1. This paragraph hardly does justice to the many attempts at modeling ethical behavior in economics; for a definitive treatment, see the relevant entries in Peil and van Staveren, Handbook of Economics and Ethics.


4. Frankena, Ethics, 15; see Gaus’s two-part article “What Is Deontology?” for a critical exposition of the term.

5. Such allowances are sometimes called agent-relative prerogatives, in that they permit an agent to say that doing x may be good, or that it would be good for someone to do x, but I shall not do it myself; contrast this with agent-neutral obligations, which bind any agent and do not allow for personal exceptions. (See Scheffler, Rejection of Consequentialism.) Sen deals with a similar concept, which he calls evaluator relativity, in “Rights and Agency” and “Evaluator Relativity and Consequential Evaluation.”

6. The “trolley problem” is a famous philosophical thought experiment that illustrates this dilemma (and many more). See Thomson, “Trolley Problem,” and Realm of Rights, chap. 7, for the most influential treatments; the original idea is credited to Philippa Foot (Virtues and Vices, 19–32).


8. Indeed, while Kant’s moral theory is usually put forth as the archetype of deontological ethics, there are scholars who instead call it teleological (Ward, “Kant’s Teleological Ethics”) or at least compatible with consequentialism (Cum- miskey, Kantian Consequentialism); see Guyer, “Ends of Reason and Ends of Nature,” for a thorough treatment of the role of teleology in Kant’s ethics. While teleology certainly plays a large role in motivating Kant’s ethical system on the whole, the moral imperatives given to individual agents are nonetheless essentially deontological in nature.

9. The first term in the title of this book, Grundlegung in the original German, is translated variously as “groundwork,” “grounding,” or “foundation.” I prefer and use the first, though the translation on which I rely uses the second.


11. I am not alone in this; for instance, Allen Wood also takes dignity as his starting point in his book Kant’s Ethical Thought.

12. Korsgaard, Self-Constitution, 108. (We will see much more of Professor Korsgaard and Self-Constitution in Chapter 3.)

13. For discussions of the various meanings of autonomy (including Kant’s), see Feinberg, Harm to Self, chap. 18; Gerald Dworkin, Theory and Practice of Autonomy, chap. 3; Hill, “Autonomy and Benevolent Lies,” 29–37, and “Kantian Conception of Autonomy”; and Irwin, “Kantian Autonomy,” 139–40.
14. This is hardly surprising, since many aspects of Kant’s moral and political theories overlap (by intention). Sullivan’s *Introduction to Kant’s Ethics* even begins with Kant’s political theory, and then draws analogies to it when detailing his moral theory; see Chapter 4 for more on this.

15. The concept of an agent’s proper relationship to outside influences, particularly as it relates to social identities, will be discussed further in Chapter 3.

16. And even when an agent does follow her desires unthinkingly, she still made a choice to do so, assuming she has the capacity for autonomous choice, and so forth; see Oshana, “Wanton Responsibility.”


21. See Hill, “Importance of Autonomy,” and Guyer, “Kant on the Theory and Practice of Autonomy.” In Chapter 3 in this book I discuss how an agent’s actions constitute her identity, but only if she acts autonomously.


23. Ibid., 434; the persons/things distinction itself appears on 428. (See also Sullivan, *Immanuel Kant’s Moral Theory*, 195–8.)


26. Hill, “Social Snobbery and Human Dignity,” 170. Also, Sullivan writes that “Kant’s entire moral philosophy can be understood as a protest against distinctions based on the far less important criteria of rank, wealth, and privilege, and perpetuated by religious and political force and fear” (*Immanuel Kant’s Moral Theory*, 197).

27. Hill, “Humanity as an End in Itself,” 49.

28. This point is especially pertinent in reference to the discussion of Kaldor-Hicks efficiency in Chapter 4 of this book.

29. McCloskey, *Bourgeois Virtues*, 263. Later in the book (353), she attributes this caricature to (some) modern followers of Kant, not the magister himself, a statement which (as you can well imagine) I found wholly gratifying.


31. Naturally, much scholarly attention has been given to analyzing the categorical imperative. See Sullivan, *Introduction to Kant’s Ethics*, chaps. 2–6, for an excellent introduction; Sullivan, *Immanuel Kant’s Moral Theory*, part III, for a detailed summary; and Paton, *Categorical Imperative*, for an in-depth analysis.

