior (such as voting, tipping while traveling, or heroic acts) results in unsatisfactory, ad hoc assumptions of new preferences. Assuming that people behave ethically only out of expectation of future benefits, or even out of an unconscious, evolved disposition toward reciprocity, seems overly cynical, and cheapens the true ethical acts of persons, whose conscious, reflective, and deliberate behavior deserves to be explained in ways that emphasize its moral nature without cynically degrading it to self-interest.

Perhaps the most important contribution toward this end came from Amartya Sen, a Nobel laureate in economics and a prominent philosopher as well, who introduced the concept of commitment into the discussion of rational choice in his classic 1977 paper “Rational Fools.” By suggesting the possibility of commitment as an alternative method of motivating and
explaining choice, he opened the door for deontological values and principles to be incorporated into economics. Using Sen’s commitment framework, agents can act on principles, duties, or values without representing them as preferences which can be substituted for others as relative (subjective) costs dictate. Sen’s approach allows for absolute considerations in economic models of choice, influences which are not subject to standard economic factors such as prohibitively high opportunity cost or diminishing marginal benefit. Furthermore, as Sen noted, commitment “drives a wedge between personal choice and personal welfare,” since an agent’s choices cannot be assumed to maximize her own well-being if it is possible that they were motivated by some conviction or principle.2

Consistent with Sen’s theme, if economics strives to explain human behavior, it must recognize there are other ways for agents to behave other than according to consequentialist logic. As Vivian Walsh, another prominent economist and philosopher, writes:

It has always been thought that some actions were wrong despite their consequences. These claims must have a place in any serious moral theory, and economic theory and decision theory should not be allowed to foreclose what is properly an issue for moral philosophy, simply by adopting particular formal structures as constitutive of rational choice without explicit dialogue on the philosophical issues raised by doing so.3

In this chapter—and, more broadly, this entire book—I propose to construct an economic model of decision-making based on nonconsequentialist ethics, specifically the moral theory of Immanuel Kant, in which the nature of actions themselves, rather than their consequences, determines their moral worth.

This type of ethical theory is often called deontological, as opposed to consequentialist or teleological (goal-oriented); these terms are heavily debated among philosophers, but William Frankena’s definition is often taken to be representative:

Deontological theories . . . deny that the right, the obligatory, and the morally good are wholly, whether directly or indirectly, a function of what is nonmorally good or what promotes the greatest balance of good over evil for self, one’s society, or the world as a whole.4

Like most scholars, Frankena defines deontology negatively, in terms of
what does not determine the right thing to do—namely, the goodness of the outcomes of actions. Understood practically, this means that deontological ethics may allow, in some cases if not all, for persons to act in ways that do not maximize the goodness of consequences, and may even demand it. For instance, a deontologist may not allow the intentional killing of one innocent person to save two (or more) others; even though the number of lives lost would be smaller if the one innocent person were killed, and therefore the single killing would likely be recommended by a consequentialist, a deontologist may judge the act of killing an innocent person to be wrong regardless of the consequences. Positive definitions of deontology are harder to come by, mostly because every deontologist has a different idea of what makes certain actions right, but most agree that it is not a consequential measure.

Philosopher John Broome once wrote that “if deontological moralities affect people’s behaviour in important ways, then economics is in for a shock.” Well, among moral philosophers, Immanuel Kant is widely held to be the paradigmatic deontologist, and in this chapter, at risk of shock, I use his conception of duty to incorporate deontological considerations into the economic model of choice. I begin with a summary of key aspects of Kant’s moral theory, including autonomy, dignity, the categorical imperative, perfect and imperfect duty, and judgment. Next, I illustrate these ideas by applying them to the classic prisoners’ dilemma of game theory. Finally, I present a Kantian-economic model of decision-making, which shows one way in which Kant’s ethics can be incorporated into the economic optimization framework. My contention is that with the proper understanding of duties, preferences, and constraints, the standard economic model can describe deontological choice along Kantian lines; if I’m right, the shock should be a little less painful.

Kantian Ethics

Despite his high degree of name recognition, Kant is widely misunderstood as a moral philosopher. He is often accused of being cold, rigidly logical, and uninterested in the realities of human existence, and much of the blame for this must be laid at the feet of the great magister himself. Many people’s exposure to Kantian ethics starts and ends with his slim
Kantian Ethics, Economics, and Decision-Making

1785 work, the *Groundwork for the Metaphysics of Morals*, especially if it is taught as part of an introductory ethics course alongside the work of other moral philosophers. The more approachable sections of the *Groundwork* detail, in a relatively straightforward fashion, the three formulae of the categorical imperative in the context of four examples of immoral behavior (making false promises, committing suicide, failing to develop your talents, and neglecting the needs of others). But the treatment of autonomy, dignity, and the will in the *Groundwork* is much more abbreviated, incomplete, and difficult for students to grasp without lengthy elaboration from their teacher (or secondary readings), and topics like strength, virtue, and judgment are all but ignored. It is not until 1797’s *The Metaphysics of Morals* (especially the second half, known as the *Doctrine of Virtue*) that Kant explains the real-world implications of following the moral law. There, alongside lengthy discussions of the nature of virtue, vice, and other general ethical topics neglected in the *Groundwork*, he provides a systematic listing of duties to others and to ourselves, along with “casuistical questions” for the reader to ponder.

Furthermore, Kant himself regarded the more formalistic universalization formula of the categorical imperative as more applicable than the other two more humanistic versions, given its more technical, algorithmic nature: “One does better if in moral judgment he follows the rigorous method and takes as his basis the universal formula of the categorical imperative.” I regard this as unfortunate, because it is from the other formulae—especially the Formula of Respect for Humanity—that what I consider the “heart” of Kant’s ethics emerges, and it is with this heart, namely dignity and the autonomy from which it derives, that I begin. After explaining these more general aspects of Kantian ethics, only then do I turn to the categorical imperative, which operationalizes autonomy and dignity, and then to the nature of the duties that result from it and the role of judgment in moral decision-making. Duties will then be the main focus of the remainder of the chapter, which lays out a Kantian model of decision-making, and the next chapter will emphasize autonomy and the will at length, at which point the model of decision-making—or what I will call judgment—becomes a true model of choice.
Autonomy and dignity

Kant’s moral philosophy is ultimately based on autonomy (or inner freedom), the capacity of every rational agent to make choices according to laws that she sets for herself, without undue influence from either external pressures or internal desires. Thus considered, autonomy has equally important and interrelated negative and positive aspects. The agent is not bound to either internal desires or external authority, and therefore she is free to make choice according to her moral judgment. Nonetheless, she is bound by the laws she determines for herself (and which will rationally accord with the moral law), but since these laws are imposed on the agent by the agent herself, she is acting freely. If she sets herself the rule “I shall not lie,” then she is not limiting her freedom or autonomy—rather, this is the ultimate expression of her autonomy, because the rule is of her own making and imposition. To put it another way, we are free to restrict ourselves. “In one sense,” philosopher Christine Korsgaard writes, “to be autonomous . . . is to be governed by principles of our own causality, principles that are definitive of your will. In another, deeper, sense to be autonomous . . . is to choose the principles that are definitive of your will.”

Of course, the word autonomy has a number of meanings, in both the personal and political realms, and in common usage as well as philosophy. Perhaps the most familiar usage comes from international politics, in which national autonomy is roughly synonymous with sovereignty, which applies when no other power can legally compel a nation-state to action (or inaction). While the United States (for example) can be persuaded or given incentive to enter into a treaty with another nation, whether to do this is the choice of the appropriate government official or body in the U.S., not the other nation. By contrast, its fifty constitutive states have limited sovereignty or autonomy, given that their actions are limited by the U.S. Constitution; for example, the states cannot enter to treaties with foreign countries, and to some extent must follow the policies of the federal government. But the U.S. itself in not subject to other nations’ laws, so it is autonomous in this sense. Nation-states, particularly those with constitutional legal systems, are autonomous in the other sense as well, that of determining the laws or principles by which they will operate. Constitutions are laws that a nation sets for itself, to set procedural guidelines for its activities (such as setting up legislatures, executives, and
judiciary), as well as to limit those activities more substantively (such as the Bill of Rights in the U.S. Constitution).

