

Chapter 1

Why Integrity?

Consider the proposition “lying is wrong.” True or false? My experience has shown that most people quickly answer “true.” Not surprisingly, the question tends to be a little more problematic for economists.

The first time I asked someone whether lying is wrong, it was of a job candidate. When she dutifully asked what I was working on I innocently asked the question above to help explain. The look on her face was a mixture of confusion and terror. I wasn’t trying to trick her, but I could sense her apprehension in weighing her options. If she answered “true,” she might have thought I would not think her a good economist, while if she answered “false,” I might spring some fancy ethical trap. Had she answered “false” it likely would have been because she realized that economists don’t deal much with ethics. We deal with preferences and rationality, and thus verbal claims against preferences are just noncredible “cheap talk.” Her hesitancy signaled she was not a certain type of naturalist, for whom ethical questions like the one posed are nonsense, in which case she could have immediately blurted out “false.” On the other hand, she may have had some faint recollection about the notions of right and wrong from before her graduate training and thought that somehow she should answer “true.”¹ Her dilemma was apparent. She decided against either response, and after some awkward mumbling we talked instead about New England’s strange weather.

Since the question refers to a moral principle, it asks about one’s understanding of the world. Were one to thoroughly embrace the principle, so much so that it formed part of one’s identity, we might be able to say that the person was a person of integrity. In that case, to violate a moral principle the person takes to be true (by lying) would be to act contrary to her own under-

1. She might also have objected to the categorical nature of the question. Many I have asked wanted to say “true,” but could think of exceptions like lying to prevent harm or “white lies.” I will discuss those qualifications in detail later, but when I modify the proposition to “lying is wrong in normal circumstances,” it really doesn’t seem to change the number of “true” and “false” answers.

standing of the world and thus her understanding of her own self as well, which would be incoherent. Even in a weaker case, one more consistent with current economic theory, if the person did not identify with the moral principle but instead had a strong preference over it, she would still have reason not to lie.

In this book I will develop the notion of two different kinds of integrity. One focuses on identity-conferring commitments to principles, the other on preferences for principles. But while they differ in their particulars, most fundamentally, both refer to honest behavior. The possibility of integrity and honest behavior is important any time there are economic agreements.² That covers a lot of economics. This book will cover informal interactions that characterize social dilemmas, formal legal agreements between buyers and sellers, political agreements, employment agreements, religious agreements, and the social contract. In each of these cases, the possibility of integrity changes the answers given by neoclassical economics and instrumental rationality (rational decision making in one's own self-interest) because *to make an agreement that one has no intention of honoring is to lie*. This simple but powerful fact seems to have eluded most economic analyses.

The analysis here will not require heroic assumptions. We will not focus on the few individuals who might embrace myriad moral principles and possess the kind of strength of will that, like Don Quixote, leads them to try to right every wrong. But neither will we confine ourselves to the kind of lead character presented in Ben Jonson's play *Volpone*. While nothing prevents a rational economic man from *also* being moral, when it comes to economic agreements, he is most often characterized as opportunistic.³ More beautifully articulated, *Volpone* is ruthlessly selfish and engages in

. . . sanctimonious speeches, lust and possessiveness poorly disguised as love and marriage, cynical legalism passing itself off as pure justice, boastful name-dropping that pretends cultural sophistication, snobbery congratulating itself that it is decorum, and greed deluding itself that it is really prudence, responsibility, even religion.⁴

2. Earlier works that touch on the importance of honesty in economic interactions include Sen 1978, Adler 1992, and Bowles 1998.

3. One might protest that self-interest is fundamentally moral in and of itself. Indeed, Holmes (1990) persuasively argues that self-interest as developed by Adam Smith and his contemporaries did provide a contrast to the prevailing doctrines of honor and original sin, and that universal self-interest is egalitarian and democratic because it requires respect for everyone's interests. Nevertheless, moral philosophers usually maintain that when it comes to the moral behavior of an individual, the moral rightness of any act is independent of what the person might prefer.