32. Paton lists five formulations of the categorical imperative (*Categorical Imperative*, 129), but two of them are variations on two of the three “standard” ones. (I
discuss the variant of the first one below.) Also, Kant gave no official names or titles to his formulae, so I base mine on Sullivan, *Immanuel Kant’s Moral Theory*, 149–50.

34. Ibid., 402–3, 422.
35. Through his innovative and original analysis of the Formula of Autonomy, Binmore seems to arrive at the same conclusion: that logical consistency alone cannot ground the categorical imperative (*Playing Fair*, 153–9).

38. *Groundwork*, 427; see also Herman, “Obligatory Ends.”
39. For the opposite understanding, see McCloskey, *Bourgeois Virtues*, 322.
40. Paton, *Categorical Imperative*, 76. Later, he writes that “it is no part of Kant’s doctrine that the moral law can be applied without any regard to empirical knowledge of the facts of human life” (91). See also Sullivan, *Immanuel Kant’s Moral Theory*, 159–60.

41. *Groundwork*, 412 (emphasis mine); see also *Metaphysics of Morals*, 216–7.
42. *Groundwork*, 429.
43. For instance, in his discussion of price and dignity in the *Groundwork*, Kant writes that “skill and diligence in work have a market price” (434); in the *Anthropology*, he writes that “all other good and useful properties of character have a price in exchange for others which have just as much use. Talent has a market price, since the sovereign or estate-owner can use a talented person in all sorts of ways. . . . But character has an inner value and it is above all price” (292). For more, see Sullivan, *Immanuel Kant’s Moral Theory*, 196–7.
44. For more on this point, see the prisoners’ dilemma discussion later in this chapter, as well as Chapter 5.

47. Ibid., 436.
50. Herman, “Murder and Mayhem,” 117.
51. Ibid., 118.

52. *Groundwork*, 421n12.
53. Much of this discussion of perfect and imperfect duties admits of exceptions; for instance, we saw above that “do not kill,” certainly a narrow, perfect duty, cannot result from the Formula of Autonomy.

55. See Gregor, *Laws of Freedom*, chap. 7, and Hill, “Imperfect Duty and Supererogation,” for an in-depth discussion of perfect and imperfect duty; we will revisit the topic of the latitude involved with imperfect duties in Chapter 3.
57. As we will see when discussing the prisoners’ dilemma game, in a way this is true of perfect duties as well: they tell us what not to do, but they do not tell us what to do instead. Do not lie, certainly, but don’t feel you must necessarily tell the truth either. (If you listen carefully, you can hear the collective sigh of relief from lawyers and politicians everywhere.)

58. *Metaphysics of Morals*, 454. This contrasts significantly with caricatures of Kant in which he is supposed to advocate an extreme version of altruism, where others’ interests take absolute precedence over one’s own, and extreme self-sacrifice is mandated. Instead, he wrote that “when I say that I take an interest in this human being’s well-being only out of my love for all human beings, the interest I take is as slight as an interest can be. I am only not indifferent with regard to him” (*Metaphysics of Morals*, 451), hardly an endorsement of ascetic generosity. (We will return to this point in Chapter 3.)

59. For instance, Aristotle wrote that it is not enough for a virtuous person to do acts which are generous or honest, but she must actually *be* generous and honest; she must embody, not merely exhibit, virtues. (See Hursthouse, *On Virtue Ethics*, chap. 4, for a comparison of Aristotle and Kant on this point.)

60. *Groundwork*, 393.

61. Ibid., 394.

62. Ibid.

63. Ibid., 398. As you may well imagine, this passage is often cited by those who accuse Kant of neglecting the role of emotions in human life. Kant did acknowledge some instrumental role of moral sentiment, mainly to help promote the performance of duty in cases where self-love may be too strong, for these feelings can support—but never replace—the requirements of the moral law. In the best case scenario, moral sentiments are psychological manifestations of the moral law, and if so can be regarded as truly moral. But otherwise, if these sentiments are not identical to the moral law, it is possible they can lead us astray from what duty requires, and this is the danger of relying on them to guide moral action. On these points, see Sullivan, *Immanuel Kant’s Moral Theory*, 132–3; Baron, *Kantian Ethics Almost Without Apology*, chap. 6; Sherman, “Place of Emotions in Kantian Morality”; and Sabini and Silver, “Emotion, Character, and Responsibility.”