Kant’s use of the word “autonomy” in relation to an individual agent is very similar to the meaning from international politics. An autonomous person does not allow external factors—especially outside authority—to determine her choices or actions. This is not to say, of course, that she cannot be influenced by external factors, or even decide to do what someone tells her; both happen all the time, and neither by itself implies any lack of autonomy. But she cannot cede her own decision-making authority to another, and if she chooses to follow someone’s direction or order, she must make a conscious, deliberate choice to do so. In this way, the ultimate decision is hers, not the other person’s, for she must endorse the external reason for action and thereby make it her own.

This aspect of autonomy, that of resisting external compulsion when making decisions, is familiar to most, and corresponds fairly closely to how we think of autonomy in personal situations (and in fields like medical ethics). While we take this aspect for granted most of the time, the other aspect, that of resisting one’s own preferences and desires, is more particular to Kantian ethics, and also more counterintuitive. Most of the time, following one’s desires and preferences—or inclinations, to use Kant’s term—is unproblematic, since there is no moral conflict involved. (Even so, one must make a conscious choice to follow them, rather than do it unthinkingly.) But in an ethical choice situation, Kant held that one’s preferences are not a reliable guide to proper decision-making; what we want to do, even if it may seem ethical, is not necessarily the right thing to do. When the moral choice differs from our preferred choice, autonomy grants us the power to deny our inclinations and do the right thing instead; in fact, autonomy implies the responsibility to follow the dictates of one’s own moral judgment. This is in clear opposition to Hume’s famed position that “reason is, and ought only to be the slave of the passions,” a view against which Kant aligned himself from the start.

While Kant held that all rational agents have the capacity for autonomous choice, exercising this capacity is not automatic, nor it is always easy. If an agent allows either her preferences or desires, or the wishes or commands of other persons, to influence her choice without adequate reflection and endorsement, she is said to have acted heteronomously: she has
allowed her will to be co-opted by a force other than her own judgment. But even in such cases, the agent has made a choice—she has chosen to sacrifice her autonomy or inner freedom. As Irwin puts it, “the difference between heteronomy and autonomy does not consist in the difference between compulsion and free acceptance, but in the source of the principles that we freely accept. We become evil not by being overcome by an evil principle, but by freely incorporating such a principle in our maxim.”

Korsgaard, drawing from Kant’s *Religion*, writes that “we learn that a bad person is not after all one who is pushed about, or caused to act, by his desires and inclinations. Instead, a bad person is one who is governed by what Kant calls the principle of self-love. The person who acts on the principle of self-love chooses to act as inclination prompts.” Finally, according to Thomas Hill, Kant held “that all have autonomy, that this implies commitment to certain rational constraints, and that some live up to these commitments while others do not.” In this way, autonomy is not just a property of rational beings, but also, in a normative sense, a goal: a person should always try to assert her judgment and her will without blind obedience to either her internal desires and preferences, or external authorities and influence. Only by doing so can she be true to herself, preserving her integrity and respecting her dignity.

Kant famously contrasted things and persons, the former having a price and the latter possessing a dignity above price: “whatever has a price can be replaced by something else as its equivalent . . . whatever is above all price, and therefore admits of no equivalent, has a dignity.” Things are instrumental, simply means to an end, and therefore their value is contingent on their usefulness; whereas persons, who are to be regarded as “ends in themselves,” possess an intrinsic worth which is incalculable and incomparable, resisting summation or substitutability, and which “admits of no equivalent.”

Kantian dignity is a relatively simple concept, and a very appealing one to the modern person (though shocking in his day). As Hill puts it, “the root idea of dignity is simply that virtually everyone, regardless of
social station, talents, accomplishments, or moral record, should be regarded with respect as a human being.” Despite its simplicity, the Kantian conception of dignity has very strong implications for how persons may be treated, by other persons as well as by the state (and themselves). This is of paramount importance to mainstream economics, where the concept of trade-offs is a central one, because Kantian dignity cannot be “exchanged,” since it has no price. Furthermore, one source of dignity—in other words, a person—cannot be traded off against another source of dignity, even if the latter seems to represent “more” dignity (such as two persons). As Hill writes,

this may seem to imply that there can never be a justification for impairing the rationality or sacrificing the life of any human being, but this is not necessarily so. What is implied, strictly, is only that one may not sacrifice something of dignity in exchange for something of greater value. Thus, if the sacrifice of something with dignity is ever justified, the ground for this cannot be “this is worth more than that” or “a greater quantity of value is produced by doing so.”

This may seem to be only a symbolic difference—a person’s dignity may be sacrificed, but not “in exchange”—and to some extent it is, but nonetheless it is an important one, for when we find we have to impose harm on someone undeserving of it, we want to ensure that it is done with regret and stark acknowledgment of what is being done, not the cold, smug satisfaction of efficient exchange.

As I said above, despite the formalistic gloss with which it is presented in the *Groundwork*, I maintain that the true heart of Kant’s ethics lies in his belief in the essential dignity and autonomy of rational beings. The categorical imperative and the duties that are derived from it, which I discuss next, ultimately can be traced back to dignity and autonomy, and thereby can be understood as simply operationalizing the respect that is owed every person due to his or her capacity for free choice. And as we shall see below, real-world ethical decision-making often cannot be conducted by straightforward recourse to duties or rules, but is rather a matter of judgment. Such judgment depends not on an encyclopedic knowledge of Kantian duties (wonderfully caricatured by Deirdre McCloskey as ethics by “pocket-sized, three-by-five inch card”), but rather on a more holistic appreciation of dignity and autonomy, and the responsibilities which they imply for persons in an ethical community.
The categorical imperative

Based on the autonomy of rational agents, and the equal respect it demands persons pay to each other (and themselves), Kant developed the categorical imperative, his version of the “moral law” that all rational agents freely impose on themselves. Kant maintained that the categorical imperative merely formalized the ethical decision-making of the common person: “The ordinary reason of mankind in its practical judgments agrees completely with [the categorical imperative], and always has in view the aforementioned principle. . . . To be sure, such reason does not think of this principle abstractly in its universal form, but does always have it actually in view and does use it as the standard of judgment.”

In Kant’s view, a person is to apply the categorical imperative—preferably the universalization formula—to her plans of actions or maxims; if a maxim “fails the test,” then that planned action is rejected as immoral, and if it passes, it is judged as permissible (not necessarily moral, and never demanded).

This belies an all-too-frequent characterization of Kant’s ethics, that it is excessively demanding and rigorous. As I will explain later in this chapter (and further in Chapter 3), Kantian ethics are demanding only in what one must not do, not in what one must do instead; there is much room for judgment in what the agent actually chooses to do in fulfillment of her duties.