4. Jonson 2003, ix.

Even though Volpone provides a colorful foil for the ruthlessly moral Don Quixote, we will instead focus most of our attention on a normal person in normal economic circumstances. Like most of us, she takes at least some principles seriously, but she also has to grapple with other considerations, most notably her own preferences and tempting rationalizations. But we do know that average people in real life do keep their agreements *even when they could cheat*. Plumbers sometimes do as they say, workers do provide good efforts, business partners don't cheat each other, contractors keep their agreements, people follow their religious vows, and even politicians sometimes do what they promise. These kinds of honored agreements are what make the economy and society work.

It will be necessary to aim a fair amount of attention at the existence of principles. Sometimes principles are taken to be truths or laws or as starting points for reasoning. As I have already suggested, I characterize principles as general propositions that can be taken as either true or false. Once a principle is accepted as true, it provides a sort of universality that applies to beliefs and actions. If I accept "lying is wrong" as true, I am not free to interpret it as meaning I have to tell the truth only to right-handed people but not necessarily to left-handed people. There may be exceptional cases like lying to prevent great harm, but most candidates for such exceptions fail a meaningful-relevance test. Moreover, this universal character also applies no matter what my preferences are. I may not *want* to tell the truth to some left-handed person, but that doesn't mean that "lying is wrong" suddenly becomes false. The principle holds in spite of my own preferences. Thus principles and preferences are different, and so both will be included in the analyses in this book.

While preferences and principles are independent from one another, individuals still possess preferences over principles. I possess a preference over "lying is wrong" to the extent that if I act on that principle I may receive positive utility. The more important point now is that it is the existence of principles that forms the basis for one kind of integrity, *commitment-integrity*, and the preference for principles that forms another kind, *preference-integrity*. The former requires individuals to make judgments about the truth or falsity of principles, while the latter only requires a complete preference mapping.

If individuals recognize the existence of principles and include them in their decision-making process, the next question becomes which one or ones should we focus on. While some of the analysis will consider additional principles, the principle I have already referred to, "lying is wrong," turns out to be both fundamental and also the principle most relevant for economic agreements. Virtually all religious and philosophical traditions counsel against ly-

ing, even if they differ in their justifications and exceptions.⁵ It holds the same elevated status as “murdering innocents is wrong” in most traditions, in the sense of its barely qualified condemnation. The principle is also most relevant for economic agreements because all agreements are representations, and lying is an intentional misrepresentation. Therefore, most of the analysis centers on what I call the *minimalist principle*, “lying is wrong.” And by basing it on just one fundamental principle, the analysis stays more tractable.

To what extent do people base their acts on principled decision making? As an empirical matter we do know that the behavioral postulate of instrumental rationality is violated with great regularity. Most commentators on the subject have pointed to commonplace behaviors not involving agreements, like tipping, tax paying, voting, contributing to public goods, and helping others when there is no chance of reciprocity. Sometimes people are even heroic.⁶ Overwhelming experimental evidence also shows that people often cooperate against their material interests. Experimenters use social dilemmas like a Prisoner’s Dilemma or voluntary contribution to a public good where, based on instrumental rationality, the dominant strategy is to defect. But consistently great numbers do not defect, even where there is anonymity and the game is played only once.⁷ In his meta-analysis of 37 different studies consisting of 130 distinct social dilemma experiments, David Sally calculates a mean cooperation rate of 47.4 percent for the entire pooled sample.⁸ The 2002 award of the Nobel Prize in Economics to psychologist Daniel Kahneman decisively legitimated the behavioralist school with its critique of instrumental rationality.⁹ That does not mean, however, that instrumental rationality is not still a useful theoretical construct or a good first empirical approximation. The findings of the behavioralist school (and perhaps others) may yet be incorporated. One of the virtues of neoclassical economics is that it can fold new ideas and findings into its considerable maw, sometimes almost seamlessly (witness models of imperfect information).