65. For a much more nuanced view of suicide from a Kantian point of view, see Hill, “Self-Regarding Suicide.”


69. O’Neill, “Kant After Virtue,” 161. A characteristic passage from MacIntyre is: “In Kant’s moral writings we have reached a point at which the notion
that morality is anything other than obedience to rules has almost, if not quite, disappeared from sight” (After Virtue, 219); see also Louden, “Kant’s Virtue Ethics,” in response to McIntyre and other critics.

70. Groundwork, 404.
71. Critique of Pure Reason, A133/B172.
74. Wood, Kant’s Ethical Thought, 107. He adds, parenthetically, “Any moral theory that purports to offer one should precisely thereby have discredited itself in the eyes of any person of good judgment.”
75. Louden, Kant’s Impure Ethics, 25; on the interaction of Kant’s anthropology and his ethics, see also Frierson, Freedom and Anthropology in Kant’s Moral Philosophy.
76. Herman, “Practice of Moral Judgment,” 77.
77. Ibid., 75.
79. Groundwork, 389. (Yes, even the Groundwork discusses judgment.)
80. Metaphysics of Morals, 224.
81. In Chapter 3, I argue that a person’s unique faculty for judgment (along with her will, to be discussed in the next chapter) defines who she is, and together they comprise her character, which serves as her most essential identity and renders her unique.
82. This wording is inspired by the jurisprudence of Ronald Dworkin (Taking Rights Seriously, Law’s Empire), which informs my take on judgment; I plan to explore this much further in future work.
83. For more on social proximity, see Chapter 3.
84. Again, this is analogous to a Dworkinian judge balancing various principles relevant to a case at hand; see “Hard Cases.”
85. Herman, “Making Room for Character,” 54; see also Hill, “Moral Dilemmas, Gaps, and Residues,” arguing that “having gaps . . . is not unqualifiedly a defect in a [moral] theory. Gaps may reflect important features of our moral experience that closure would distort” (382).
86. This game is alternatively referred to as the prisoner’s dilemma and the prisoners’ dilemma. I choose the latter to emphasize that the problem arises only in strategic, multi-player situations.
87. Sen, “Choice, Orderings, and Morality,” 77. For a similar discussion of ethical behavior and the prisoners’ dilemma game, see Minkler and Miceli, “Lying, Integrity and Cooperation.”
88. Keeping the description of the game general avoids the complication of judging the intrinsic morality of the acts that characterize deviation and cooperation in any specific game. For instance, deviation in the standard bank robber
tale—telling the truth and cooperating with the authorities—is the more ethical choice from society’s point of view, but deviation in a public good financing problem—free-riding—is not.

89. Binmore sees some inconsistency here, but it is not clear to me why; he writes that “the maxim [never use a strongly dominated strategy] presumably fails Kant’s test because, if it were adopted as a universal law, people would not cooperate” (Playing Fair, 149).


91. Indeed, Kant explicitly criticized the Golden Rule—“do unto others as you would have them do unto you”—for the same reason, that it is grounded in subjective inclination rather than objective duty (Groundwork, 430n). (See Wattles, Golden Rule, 83–6, for a discussion of Kant’s critique of the Golden Rule.)


94. Groundwork, 402–3; see Hill, “Promises to Oneself,” for more detail on the Kantian ethics of promise-keeping, to oneself as well as to other people.

95. Groundwork, 429–30 (emphasis mine).


97. This statement holds even if the choice situation reflects incomplete information, or constraints resulting from legal or institutional coercion, neither of which are an agent’s responsibility towards others, and therefore do not affect the respect with which one treats another person.

98. For an extension of this logic to antitrust law, see Chapter 4.

99. Interestingly, this is the view also taken by most modern utilitarian philosophers, with the prominent exception of Richard Brandt (see Theory of the Good and the Right, chap. 13). For a defense of the desire-based account of utility, see Griffin, Well-Being, chap. 1.

100. The common assumption of the self-interested economic agent is usually not recognized as an essential part of the rational choice model, but is undeniably useful in showing that some socially beneficial outcomes can arise out of the interaction of narrowly self-interested actors (captured in Adam Smith’s famous “invisible hand” metaphor and modern general equilibrium theory, for example). See Walsh, Rationality, Allocation, and Reproduction, chap. 5, for an fascinating discussion and history of the self-interest assumption in economics.