Kant laid out three formulations of the categorical imperative, which he claimed are equivalent ways of stating the basic principle from difficult angles. The Formula of Autonomy or of Universal Law (henceforth the “Formula of Autonomy”) is the most commonly known version of the categorical imperative—and perhaps, as I mentioned above, the most misrepresentative. It is based on the universalization of maxims: “act only according to that maxim whereby you can at the same time will that it should become a universal law.” The standard illustration of this formula deals with lying: suppose I propose a maxim of lying to promote my interests. If I will that everyone may do the same, as the Formula of Autonomy demands, that would promote lying to such an extent that no one would believe anything anybody said, which would thereby defeat the goal of the lie (to benefit my interests). Despite its apparent reliance on logic and noncontradiction (as Kant explained it), the Formula of Autonomy is ultimately based on the equal dignity of all persons, and the implication
thereof that no one person should claim special allowances for herself that she cannot grant equally to everybody.\footnote{Understood this way, the problem with lying is not that it defeats my interests once universalized, which would imply an egoist ethic. Rather, in order for lying to work, I must be the only one (or one of very few) lying, which means I must not grant everyone else the same license I grant myself. Since I have no grounds on which to do this, given the equal dignity of all persons, I cannot will that a maxim of opportunistic lying be universalized.} Technically, there are two understandings of consistency in Kant’s ethics: consistency-in-conception and consistency-in-the-will. \textit{Consistency-in-conception} is the well-known test of logical consistency discussed above, which rules out lying because a universalized maxim of lying would destroy the trust on which successful lying depends. \textit{Consistency-in-the-will} stems from a variant of the Formula of Autonomy commonly known as the \textit{Formula of the Law of Nature}: “Act as if the maxim of your action were to become through your will a universal law of nature.”\footnote{The unique feature of this version is that it extends the teleological (or goal-oriented) basis of natural laws (including those based on human nature) to maxims; as Paton explains, \footnote{When we ask whether we could \textit{will} a proposed maxim as if it were to become thereby a law of nature, we are asking whether a will which aimed at a systematic harmony of purposes in human nature could consistently will this particular maxim as a law of human nature.\footnote{The word “will” is stressed above for a reason; the consistency-in-the-will test asks if the agent can \textit{rationally will} that her maxim be universalized, not simply whether such universalization is logically inconsistent.} This test generates duties such as beneficence by ruling out a maxim of indifference to others: since everyone will need aid at some time, no one can rationally will that everyone neglect the well-being of others. Rather than logical consistency, the consistency-in-the-will test demands that universalized maxims not contradict “objective ends, which depend on motives valid for every rational being,”\footnote{such as our humanity (based on dignity) and our own survival (based on the teleology of human nature), but not our everyday preferences and inclinations. Universal indifference endangers our own survival, so it fails the consistency-in-the-will test. As we shall see in our discussion of the prisoners’ dilemma game below, this}}
version of consistency is often misunderstood (when recognized at all). It introduces considerable vagueness into the universalization formula (since logical consistency is no longer enough), and also compromises its formalism (which is exaggerated anyway), with the effect of humanizing one of the aspects of Kant’s ethics which most often draws accusations of coldness and inhumanity, but at the cost of additional ambiguity.

Based on these first formulae, we can see exactly how Kant does, and how he does not, allow consequences, circumstances, and context to enter into moral considerations. It is true that specific, personal consequences of actions have no influence on their moral status, because such features cannot, by definition, be universalized without denying the equal dignity of all in order to carve out exceptions just for oneself; this is Kant’s version of the “impartial spectator” or “disinterested viewpoint” of Smith and Rawls, and it justifies universalization (not the other way around). Some take this too far, claiming that Kant pays absolutely no attention whatsoever to consequences or context, but this is a distortion, because some empirical knowledge of human behavior (and the consequences thereof) is necessary to derive the results of universalizing a maxim. For instance, to say that universal lying is self-defeating, we must know what a lie is, what its purpose is, how people react to it, how it affects the trust they have in communication, and so on. Also, context can be incorporated into a maxim; while a maxim of killing for advantage is forbidden, a maxim of killing in self-defense would not be. In Paton’s words, “if Kant had said merely that we must not allow our desires for particular consequences to determine our judgment of what our duty is, he would have avoided a great deal of misunderstanding.” The categorical imperative itself can be derived a priori, but the duties derived from them depend on empirical knowledge of human life: even in the *Groundwork*, Kant wrote that “all morals . . . require anthropology in order to be applied to humans.” As we shall see later in this chapter, judgment is essential to derive specific moral commands from the categorical imperative, even after general duties like “do not lie” are derived from it, as well as to settle conflicts among obligations, and empirical facts are indispensable in this judgment.

While, as we saw above, Kant recommended the Formula of Autonomy as the most easily applied version of his moral law, I believe it is the second version of the categorical imperative, the *Formula of Respect for the Dignity of Persons*, that better captures the heart of his moral philosophy:
“act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means.” As is evident from both the title and language, this formula is explicitly grounded in the essential dignity possessed by rational, autonomous persons, and is a much more humanistic formula than the Law of Autonomy (and more clearly so than the Formula of the Law of Nature). To continue with the example of lying, when a person lies, she is using the person to whom she lies, as well as the humanity in her own person, as a means to the end of this deception (whatever she stands to gain for it). She is using the other person as a means to her end because she is not treating him as an end-in-himself; she is not treating him with the respect he is owed as a rational, autonomous person. To look at it another way, the other person is not an equal participant in the situation, because he literally cannot agree to be lied to (a lie, by its nature, must be concealed). It is for this reason that coercion and deception are considered the two paradigmatic ways that one can be treated merely as means (while not at the same time as an end). Furthermore, the liar is also using herself as a means to her end, relying on her good name and the trust the other person has in her to further her ends; she thereby fails to respect the dignity in her own person, demeaning herself for the sake of momentary advantage.

A common misunderstanding regarding the Formula of Respect (for short) is that it prohibits using other persons under any circumstances. If taken this way, the categorical imperative would forbid all commerce, whether market-based or informal, in the form of contracts, promises, or favors—in short, any transaction in which one person makes use of another for any reason. But Kant’s wording is masterfully precise: one must never use persons simply as means, without at the same time treating them as ends. So we are free to use each other for our own ends, provided that in doing so we treat each other with respect, since we are all rational, autonomous agents possessed of dignity. More precisely, it must be possible (not necessarily likely) for any person with whom we deal to consent to the way in which we use them. On this understanding, coercion and deceit are the paradigmatic ways in which persons can be used merely as means: coercion obviously denies the person any opportunity to consent or dissent, and deceit implies that the person is not aware of the deceiver’s true intentions, and therefore does not have the chance to agree with them.
The Formula of Respect rules out these actions, but leaves enormous room for mutually voluntary and honest transactions.\(^4^4\)

Finally, the Formula of Respect is not merely a negative principle that commands us to refrain from using persons simply as means; it also instructs us to take other people’s ends as our own, and thereby generates duties such as beneficence. As Kant wrote, “It is not enough that he is not authorized to use either himself or others merely as means (since he could then still be indifferent to them); it is in itself his duty to make man as such his end.”\(^4^5\) As we will see below when we discuss the two types of duty, this type is different than that which prohibits using persons merely as a means, as it is much more general and “wide,” allowing of more latitude than the latter. Also, in our discussion of sociality in Chapter 3, the negative and positive conceptions of respect invoked in the Formula of Respect will help us frame the stages of human sociality. While the negative aspect of the formula is much better known, and is certainly an important first step to civil relations, the positive aspect is needed if human society is to flourish in any meaningful sense.

Our later discussion of sociality also invokes the third version of the categorical imperative, the *Formula of Legislation for a Moral Community*. While not used for judging maxims, the Formula of the Kingdom of Ends (as it is also known) is usually considered to be a combination of the first two, but with a unique emphasis on legislation and the teleological nature of Kant’s broader ethics: “every rational being must act as if he were through his maxim always a legislating member of the universal kingdom of ends.”\(^4^6\) This version reminds us that each person sets the moral law to and for herself, keeping in mind that that law must also be universalizable to all rational persons. It also makes clear the ultimate goal of ethical behavior: the attainment of the “kingdom of ends,” an ideal state of the world in which every person’s individual ends coexist in a harmonious, moral community.

Kant contended that the three formulae of the categorical imperative were equivalent: “the aforementioned three ways of representing the principle of morality are at bottom only so many formulas of the very same law: one of them by itself contains a combination of the other two.”\(^4^7\) However, he neglected to explain exactly why or how this is true (as occasionally was his wont). Certainly, the third formula is easily reconciled with the first two, as it
Kantian Ethics, Economics, and Decision-Making

can be seen as a combination of them, but the relationship between the first two may be more difficult to grasp. Such esteemed Kant scholars as Barbara Herman and Christine Korsgaard (just to name two) have questioned the consistency of the first two formulae in judging even basic moral concepts such as murder and lying, finding that the Formula of Autonomy is much more “flexible” than the Formula of Respect and even the former’s close cousin, the Formula of the Law of Nature.\(^{48}\) In large part, this is due to the fact that the Law of Autonomy (and its consistency-in-conception test) is much more sensitive to the way maxims are stated, which directly affects their logical consistency when universalized. For instance, Korsgaard attempts to reconcile the blanket prohibition on lying that stems from the Formula of Respect with the more elaborate constructions of maxims of lying that can “not so much pass as evade universalization.”\(^ {49}\) The more general duty not to lie seems more in line with the spirit of Kant’s ethics, where the possibility of cleverly crafted maxims that can slide by the consistency-in-conception test abuses the concept of equal dignity that is at the core of the Formula of Autonomy (and which should make it equivalent to the Formula of Respect).