It also seems beyond question that many normal people in normal circumstances have an aversion to lying for material gain. Again, the limited ex-

5. See Bok 1978, which I will often refer to and will discuss in some detail in chapter 5.

6. Evidence and analysis of rescuing and heroic behavior is discussed in Mansbridge 1998.

7. See, for instance, Marwell and Ames 1981; Schneider and Pommerehne 1981; Caporael, Dawes, Orbell, and van de Kragt 1989; Davis and Holt 1993; Frey and Bohnet 1995; and Ledyard 1995.

8. Sally 1995, 62. Also see Ledyard’s (1995) review for similar findings. Caporael, Dawes, Orbell, and van de Kragt (1989) provide particularly good evidence against self-interested behavior as normally conceived.

9. See Kahneman 2003 for a brief review of the evidence against “selfishness,” rationality, and unchanging tastes.

isting evidence is supportive. One empirical regularity in experimental settings is that communication significantly increases cooperation.¹⁰ More to the point, David Sally estimates in his meta-analysis that (noncredible) promises to cooperate elicited by experimenters increase cooperation by 12 to 30 percent, depending on the regression model.¹¹ That finding in particular suggests that even people in contrived situations recognize and act on the minimalist principle. More recently, Uri Gneezy experimentally tested the propensity to lie by varying the harm that lying causes. His main finding is that “the average person prefers not to lie, when doing so only increases her payoff a little but reduces the other’s payoff a great deal.”¹² Social psychologists similarly find that most lies are “white lies” aimed at affecting the perceptions of others rather than lies with the purpose of material gain.¹³

Finally, for admittedly unconventional evidence, consider the offerings of classical literature. I have already mentioned Don Quixote, Cervantes’s beloved but doomed character who unrelentingly champions good. Don Quixote embraces all moral principles, delusionally perhaps, and repeatedly puts his and his squire’s life on the line to defend them. Or consider Joseph Conrad’s *Lord Jim*. Jim is the confident, flawlessly principled sea ship officer. Yet in one inexplicable moment, he abandons his ship loaded with passengers when he believes it to be hopelessly doomed. Jim then spends the rest of his life trying to recover the sense of himself that he himself deserted. In one case we affectionately root for the character who fights for the impossible, in the other we sympathetically mourn for the one who abandons that which seems so possible. The kinds of affection and sympathy that these characters have evoked in generations of readers is deeply felt because integrity matters. Don Quixote represents the flawed man who reaches for perfection—unattainable integrity; and Jim, the ideal man who falls from grace—lost integrity. We are

10. John Ledyard in his 1995 chapter on public goods in the *Handbook of Experimental Economics* concludes that the evidence on pre-play communication counts as a “strong effect” that increases cooperation. For instance, Dawes, McTavish, and Shaklee (1977) report that payoffs increased from 31 to 71 percent when relevant communication was allowed in one-shot games. Isaac and Walker (1988) find contribution rates of over 80 percent in one-shot public goods type games with communication, and over 90 percent in repeated games.

11. Sally 1995, 78.

12. Gneezy 2005, 385.

13. DePaulo et al. (1996) examined the daily diaries of 77 college students and 70 communities to come to this conclusion. While they do find evidence that some lies are told for personal advantage, they “think that lies are less often told in the pursuit of goals such as financial gain and material advantage and instead are much more often told in the pursuit of psychic rewards such as esteem, affection, and respect” (981). Such lies are also told to minimize the tension with, and hurt feelings of, others. Thus, these kinds of lies are not seen as serious by the liars.

moved by their stories time and again because we think integrity is real and important.

Still, I cannot claim that the evidence suggests that principled behavior or integrity is ubiquitous. People are certainly interested in integrity—in 2005 it was the most looked-up word on Merriam-Webster's Web site.¹⁴ And it may turn out that integrity is more pervasive than we know, we just haven't really looked carefully enough. To the extent that it is desirable, we might look into ways of increasing its frequency (the topic of chapter 9). Of course, adding integrity to the motivational mix might only add to the cost of complexity with little or insufficient benefit. That could be the methodological position of those embracing Milton Friedman's positivism, or some version of it.¹⁵ On that view, a good theory is one that predicts well, quite apart from the realism of the theory's assumptions. When combined with Occam's razor, the best theory becomes the one with the simplest assumptions that also predicts best. I would call this view *methodological simplicity*, and, especially when considering the already staggering complexity of some economic models, simplicity is a virtue.