101. Obviously, this is a very abbreviated summary of economic models of choice, as the myriad fine details are not essential for our purposes; for an insightful and critical treatment, see Hargreaves Heap et al., Theory of Choice.
102. Another argument against lexicographic preferences is based on the standard use of preferences in welfare economics to measure well-being, which is obviously complicated by having “supreme” preferences that must be satisfied before even the least expensive normal ones. But since I argue against the equivalence of well-being and preferences later, I choose not to emphasize this argument.


104. Sen makes a similar point, writing that even though a constraint can be modeled formally as a preference, such a preference is “a devised construction and need not have any intuitive plausibility seen as preference. A morally exacting choice constraint can lead to an outcome that the person does not, in any sense, ‘desire,’ but which simply mimics the effect of his self-restraining constraint” (“Maximization and the Act of Choice,” 193).

105. Gintis, *Bounds of Reason*, 73. To similar effect, Minkler writes: “While preferences and principles are independent from one another, individuals still possess preferences over principles . . . to the extent that if I act on that principle I may receive positive utility” (*Integrity and Agreement*, 3); as we’ll see in the next chapter, Amitai Etzioni also posits a (meaningful) utility gain from moral behavior. In “Limits of *Homo Economicus*,” Gaus poses the same problem more generally as a conflict between principles and values, but also assumes a rate of exchange between them.

106. For further discussion of moral preferences versus constraints, see Goldfarb and Griffith, “Amending the Economist’s ‘Rational Egoist’ Model, Part 2.” In a more recent paper with Dowell (“Economic Man as a Moral Individual”), they offer a model which represents a compromise between normal moral preferences and lexicographic ones. In their model, a preference for morality enters the utility function in such a way that the “utility from consuming a given set of goods varies in a ‘lumpy’ or ‘discontinuous’ way with the concurrent moral content of the individual’s behavior” (649). As a result, incentives below a certain level will not produce changes in moral behavior, but large ones may, depending on the size of the discontinuity representing the agent’s aversion to immorality. (In a way, this mimics the outcome of the model of the will to be described in the next chapter.)

107. Goldfarb and Griffith, “Amending the Economist’s ‘Rational Egoist’ Model, Part 2,” 65. Or, as James Buchanan put it, “the ordering over goods cannot be separated from the means through which goods are expected to be secured” (“Choosing What to Choose,” p. 85). Sen calls this process *significance*; see his “Maximization and the Act of Choice.”

108. Kant, *Metaphysics of Morals*, 451; I think “permit” here is misleading, since treating yourself as an end would mandate, not simply allow, some degree of self-interested action. As Herman writes, “the Kantian charge against self-interest is not that it is inherently contrary to morality; it need not be. The problem lies in its tendency to be presumptive” (“Making Room for Character,” 52). Con-
trast this with virtue ethicist Michael Slote, who writes that “Kant’s conception of morality . . . provides no basis for assigning positive moral value to actions to the extent they merely promote the happiness or well-being of their agents” (“Some Advantages of Virtue Ethics,” 440).


110. See also Hill, “Moral Dilemmas, Gaps, and Residues.”

111. Kant, “On a Supposed Right to Lie.” I am afraid Kant shot himself in the foot with this one, and wasted innumerable hours of good scholars’ time in attempting to hold his moral theory together in light of it.

CHAPTER 2


2. Ibid.


4. “Human Freedom and the Self,” 34. Chisholm contrasts this view with the “Kantian approach” (with which he agrees), in which “there is no logical connection between wanting and doing, nor need there even be a causal connection” (34–5).


6. On bounded rationality, see Simon, “A Behavioral Model of Rational Choice”; and on cognitive biases, see Kahneman and Tversky, “Prospect Theory.”

7. Behavioral economics, inspired by Kahneman and Tversky’s work among others, will come up again in Chapter 5.

8. See the introduction to Stroud and Tappolet, *Weakness of Will and Practical Irrationality*, for a summary of research on weakness of will; more on this topic later in this chapter.


12. Davis, “Identity and Individual Economic Agents,” 73. Davis’s ideas regarding identity play an important role in the next chapter.


16. Wallace, “Addiction as Defect of the Will,” 172. This is reminiscent of Thorstein Veblen’s criticism of “the hedonistic conception of man [which] is that of a lightning calculator of pleasures and pains who oscillates like a homogeneous globule of desire of happiness under the impulse of stimuli that shift him about the area, but leave him intact” (“Why Economics Is Not an Evolutionary Science,” 389).