Also, some maxims, even those that describe acts which are clearly immoral in commonsense terms, may not be logically inconsistent when everyone adopts them. Herman gives the example of the maxim “to kill whenever that is necessary to get what I want,” which I hope the reader will agree is an immoral plan of action.\(^ {50}\) If this maxim were universalized, the world would definitely be a worse place in which to live, but that is hardly inconsistent as a matter of pure logic; certainly such states of the world have existed through human history (and even today in some parts of the world). The Formula of the Law of Nature (which demands consistency-in-the-will) would judge this maxim immoral, since no one—not even a opportunistic murderer—could will a world in which her life is constantly imperiled by (other) opportunistic murderers. Herman flatly admits that “this result surprised me,” as she had earlier regarded the consistency-in-conception test as more demanding, and its prohibitions more important, than the consistency-in-the-will test.\(^ {51}\) But this result supports my contention that, even though the Formula of Autonomy is ultimately grounded in equal dignity, the more clearly humanistic versions of the categorical imperative better represent the heart of Kant’s ethics, even though they are less formulaic and deterministic in their judgments.
Duties perfect and imperfect

As we know from above, when a maxim is rejected by the categorical imperative, the result is a duty prohibiting such a plan of action. (If the maxim is not rejected, it is judged to be permissible but not required, and not necessarily moral in any affirmative sense.) For instance, a maxim of opportunistic lying is rejected by the categorical imperative, resulting in a duty not to lie. Similarly, a maxim of indifference to the suffering of others would be rejected, resulting in a duty not to be indifferent to suffering (often understood as a duty of beneficence).

These are standard examples of the two types of Kantian duties, perfect and imperfect. A perfect duty (also called narrow or due duty), such as the duty not to lie, is one “which permits no exception in the interest of inclination,” and is usually negative in nature, such as “do not kill” or “do not steal.” Generally, perfect duties are derived from the Formula of Autonomy (based on the logical inconsistency of universalization), and also from the negative portion of the Formula of Respect (which forbids using humanity merely as a means). An imperfect duty (also called a wide or meritorious duty) is one that the agent has some latitude in executing, both in degree and method: “the law cannot specify precisely in what way one is to act and how much one is to do.” Such duties are mostly represented in positive terms, such as the duties of beneficence and cultivation of one’s talents, but technically they are negative—do not be indifferent to others, do not neglect your natural abilities—because, like perfect duties, they too result from the rejection of maxims. Imperfect duties demand that we include certain ends in our decision-making processes, but they do not require any particular action in service to those ends (or inaction, as perfect duties normally do). Imperfect duties are most easily derived from the positive part of the Formula of Respect (treat humanity always as an end), and also the Formula of Law of Nature (requiring consistency with ends-in-themselves).

Another understanding of this distinction is that perfect duties are duties of action, while imperfect duties are duties of ends: “the distinction which Kant has in mind is that between a law commanding (or prohibiting) an action and a law prescribing the pursuit of an end.” Perfect duties are precise in their requirements regarding forbidden acts: do not lie, do
not kill, and so forth. Perfect duties, insofar as they are negative duties (as most are), constrain the agent from using certain means in pursuit of her inclinations. Imperfect duties are less precise, merely spelling out attitudes (such as beneficence) that should be adopted, with no specific instructions on how to express them in action. Kant leaves the rational agent some discretion regarding how heavily to weigh these dutiful ends against other duties, and even against one’s self-interested ends, and he suggests that they should be pursued only when doing so would not lead to excessive hardship or sacrifice on the part of the agent: “How far should one expend one’s resources in practicing beneficence? Surely not to the extent that he himself would finally come to need the beneficence of others.”

Though the substantive content of perfect and imperfect duties ensures that actions performed according to them are moral, Kant holds agents to an even higher standard to be moral themselves: they must perform dutiful acts not merely according to the moral law, but out of respect for it. This adds an essential motivational standard for Kantian agents that parallels classical virtue ethics (as described in the Introduction above): doing the right thing for the wrong reason does not make an agent moral or virtuous. A bad person can do a good thing—unintentionally or for the wrong reason—just as a good person can do something that turns out badly. This accords fairly closely with common intuitions about morality, where intentions are seen as more important (in terms of moral evaluation of character) than results or consequences. The common sayings “you tried to do the right thing” and “it’s the thought that counts” are two common reflections of this point of view.

This requirement is a reflection of Kant’s doctrine of the good will; as the first line of the *Groundwork* reads, “there is no possibility of thinking of anything at all in the world, or even out of it, which can be regarded as good without qualifications, except a *good will*. Many other human capacities are useful, or even admirable, but nonetheless they can be used for ignoble ends; a good will is necessary to ensure those capacities are used correctly. Kant believed that intentions are the only things that are completely under an agent’s control; once an action is initiated, the laws of the physical world take over, as well as possible intercessions from other persons. “A good will is good not because of what it effects or accomplishments, nor because of its fitness to attain some proposed end; it is good only
through its willing, i.e., it is good in itself.\textsuperscript{61} So ethically correct intentions formed out of respect for duty, even if they fail to culminate in the intended actions or results, still earn the person moral acclaim, but good actions done with impure intentions do not: “if with the greatest effort it should yet achieve nothing, and only the good will should remain . . . yet would it, like a jewel, still shine by its own light as something which has its full value in itself. Its usefulness or fruitlessness can neither augment not diminish this value.”\textsuperscript{62}

Even moral sentiments cannot ensure the good will, as Kant explained in a particularly notorious passage:

There are many persons who are so sympathetically constituted that, without any further motive of vanity or self-interest, they find an inner pleasure in spreading joy around them. . . . In such a case an action of this kind, however dutiful and amiable it may be, has nevertheless no true moral worth.\textsuperscript{63}

In another part of the \textit{Groundwork}, Kant even says that although we have a duty to preserve our own lives (and not commit suicide), most of us do so not out of duty, but rather out of an obvious inclination (to live!) which only happens to be in accordance with duty.\textsuperscript{64} Only those who refrain from committing suicide because they believe it to be wrong are truly moral. Of course, this does not imply that following one’s inclination to live is immoral, but merely amoral or not deserving of merit or esteem. But, on the other hand, we would safely say that a person whose only desire was to end her life, but nonetheless did not, was moral, since her respect for the moral law outweighed her (presumably very strong) inclination to die.\textsuperscript{65}

This is a rare case where we may be able to infer the true motivations behind an action: when it is clearly in opposition to a person’s inclination. But most of the time, Kant recognized, we very rarely know the true motivation of our actions—whether they were performed out of duty, inclination, or both—much less anybody else’s. If someone sacrifices something of value to herself to act dutifully—say, running into a burning building to save a stranger’s child—then it is fairly safe to assume that action was done out of duty rather than out of an inclinational motive. However, if the person were hoping for adulation and a six-figure book deal, that would change our assessment of his motivation and character—as I think it would for most people, not just Kantians. It seems much more likely that
we act with mixed motivations in most cases: partly to do the right thing, but also because we want to help others, be honest, and so forth. Kant’s position, partly in response to sentimentalists like Hume and Smith, was that persons should not behave morally because it feels good, but rather it should feel good that they behave morally. In other words, the moral assessment of the act comes first, and the “warm glow” or satisfaction (as long as it does not lapse into vanity) derives from that assessment.  

Judgment

It is unfortunate that, with regard to his ethical system, Kant developed the reputation of being demanding, uncompromising, rule-obsessed, and insensitive to context and circumstances. Once again, if one reads only the *Groundwork*, it is clear why this impression is so pervasive; indeed, even a wider reading of Kant’s ethical writings tends to reinforce this. But in truth, none of these characterizations is deserved, because the categorical imperative and the duties that result from it provide merely a basic framework for ethical deliberation in real-world contexts, with significant room for flexibility to accommodate the context of individual decision-making situations. As Onora O’Neill explains:

Discussions of judgment . . . are ubiquitous in Kant’s writings. He never assumes agents can move from principles of duty, or from other principles of action, to selecting a highly specific act in particular circumstances without any process of judgment. He is as firm as any devotee of Aristotelian phronesis in maintaining that principles of action are not algorithms and do not entail their own applications.  