But I think *methodological realism* is better. In the version that I favor, we should aim for a theory whose assumptions and predictions are *both* true. That is, both the theory's assumptions *and* predictions should aim to correspond to the way things really are, or possibly could be. On this view, the real world exists independently of what we think about it. Our histories, values, and biases may affect the selection of our theoretical analyses, but the intention is to bring our thinking into correspondence with the objective, existing state of affairs, at least to the extent that we can.¹⁶ Of course the goal of some theorizing may be simply to develop models that are logically and internally consistent without reference to the world. The goal of those kinds of analysis too is a kind of logical truth. But if the pursuit of truth contains some reference to the real world, it seems incongruent to insist on truth in one aspect of

14. Associated Press report, *Hartford Courant*, December 11, 2005.

15. Friedman 1953.

16. One can find many different brands of realism in the literature on methodology (e.g., Maki 1988; Lawson 1997). There are a lot of interesting epistemological and ontological issues. My view is that there is an objectively existing real world, we can possibly know that world, the aim of economic theory is to accurately represent that world, and a theory is better than another if it *more* accurately represents the world. Because such representations are subject to human comprehension, to say that one theory more accurately represents the world leaves open the possibility of interpretation and thus persuasion. Moreover, our thinking can also *change* the world because thinking precedes acts and acts change the world. One kind of thinking has changed the world by introducing genocidal acts; another has changed it by reducing or eliminating disease. But to the extent that we truly understand the world, we may be able to intentionally affect it with corresponding policies.

a theory (e.g., implications) but not another (e.g., assumptions). If we are committed to seeking the truth about the world, we should not partition it selectively.

The concern about the added complexity of real assumptions is legitimate, however. The assumptions that we do make should be true, but that does not answer the question about how many true assumptions we should include in our analyses. The issue becomes one of balancing the costs and benefits, and reasonable people may disagree. In this book, beyond instrumental rationality—an assumption I think is true for many people at many different times—I will only add one further assumption in the case of preference-integrity, namely, that a preference for honesty exists, and one set of assumptions in the case of commitment-integrity, namely, those relating to moral principles. This added realism generates novel implications.

As a preview, here are a few of the more interesting implications generated by incorporating one version or another of integrity into analyses of economic agreements. By including the possibility of preference-integrity in Prisoner's Dilemma-type interactions, modeled as a continuum of types, we will see that three distinct types of players emerge. Perhaps most important, a sufficient amount of integrity results in higher payoffs by promoting a critical amount of cooperation. Trust results from honesty, not vice versa. With respect to legal contracts, integrity can be consistent with contractual breach, but not fraud. In the political arena, either kind of integrity reduces political shirking (voting against constituents' interests) and can help explain why legislators don't change their votes in their last term of office. In the employment context, adding commitment-integrity leads to an implication that contrasts starkly with the literature on the theory of the firm. Whereas the received literature focuses on employee (or contractor) shirking and then recommends monitoring, incentive contracts, and hierarchy, the analysis presented here implies that creating clear, fair agreements will take care of misaligned expectations, which leaves the door open for focusing on the coordination of different input contributions, especially knowledge ones—the real essence of the firm. We will also see that adding integrity can help to solve some of the anomalies in the economics-of-religion literature. For instance, a person of religious integrity has reason to engage in counterpreferential behaviors like refraining from eating forbidden foods, even in private. As a last example, we will see that, based on the minimalist principle, a person of commitment-integrity also has reason to support a more general social contract like human rights, under some fairly routine conditions.