A related critique is that Kant’s moral theory is excessively focused on rules, but Kant opposed rule-worship as the enemy of true freedom and autonomy: “Dogmas and formulas, those mechanical instruments for rational use (or rather misuse) of [man’s] natural endowments, are the ball and chain of his permanent immaturity.” As O’Neill writes, in her response to virtue ethicist Alasdair MacIntyre’s criticisms of Kant along these lines, “Kant provides us primarily an ethic of virtue rather than of rules,” in part because “Kant offers us a form of rationalism in ethics that . . . does not generate a unique moral code, but still provides fundamental guidelines and suggests the types of reasoning by which we might see how to introduce these guidelines into the lives we lead.” At most, the categorical impera-
Kantian Ethics, Economics, and Decision-Making 33
tive, and the duties it generates, are intended to provide rough guidelines for moral intention and action, a sort of moral “compass”\(^70\) that we rely on to help us through the trickier ethical corridors of life.

Despite the apparently rule-oriented nature of the *Groundwork*, Kant had tremendous respect for judgment: “though understanding is capable of being instructed ... judgment is a peculiar talent which can be practiced only, and cannot be taught. It is the specific quality of so-called mother-wit; and its lack no school can make good.”\(^71\) Sullivan emphasizes the never-ending development and growth of our judgment:

Through simply living, facing ordinary moral problems day by day, we all accumulate a store of moral experience to help us judge how to act; we all develop some sensitivity to the features to which we should attend. Moreover, most of the situations in which we find ourselves are familiar ones, and we do not need to deliberate over how to act. We simply act on maxims that reflect our long-standing commitments and values.\(^72\)

Kant refuses to elaborate on judgment in terms of higher-order principles or rules, arguing that doing so would lead to infinite regress: since no rule, even higher-order rules, will be determinate in all situations, one would need yet higher rules to show how apply them, and then even higher-order rules for those, and so on.\(^73\)

Judgment proves essential to moral decision-making in several ways. For one, judgment is necessary to apply the general guidelines provided by the categorical imperative to day-to-day decisions. Of supposed “universal decision procedures” that are imputed to Kantian ethics, Allen Wood writes that “human life and moral deliberation are too complex for any such procedure ever to exist.”\(^74\) Robert Louden argues that moral anthropology is need to “strengthen agent’s powers of judgment,” which it does by “organizing and presenting relevant aspects of human experience to agents to reflect on under controlled circumstances.”\(^75\) Barbara Herman refers to this function of judgment as providing “rules of moral salience,” which allow the agent “to pick out these elements of his circumstances or of his proposed actions that require moral attention.”\(^76\) “Then, because the categorical imperative “is not itself a moral rule—it is an abstract formal principle,”\(^77\) judgment is required to understand exactly how to construct a maxim, itself “a principle that expresses a complex volitional judgment,”\(^78\) and also how to determine its validity vis-à-vis that formal
principle, as it is rarely obvious (even, as we saw above, in the supposedly straightforward and logical contradiction-in-conception test). And after the relevant moral aspects of a situation have been determined, judgment is necessary to determine how to act dutifully in accordance with moral laws—especially with regard to imperfect duties—for “to be sure, these laws require . . . a power of judgment sharpened by experience, partly in order to distinguish in what cases they are applicable, and partly to gain for them access to the human will as well as influence for putting them into practice.”

Perhaps most important for practical purposes, judgment is essential for solving the problem of conflicting obligations or rules. Obligations often conflict through no fault of the conflicted person: she may be called suddenly to help a family member at the same time she has promised to meet a friend. Kant believed that a person only had one operative duty at any given time, so one obligation must be pronounced to be of “stronger ground of obligation” than the other. As he sometimes did, Kant elaborated little on what exactly he meant by this, but rather than reflecting oversight on his part, this may have been intentional. He may have meant to imply that the agent must use her judgment to decide which obligation is more important, based on her knowledge of, and respect for, the moral law, as opposed to deterministically following an established set of rules. It is entirely possible that two people, both dedicated Kantians facing identical circumstances, would make two different judgments regarding the best action in that particular case, because each person’s judgment, based on previous experiences and choices, is unique to her. In other words, each person will make the choice, based ultimately on the moral law, that preserves the integrity of her character.

While, as in the case with judgment in general, Kant provided no definite rule or algorithm for comparing the “pull” of conflicting obligations, the outcome in some cases may seem clear: when a perfect (narrow) duty is at odds with an imperfect (wide) one, the latter would seem to give way, since it is flexible in its execution whereas the perfect duty is not. But recall that all duties are essentially negative, so “do not x” is roughly comparable with “do not y,” even if the first command generates a strict obligation (such as “do not lie”) and the second a wide one (such as “do not be indifferent to others”). Consider this example: Alicia makes a
promise to Bill to drive him to a blind date, but while en route to pick him up, Alicia sees Carl by the roadside, in urgent need of assistance because his wife Debbie is in labor. The duty to keep promises is a narrow one while the duty to offer assistance is a wide one, but favoring the former does not seem right in this case: most would agree that Carl and Debbie are in much direr straits than Bill (no matter long it has been since he last had a date). It is not hard to imagine that, whatever “stronger ground” may mean, it could certainly mean the more important obligation, where “importance” may be defined in terms of consequences (helping deliver a baby will likely do more good than furthering Bill’s love life), social proximity (saving a friend rather than a stranger when both are in danger), and so on.  

Of course, none of these empirical factors enter into the determination of a duty itself, but since there is no rule to pick the more important obligation, such aspects of the situation, filtered through the faculty of judgment, may be necessary to break the logjam. We can go even further and recognize that conflicts among obligations are very common in real life, in which case even perfect duties are often not as rigid and demanding as they appear in isolation. There may be many obligations facing us in any given choice situation, and the responsibility for balancing their pull, or comparing the strength of their grounds, falls to judgment. In a discussion of moral dilemmas, Barbara Herman cites “the hard work of moral deliberation that is central to a moral life: the engagement with multiple moral considerations present in an agent’s current or anticipated circumstances of action.” This is the sense in which the categorical imperative and the various duties that come from it are merely guidelines for ethical behavior, reasons for action that we must take into account when facing moral dilemmas, but ultimately must be weighed and balanced by our judgment. If Kantian ethics were truly a strict rule-based system, none of this would be necessary.

The Prisoners’ Dilemma

At this point, it may be useful to provide an example applying much of what was presented above, so we will now unleash Kant’s ethics on the most famous and widely discussed situation in game theory, the prisoners’ dilemma game. The essence of the prisoners’ dilemma game is that both
players have dominant strategies (actions that earn them the highest pay-off regardless of the other player’s action), but when they both choose their dominant strategies (as it is assumed they will), they both end up with lower payoffs than if they had both chosen the dominated strategy instead. The game derives its name from a tale of two criminal suspects (or “prisoners”) being interrogated in separate rooms; if they both kept quiet about their recent activities, they would get a small jail term, but each has the opportunity to sell the other out (and receive no jail time) if he or she confesses (and implicates the other as the mastermind), whether or not the other is likely to confess, which is therefore each suspect’s dominant strategy. In the end, both suspects confess, earning them both longer jail terms than if they had both kept quiet. The prisoners’ dilemma stands as the prototypical example of conflict between individual and collective rationality: through mutual pursuit of self-interest, all players end up worse off than if they had behaved otherwise.

It has been suggested that ethical behavior, as opposed to self-interested decision-making, would help solve prisoners’ dilemma problems. Most prominently, Amartya Sen, referencing the classic interpretation of the game, has written that “it is indeed easy to see that it will be difficult to find a moral argument in favor of confession by the prisoners,” implying that any ethical system (other than ethical egoism) would demand cooperative behavior in prisoners’ dilemma situations, including Cournot oligopolistic competition, private contributions to finance public goods, and arms races, to mention just a few. As an example of such an ethical system, many scholars (including Sen) have focused on Kant, arguing that the application of Kantian ethics would require that the players in such games cooperate rather than deviate, and therefore reach the Pareto-superior outcome. But I will argue against this, based on the description of Kantian ethics given above.

To begin, we can simply follow Kant’s own advice and apply the Formula of Autonomy to the players’ behavior in the game. I will use the term “deviation” to refer to the dominant strategy in a prisoners’ dilemma game (which leads to the suboptimal outcome), and “cooperation” for the dominated strategy (which leads to the optimal outcome). Can a maxim of deviation be universalized without inconsistency? First, we will apply the consistency-in-conception test: is one player’s deviation logically
inconsistent with the other player doing the same? Certainly both players can, and often do, deviate simultaneously—that is the crux of the problem, after all—so there is nothing logically inconsistent there. (Of course, this behavior results in inferior payoffs, but that does not affect the logical consistency of mutual deviation.) But, as we saw above, consistency-in-conception does not rule out occasional murder either, so we turn to consistency-in-the-will, which asks if the agent can rationally will that her maxim of deviation be universalized. Certainly, universal indifference to suffering cannot be willed rationally, but unless serious harm is threatened, it is difficult to see how universal deviation is inconsistent with any ends-in-themselves. So if we rule out such extreme circumstances, neither understanding of consistency derived from this formula of the categorical imperative prevents deviation in the prisoners’ dilemma game.

Sen comes close to the contradiction-in-the-will test when he writes that “certainly neither prisoner would like that confessing becomes a universal practice, and the only universal law that each prisoner would like is that everyone should refuse to confess.” But here he conflates “willing” with “liking” or “wanting,” introducing inclination into the universalization procedure. A similar misunderstanding of the universalization formula of the categorical imperative has been repeated often in the literature on altruism and public good financing. For instance, Bilodeau and Gravel understand Kantian ethics to mean “compelling an individual to undertake any action which we would want everyone else to undertake,” and cite many papers that share this conception, including Laffont’s “Macroeconomic Constraints, Economic Efficiency and Ethics,” one of the earliest attempts to integrate Kantian ethics explicitly into economics. But Bilodeau and Gravel recognize that this understanding of Kant’s moral theory may not be accurate; in fact, they point to other papers that emphasize the frequent misuse of Kantian ethics in economics (an august grouping to which your author would add this humble monograph).

Perhaps we should look at the nature of duties themselves for a different perspective on the problem. Is there a perfect duty not to deviate in prisoners’ dilemma games? We saw above that such action cannot be ruled out by the consistency-in-conception test, which normally results in perfect duties, so “do not deviate” is likely not one. But what about imperfect duty: could a duty not to deviate in prisoners’ dilemma games be a specific...
instance of the duty of beneficence? Not deviating will increase the other player’s payoff regardless of her action, so it would certainly be considered helpful, and may even be considered a virtuous act. But at the same time it comes at a sacrifice to yourself (since you would be choosing the dominated strategy), and you are not required to make any given sacrifice to perform a wide duty. So while cooperation in a prisoners’ dilemma game may be nice, kind, admirable, or noble, it is not required; even a concern for the other player based on treating her as an end (as the Formula of Respect dictates) does not mandate that one sacrifice his own well-being by avoiding the self-interested dominant strategy of deviation. This highlights the essential problem with relying on Kantian ethics to prevent prisoners’ dilemma outcomes: while the categorical imperative rules out many actions as immoral, it does not specify any precise actions that must be taken otherwise. Since imperfect duties do not demand specific action, we cannot thereby derive a strict moral obligation to cooperate in prisoners’ dilemma games based on them.

But if we expand the strategies available to the players, there may be a way to imply a perfect duty to cooperate, specifically by allowing both players to promise cooperative behavior before the game is played. Kant uses promise-keeping as one of his four examples of using the categorical imperative, and the argument based on consistency-in-conception is straightforward: as with lying, if promises are not kept, trust will be compromised and promises will lose their efficacy. So if the players make promises to each other to cooperate, then they are bound by their duty to keep these promises, thereby avoiding the prisoners’ dilemma outcome. However, there is no duty to make such a promise in the first place (because such a duty would be based on a perfect duty to cooperate, which as we have seen does not exist), so there is still no perfect duty that requires the players to choose the cooperative action in prisoners’ dilemma games.

Even though we have ruled out a perfect duty to make cooperative promises, such a practice can nonetheless provide a way for players to avoid the prisoners’ dilemma outcome: if each player promised to cooperate, contingent on the other player making the same promise, then both players would be bound by their promises and would therefore cooperate. But such a promise to cooperate is effective in eliciting similar behavior from other players only insofar as the player making the promise is known to be a
Kantian—after all, anyone can promise to cooperate, for doing so is “cheap talk” in a single-play game (and may be generated by purely self-interested behavior in an infinitely repeated game). For these reasons, the Kantian duty of promise-keeping cannot be relied upon to ensure cooperative behavior in prisoners’ dilemma games, but it can provide a method for Kantian agents to resolve prisoners’ dilemma games in their own self-interest.

Maybe we need a more general, humanistic perspective, one given by the Formula of Respect. Is there any sense in which deviation in a prisoners’ dilemma game treats the other players simply as a means while not at the same time as an end? It is true that one player’s deviation always raises his payoff and lowers the other player’s payoff (regardless of the other player’s action), but does that imply that the first player used the second simply as a means to his benefit? If it did, it would imply that one firm uses its competitor when it lowers its price to gain market share, increasing its own profit “at the expense” of the other firm’s profit; or that a firm in an imperfectly competitive industry uses its customers due to its ability to raise price above marginal cost. Would these actions be forbidden by Kant’s injunction to respect the dignity of persons (assuming we can extend this concept to firms as well)?

While these actions may fall into some readers’ personal conception of using someone, Kant had a much more precise notion of what it means to use someone merely as a means and not at the same time as an end. To respect another person’s dignity, Kant wrote that you must treat her as an autonomous, rational being; specifically, as we saw above, she must be able to—though not necessarily want to—rationally assent to your ends. For instance, in the context of the duty not to make false promises, Kant described what it means to treat someone otherwise:

[T]he man whom I want to use for my own purposes by such a promise cannot possibly concur with my way of acting toward him and hence cannot himself hold the end of this action. . . . [A] transgressor of the rights of men intends to make use of the persons of others merely as a means, without taking into consideration that, as rational beings, they should always be esteemed at the same time as ends, i.e., be esteemed only as beings who must themselves be able to hold the very same action as an end.95

Note the emphasis on possibility and ability, not preference or desire: the other person does not have to actually share my end or agree with it, and
can in fact abhor it. Kant requires only that she be able to rationally assent to it, even if she chooses not to, for it is the fact that she can choose to, not whether she would, that is essential.

For this reason, as I mentioned above, coercion and deception are the primary examples of using someone simply as a means, because they both render her incapable of making a choice regarding the other person’s end. Coercion obviously rules out choice altogether, and deception keeps the other person unaware of my true end, denying her the ability to assent to it. But if she can make a choice without coercion or deceit on my part, then I have treated her as a rational, autonomous person, and have considered her as an end as well as a means. So the firm that, without engaging in coercive or fraudulent behavior, lowers prices to steal business from a competitor treats that competitor with respect (even while lowering its profit). Asymmetry of power, as in the imperfect competition case, does not imply violation of respect, either; though she may be resentful of the higher price charged, the customer has the choice whether or not to participate in a transaction with the company and recognize its end of profit-maximization, as long as it has not acted fraudulently. The same goes for the prisoners’ dilemma game: each player’s end is assumed to be his or her own self-interest, a goal to which the other player can certainly assent (even if the players have no regard for each other, or if they despise each other). Since the standard definition of the prisoners’ dilemma game rules out coercion (through free choice of actions) and deception (through common assumptions on information), it is implied that the players are treating each other with the respect due to rational, autonomous persons.

So while Kantian ethics may encourage some degree of cooperation in a prisoners’ dilemma game, it does not require it. The categorical imperative permits players in prisoners’ dilemma games to choose their dominant strategies of deviation despite the resulting suboptimal payoffs for both. Because the categorical imperative does not demand specific actions, but only forbids certain actions (in the case of perfect duty) or requires the adoption of general attitudes (in the case of imperfect duty), it does not support a strict duty of beneficence toward the other player. Furthermore, deviation does not involve treating the other player simply as a means and not at the same time as an end, because it does not impair her autonomy or
choice through coercion or deceit. At bottom, Kantian ethics is not strict enough to guarantee cooperation in prisoners’ dilemma games, supporting the broader point that Kantian ethics provides a framework for moral judgment rather than definite, ready-to-apply rules.