Two concepts in particular drive the analysis: *mutual deceit* and *coherence*. Mutual deceit pertains to the minimalist principle “lying is wrong.” Not

surprisingly, there are some contexts in which it is morally acceptable to lie, like buying and selling at a bazaar. Economic agents themselves will treat some, but not all, situations as ones of mutual deceit. When contractual partners do not consent to lies, however, integrity does indeed matter. So a lot of emphasis will be placed on the context of economic agreements. Coherence, on the other hand, is a normative requirement of commitment-integrity. To illustrate, if one commits to the minimalist principle but then lies wildly, that is a form of incoherence. Lying to a left-handed but not to a right-handed person is also incoherent. Normatively, in either case, such a person could not be said to be a person of integrity. As a positive matter, some people surely recognize this coherence requirement and behave accordingly. Others may fail to act coherently because of a weak will, self-deception, errors in judgment, or moral exclusion—all factors that will further contribute to the analysis.

To my knowledge, no other book has used integrity as the basis to study agreements. Nevertheless, several pioneering studies have touched on the same, or closely related, issues. Perhaps most central has been Amartya Sen's illuminating use of commitments as something different from preferences.¹⁷ While philosophers have recognized the distinction for centuries, it has been a significant departure for economists only recently.

Other major works often focus on altruism or cooperation. For instance, Robert Frank, in his seminal book *Passions Within Reason* (1988), asks how mutually beneficial interactions could occur in a world populated by both selfish and cooperative people. By using evolutionary processes as a metaphor, people are predisposed to either cooperate or act selfishly. If cooperators can interact with each other they get a higher joint payoff than when selfish people interact with each other. But, as in a Prisoner's Dilemma game, the highest payoff goes to those selfish people who can interact with cooperators. The problem is, how can cooperators credibly commit to actually behaving cooperatively? Frank's answer is that cooperation is a moral trait subject to evolutionary forces (i.e., it must confer an important advantage), and that emotions both sustain and signal the presence of this moral trait. If detecting this signal in others requires a resource cost, Frank finds an evolutionarily stable equilibrium such that cooperators and selfish people exist simultaneously. Justifiably, Frank's approach and model have been enormously influential.

The real difference lies in our different approaches. While Frank locates moral behavior in the emotions, and also gives them a strategic role, my analysis locates moral behavior in the existence of a preference for honesty in the case of preference-integrity, and conscious and reflective deliberation in the case of commitment-integrity. While Frank's analysis follows philosopher

17. Sen 1978.

David Hume because moral behavior stems from irrational moral sentiments, mine features the usual notion of rationality in the case of preference-integrity and, in the case of commitment-integrity, follows philosopher Immanuel Kant because moral behavior stems more from a person's ability to reason. That is not to say our accounts are incompatible; it is to say they are conceptually quite distinct (more on this in chapter 4).

Commitment-integrity, in particular, is also conceptually quite different from those preference-based accounts that similarly emphasize cooperative, social, and/or moral behavior. For instance, in *Not Just for the Money*, Bruno Frey (1997a) builds the case that people sometimes engage in behaviors simply because they like to. People might do a good job at work because they like to; they may vote because they like civic participation; they may give blood or help another because they want to indulge their altruistic preferences. Borrowing from the psychology literature, the key insight is that sometimes people will indulge these kinds of preferences *less* often when there are incentives to act on them. Sometimes such extrinsic interventions can crowd out intrinsic motivations, particularly if the interventions are seen as controlling. So, sometimes people just enjoy doing a good job, and incentives to work hard may actually reduce such efforts. I will consider the issue more fully in chapter 6, but doing a good job because one wants to versus doing a good job because one promised to are two very different things. It turns out that the evidence suggests both are important.