Kantian-Economic Model of Decision-Making

Having discussed Kant’s moral philosophy, we are now ready to incorporate these ideas—particularly his distinction between perfect and imperfect duties—into the standard economic model of decision-making. This model assumes that the agent has fixed, stable preferences, which she maximally satisfies within given constraints and information (or beliefs). If the agent’s preferences are complete and transitive, they can be represented by an ordinal utility function, which assigns higher levels of utility to more preferred options. It is important to note that in this context, the word “utility” does not refer to a mental state such as pleasure or happiness, but simply a numerical index or ranking of options. Accordingly, an agent’s preference ranking is not assumed to be based on any psychological sense of desiring one option over another; instead, the ranking can be based on another person’s well-being, a faith-based way of life (reflected in observance of dietary restrictions, for instance), or any other source, whether or not it is based on self-interest (however widely or narrowly understood). By denying a mental-state conception of utility, economists are freed from having to place substantive restrictions on preferences, allowing agents to be self-interested or altruistic, narrowly hedonist or broadly utilitarian, all-loving or misanthropic. The constraints in the standard model of choice are traditionally given less attention, but are also assumed to be exogenous, and of a financial, physical, or temporal nature.

For the purposes of this chapter, we will accept the broad terms of the constrained preference-satisfaction model of economic choice, and go on to explain how this model can incorporate Kantian ethics as summarized above. We will approach this in two steps. First, we will assume that the agent has determined her duties as they apply in a given decision-making context, and sees how those duties fit into the model. Once we have done that, we will explore the details regarding the determination of her “best” action, which will complete the Kantian-economic model of
decision-making. Both steps necessarily involve judgment, emphasizing the point that this is the central faculty involved in modeling and practicing Kantian ethics.

Incorporating duties into preferences and constraints

The most apparent difficulty with our task is that there seems to be no room in the standard economic theory of choice for duties, especially duties that, by definition, cannot be performed because an agent wants to (for any reason, selfish or altruistic). But recall that the term “preference” in modern economic theory does not imply any basis in a mental state such as happiness and pleasure, but merely an ordering or ranking of states of the world over which the agent has some influence. So, in theory, preferences can also be based on Kantian duty as derived from the categorical imperative; the agent could rank states of the world in which she performed her duty higher than ones in which she did not, without implying any “desire” for the former over the latter. Furthermore, if the agent can rank some duties higher than other duties as well as her “normal” preferences, and can do so completely and transitively, then these duty-based “preferences” can be included in an ordinal utility function alongside any others. The resulting utility function can then be maximized within the agent’s constraints, involving marginal adjustments in response to changes in prices and income, just as in standard choice theory.

But can all duties be represented in terms of preferences in this way? This is where the distinction between perfect and imperfect duties becomes crucial. Plainly, perfect duties cannot be included among preferences, since they take precedence over inclinations and cannot be traded off amongst them if performing them becomes too expensive. Rather, perfect duties constrain the pursuit of inclinations, spelling out the means we may not use to achieve our ends. This is most clearly seen in terms of the Formula of Respect, which forbids a person from doing anything that uses other persons (or herself) merely as means while not at the same time as ends. With this understanding, we can model perfect duties as constraints, in the same way that budget constraints limit a consumer’s spending. A Kantian agent is free to pursue her own ends (including the
ends imposed by imperfect duties, to be discussed next), as long as she acts within the constraints given by perfect duty: do not lie, do not steal, do not kill, and so forth.

Certainly, explicit consideration of moral constraints such as these is not the norm in economics, but economists often implicitly assume these constraints in the background. In standard models of trade and commerce, for instance, economists normally assume that buyers and sellers make voluntary transactions, free from fraud and deceit, and with no threat of theft from either side. While such factors have certainly been added—in the literatures on corruption and crime, for instance—they are represented as aberrations, modifications of the basic modeling structure which assumes such exceptions away. Since the agents in these standard models are not free to resort to such immoral means to achieve their ends, they are behaving as if they were Kantian agents observing their perfect duties. Ironically, the nature of perfect duty seems to be the tallest hurdle to overcome when integrating Kantians ethics into the economic model of choice, but in a way, it has been there all along.

Another way to model “unusual” constraints on decision-making is to use lexicographic preferences, preferences that must be satisfied before others can be considered (similar to the stepwise process of alphabetical ordering). In other words, an agent could have a lexicographically superior preference for not-lying, such that she had to satisfy that “preference” before acting to satisfy any others. But this technique is descriptively inaccurate and practically unwieldy. It is inaccurate because it represents actions we feel are wrong as inactions which we prefer to “take.” But it is not that we rank “not lying” as a higher option than all the rest; it is something we simply will do or not do in trying to satisfy our normal preferences—a constraint. It is unwieldy because it violates the spirit of the preference-satisfaction model: to model trade-offs among preferences according to prices and income. Lexicographic preferences, by definition, cannot be traded off, and are unresponsive to prices or income—which is true of perfect duties and constraints, but not of preferences as usefully understood.102 So why model such aspects of choice as any sort of preference when they do not function as preferences, especially when we have the concept of constraints conveniently available? While perhaps an interesting theoretical construct,
lexicographic preferences misrepresent the purpose and role of moral duties and principles in decision-making, which is rather to constrain preference-satisfaction for ethical reasons.

A more moderate suggestion is to model strict moral duties and principles as preferences which can be traded off with other preferences as prices and income dictate. For instance, Gintis writes that “one might be tempted to model honesty and other character virtues as self-constituted constraints on one’s set of available actions in a game, but a more fruitful approach is to include the state of being virtuous in a certain way as an argument in one’s preference function, to be traded off against other valuable objects of desire and personal goals.” While this maintains a preference ordering allowing trade-offs, which is methodologically more workable (theoretically and empirically), there is still the problem of representing something we feel we should not do as something we want to avoid doing—this may be part of the story, but not the entire tale.

Consider a weary traveler who has arrived home after a long business trip, and the only thing on her mind, the only desire or inclination she cares about, is to see her beloved fiancé (or dog—your choice). She spots a cab, rushes towards it, but a feeble elderly man gets there just before she does. She could easily push the man aside to get the cab and satisfy the only preference she cares to act on, but does not; she recognizes one of the most basic rules of civilized behavior—“wait your turn”—and catches the next cab. Which explanation sounds more natural: that she indulges a stronger preference to not push aside senior citizens—or, at least, that particular senior citizen—to the detriment of her preference to see her fiancé a little sooner, or that she recognizes a duty not to push aside senior citizens (even if she may not particularly like them) that constrains her actions taken to see her fiancé? Again, framing moral duties or principles which tell us what not to do as preferences over what we want not to do (I dare you to say that three times quickly) obscures what is going on in such situations.

But what if persons do not regard their duties or principles as absolutes, and do in fact trade them off in response to opportunity costs? Citing common sense and experimental evidence, Gintis writes that “if the cost of virtue is sufficiently high, and the probability of detection of a breach is sufficiently small, many individuals will behave dishonestly.” The curt
response is that such persons do not recognize moral duties or principles at all, reminding one of the old joke in which a man asks a woman, “Would you sleep with me for a million dollars?” When the woman says yes, the man then asks, “Would you sleep with me for five dollars?” The woman indignantly replies, “Absolutely not—what kind of woman do you take me for?” The man answers, “We’ve already established that—now we’re just bargaining over the price.” It does not take a Kantian to see that duties or principles which are subject to compromise for mere goods or money (rather than other duties or principles) are no duties or principles at all. And if we rule true, priceless duties and principles out, we are left with persons whose virtue can be bought, which is not an attractive picture of morality (or politics, for that matter). But, of course, persons do compromise their duties or principles for self-interested reasons, but I do not think that such behavior should be treated so casually (especially in the context of modeling moral choice!), but rather as the aberration it is (or should be). (Accordingly, we will discuss it in such terms in the next chapter.) Let us not compromise our agents’ morals before we even begin to understand them.106

Some may agree with the interpretation of duties as moral constraints, but object that such factors are “merely” normative constraints, and are less binding in a physical or logical sense than the typical budget or technological constraints in economics. But, as Goldfarb and Griffith recognize, even those constraints may not be binding unless they are reinforced by normative constraints; for instance, a consumer’s budget constraint is less binding if he is open to shoplifting, and a producer’s technological constraints can be loosened by compromising workplace safety or the quality of the final product.107 But nonetheless, it is true that moral constraints, especially if self-imposed and enforced as the result of autonomous processes, can be violated just as voluntarily; we will address this in the next chapter when we discuss weakness of will and character and their role in ensuring dutiful action in the face of human weakness. But to rule out self-imposed constraints altogether is to deny an essential component of autonomy and self-determination, which must include not only what one wishes to do, but also the lengths to which one will not go to achieve one’s goals.

While perfect duties must be modeled as constraints, imperfect du-
ties can easily fit alongside other preferences in an overall ranking. Unlike perfect duties, imperfect duties do not demand specific performance of action (or inaction), but instead only mandate general ends that should be adopted, allowing latitude with respect to inclination and other duties. So an agent can order the ends imposed by imperfect duty, such as beneficence, among her inclination-based preferences in her overall ranking, since the modern economic sense of “preference” does not necessarily imply true desire in the psychological sense. Furthermore, Kant was clear that the agent’s own happiness and well-being can and should be taken into account in her decision-making, if only because she herself, as a person with dignity, is to be treated as an end along with all other persons: “lawgiving reason . . . permits you to be benevolent to yourself on the condition of your being benevolent to every other as well.” Seeing this, we can model the economic agent as following the categorical imperative by incorporating perfect and imperfect duties into her choice framework as constraints and preferences, respectively. Once the imperfect duties have been “chosen” by judgment in accordance with, and out of respect for, the agent’s duty, an ordinal utility function can be constructed from them, which she will maximize as usual. Therefore, Kantian economic agents differ from consequentialist ones not so much in how they optimize, but amongst what they optimize (rankings of imperfect duties and inclinations), against what they optimize (perfect duties and exogenous constraints), and how they choose what to optimize (autonomous judgment).

Deliberating on action

There are several problems with the simplistic model just described. First, we have not explained how the agent orders her imperfect duties against her inclinations, much less against her other imperfect duties. Second, we need to confront the possibility of a conflict between perfect duties, in which one constraint may cancel out another. Finally, there may be conflict between perfect and imperfect duties that cannot be resolved by simply compromising the performance of the imperfect duty (such as being less helpful if a lie were required to be more helpful). As we saw earlier, perfect duty cannot be assumed always to take precedence over imperfect duty, for this ranking depends on the strength
of their respective grounds of obligation, not simply on their relative strictness.

As we saw above, Kant gave no definite answers regarding these matters; the agent must use her judgment to resolve these issues, and in this way judgment plays an integral role in crafting an agent’s overall “objective function.” Once judgment has determined the operative duties, as well as their relative importance, in any given choice situation, the agent can then maximize her objective function as usual (pending the issues of will and character to be discussed in the next chapter). Furthermore, judgment is not reducible to an algorithm; as we said above, the agent must make choices consistent with her understanding of the moral law, based ultimately on the inherent dignity and equality of every person, to maintain the integrity of her character.

Nonetheless, we can expand on how this process of judgment might play out in each of the following circumstances:

1. **Ranking imperfect duties against inclination and other imperfect duties.** Imperfect duties, even if they generally fit nicely among inclinations in the agent’s preference ranking, must nonetheless be ordered somehow amongst them, as well as against each other. While there is no one “right” ordering that every agent must follow—as there would be for the simple case of perfect duty over inclination—we can suppose that there is such a “right” ordering for any given agent in any given choice situation. But from where does this “right” answer come? Again, it comes from the agent’s judgment, which can be understood as her personal interpretation of the moral law above and beyond the formality of the categorical imperative, stemming from an appreciation of the essential dignity and autonomy of human beings (from which the categorical imperative is derived).

In his theory of judicial decision-making, Ronald Dworkin argues that there is a “right answer” to any legal dispute, as determined by the principles inherent in the legal system and rights implied by them, and it is the judge’s responsibility to find it and declare it in her opinion. But this is not an ontological claim to legal certainty, for two judges may each think she has the “right answer” though they differ on what that is. Dworkin’s point is that a judge must be confident that she has taken all of the relevant factors into account, weighed them properly against each other, and found the answer she believes to be “right.” Accordingly, the
Kantian agent follows the same process, weighing all the facts of the situation and her moral obligations to arrive at the answer she feels is right—in other words, consistent with her understanding of the moral law and the concept of dignity upon which it is based. Of course, two agents may come to different decisions when facing the same dilemma, but each one must feel comfortable that he or she has made the best judgment possible, and should stand ready to defend that decision with reason.

2. Ranking perfect duties against each other. Understood as constraints, perfect duties pose less of a problem for our model of decision-making, simply prohibiting certain actions we might otherwise take to further our ends. The moral agent does not lie to gain advantage, does not cheat to win a game, and does not kill to eliminate romantic competition. In a sense, perfect duties shrink the “action space” available to a moral agent: they foreclose all the actions that involve deception, coercion, and so forth. Perfect duties always take precedence over mere inclination, perhaps the only clear ranking in this picture of Kantian decision-making. But what if several perfect obligations conflict with each other? To continue with a visual analogy, what if all actions (including inaction) are foreclosed by one duty or another, but a course of action must be chosen all the same? This is what philosophers call a \textit{tragic dilemma}, a choice from which the agent cannot emerge with “clean hands.” To be more precise, one perfect duty (or more) must be compromised, beginning with the one with the weakest ground of obligation. And once again, this decision comes down to judgment: the agent must judge which action is most consistent with preserving respect for dignity and autonomy.\textsuperscript{110}

Whichever perfect duty (or duties) the agent judges to be most important constrains her action as described above, and then her decision-making continues as before, maximizing her objective function within the remaining constraints. Of course, she should not “take advantage” of the relaxation of the compromised perfect duty; it is not a perpetual license to do wrong. If it becomes possible to act in accordance with that perfect duty again, then the conflict is no longer an issue, and the duty once again becomes operative. But while it remains in “suspension,” the agent should regret that she cannot follow it, though she should not feel guilty, for she literally could not fulfill that obligation while maintaining the other one (or ones) which she judged more important.
3. Ranking perfect duty against imperfect duty. By now, the reasoning should be clear. When perfect and imperfect duties come into conflict, two things may happen. If the perfect duty is judged to be more important, then the imperfect duty is compromised to some degree—an option not available to us in the case of the compromise of perfect duty. If you planned to donate $1,000 to charity, and then realized you had forgotten a debt of $500 that you have to repay, you can always reduce your donation without forgoing it altogether. More generally, you still hold the end of charitable giving, but you decide that the planned, specific execution of the corresponding duty must be compromised to some extent. The apparent ease of accommodation in this case may have contributed to the presumption that in the case of such a conflict, imperfect duty always bends a knee to perfect duty. 

But it may not always be the case that imperfect duty is judged less important; despite Kant’s unfortunate rhetoric in pronouncing the supremacy of the duty to not lie to the murderer at the door who seeks to kill your friend hiding inside, the duty to help your friend to escape the murderer’s wrath almost certainly does take precedence, as our judgment easily confirms. So here we have a case of perfect duty being relaxed, for just as long as this particular conflict continues, to make room for proper execution of the imperfect duty judged to be of more importance. And to be sure, again this does not give the agent license to lie to just anybody, but only when a more important end or duty is at stake, and not without a certain amount of regret. In each of these cases of conflict, the agent uses her judgment, based on her appreciation of the moral law and its basis in the dignity and autonomy of human beings, to resolve the conflict to what must be her satisfaction (if not her presumptive confidence). And from such use of her judgment, she determines her operative duties in any given choice situation, and choice continues from that point as if she were an ordinary economic decision-maker.

In this chapter, we have been presuming an agent who embodies perfect morality and rationality, one who never wavers in her devotion to the moral law, whose will is unquestionable. Since such a person has never existed outside the realm of fiction, in the next chapter we further refine the model of Kantian decision-making to incorporate the will, in all its strength or weakness, and only then will we have a model of true choice.