Another strand of the literature features multiple-preference and utility models. For instance, in *Selfishness, Altruism, and Rationality*, Howard Margolis (1982) offers a model in which individuals get two irreducible kinds of utility from two different kinds of preferences: private and social. Self-interested acts fulfill one kind of preference, altruistic acts another kind. Innovatively, Margolis formulates a rule, operationalized as a weight, that gives the ratio of spending in each category of preference necessary to result in an equilibrium. That weight, in turn, is subject to Darwinian evolutionary forces. Evolution selects not for those who are narrowly self-interested but instead for those individuals who also participate in groups. Of course that is a very different kind of account than one that relies on conscious reflection and commitment to principles, but in any case the multiple-preference and utility literature will be briefly assessed in chapter 3.¹⁸

18. Along somewhat different lines from Margolis, but in line with the arguments in this book, Amitai Etzioni, in *The Moral Dimension* (1988), offered an influential framework in which individuals are motivated by both moral principles and self-interested preferences. Etzioni takes a "moderately deontological" position with respect to moral acts: their rightness stems mostly from an individual's intent, not the estimated outcome. Ultimately, however, Etzioni conceptualizes such moral motivations in terms of moral *preferences*.

If my explanation for why people would tell the truth in economic interactions differs from that of economists, so does it differ for why people would lie. For instance, in *Private Truths, Public Lies* Timur Kuran (1995) suggests that people sometimes engage in preference falsification, a type of lie in which one misrepresents her true motivation or disposition in order to manipulate the perceptions of others. For instance, such a person might disingenuously feign a private preference in public in order to enhance her reputation. In contrast, for those who might otherwise be tempted to act with commitment-integrity, lying stems more from a failure to choose moral principles, weakness of will, errors in judgment, self-deception, or moral exclusion. Moreover, and once again, we largely focus on different contexts. Kuran looks at scenarios where social pressure might induce one to conform against one's true preferences (e.g., revealing one's true political views). Since the topic of this book is economic agreements, the social pressure of fidelity would most often reinforce the behavior one promised in the first place.

The organization of the book is straightforward, beginning with the development of the theory. Chapter 2 introduces preference-integrity, sketches a simple framework, and assesses its strengths and weaknesses. Chapter 3 does the same for commitment-integrity, but since the concept is more of a departure, the chapter also goes into some depth to include its philosophical and psychological foundations. This chapter provides the core theoretical foundation for the book. Chapter 4 considers game theory and social dilemmas. The mathematical analysis in these chapters is simple, to keep the analysis as accessible as possible.

Each of the four chapters that follow explores a different kind of agreement. Chapter 5 first goes into some detail about the notions of lying and mutual deceit in order to provide a solid foundation for the minimalist principle. The chapter then considers two different kinds of agreements, legal and political. Chapter 6 considers the very important employment agreement. As I alluded to earlier, the existence of workers of integrity is particularly consequential for both actual firms and also the literatures on the theory of the firm. Chapter 7 considers religious agreements and behavior. Religious principles are included, which necessitates investigating other issues, such as multiple identities. The chapter provides an example for what happens when we move beyond the minimalist principle. Chapter 8 investigates the normative requirement for a person of integrity who embraces only the minimalist principle, vis-à-vis human rights, a particular kind of social agreement. The role of coherence is particularly important here, and it provides the basis for thinking that persons of integrity have reason to support human rights.

Assuming that integrity is something people value, the question then be-

comes how it could be fostered. Chapter 9 addresses that psychological question with a four-part decision-making framework by specifically considering moral training and moral leadership. Finally, chapter 10 summarizes the thesis, addresses some remaining objections, and considers how integrity might rank with other dispositions worth valuing. While I try to anticipate the skeptics' criticisms throughout, in this last chapter I address the concerns of a perhaps more sympathetic audience—feminists and others, who might claim that the analysis here does not leave enough room for dispositions like emotions or caring.

The kind of integrity considered in this book is not particularly special. That's one reason why it merits further study. To achieve that goal, I will occasionally motivate chapters by deploying our friends Don Quixote and Volpone to illustrate opposite ends of the spectrum. To stand in for an ordinary person in the swirl of the complex middle, I will place Martha. Martha finds herself to be a party to different kinds of economic agreements that require decisions. Does she follow the dictates of integrity, or the tug and pull of her countervailing preferences?