

*Truthfulness, Lies, and Moral Philosophers:
What Can We Learn from Mill and Kant?*

ALASDAIR MACINTYRE

THE TANNER LECTURES ON HUMAN VALUES

Delivered at

Princeton University
April 6 and 7, 1994

ALASDAIR MACINTYRE is Arts and Sciences Professor of Philosophy at Duke University. He was educated at Queen Mary College, University of London, and at the University of Manchester. He taught at various British universities, including Oxford and Essex, until 1970. Since then he has taught at a number of American universities, most recently at Vanderbilt University where he was the W. Alton Jones Professor of Philosophy from 1982 to 1988, and from 1989 to 1994 at the University of Notre Dame, where he was the McMahon/Hank Professor of Philosophy. He is past president of the eastern division of the American Philosophical Association. His numerous publications include *Three Rival Versions of Moral Enquiry* (1990), *First Principles, Final Ends and Contemporary Philosophical Issues* (1990), *Whose Justice? Which Rationality?* (1988), *After Virtue* (1981), *Secularization and Moral Change* (1967), and *A Short History of Ethics* (1966).

When children are still quite young, they learn not one, but two rules concerning truth-telling and lying and these in very different ways. One of those rules they learn by explicit instruction, characteristically when they have first been discovered in a lie. What they are then taught is that it is wrong to lie, but what the rule is that is invoked notoriously varies from culture to culture and sometimes within cultures. For some lying as such is prohibited. For others some types of lie are permitted or even enjoined, but about which types of lie are permitted or enjoined there are also significant differences. It is not difficult to understand why. Among those types of lie that are often permitted or enjoined in different social orders are certain types of protective lie, lies designed to defend oneself or one's household or community from invasive hostility, perhaps from religious persecutors or witches or the tax-collectors of some alien power, or to shield the vulnerable, perhaps children or the dangerously ill, from knowledge thought to be harmful to them. Since who is judged to need protection from what varies from one social and cultural order to another, which of these types of lie are permitted or enjoined can be expected to vary accordingly. But these are not the only types of exception that are sometimes accorded social recognition and sanction. And, unsurprisingly, reflection upon how the rule that provides for such different types of exception should be formulated and justified commonly gives rise to controversy. Consider as one contributor to those controversies a moral tradition that belongs to the background history of our own moral culture.

One of the earlier statements of that tradition, often appealed to later on, is in Book III of the Republic (382c-d), where Socrates is represented as describing some lies as useful against enemies or for the prevention of evils. Some Greek patristic theologians, among them St. Clement of Alexandria, held similarly that on occasion untruths might be told, for example, to protect the Christian community from the invasive enquiries of persecutors. About precisely what classes of untruths were permitted they and later writers sometimes differ from each other, and they also disagree among themselves in the precise

statement of the view that they share, some saying that all lying is prohibited, but that an untruth told for a just reason is not a lie, others that some lying is not prohibited. Newman in summarizing their shared standpoint emphasized that all of them agree that the occurrence of such a just reason “is, in fact, extreme, rare, great, or at least special” (*Apologia pro Vita Sua*, note G). Modern exponents of this view, he adds, include John Milton, Jeremy Taylor, and St. Alfonso di Liguorio. None of these were, of course, consequentialists. Their position was expressed succinctly by Samuel Johnson: “The general rule is, that Truth should never be violated, because it is of the utmost importance to the comfort of life, that we should have a full security by mutual faith There must, however, be some exceptions. If, for instance, a murderer should ask you which way a man is gone, you may tell him what is not true, because you are under a previous obligation not to betray a man to a murderer But I deny the lawfulness of telling a lie to a sick man for fear of alarming him. You have no business with consequences; you are to tell the truth” (James Boswell, *The Life of Samuel Johnson*, June 13, 1784).

John Milton, Jeremy Taylor, and Alfonso di Liguorio would all have agreed with Johnson that there is indeed an hierarchical ordering of duties and obligations and that any type of exception to an otherwise universal binding rule can be justified only as required by some other binding rule that is superior in that ordering. But Johnson’s statement suggests at the very least consequentialist questions. If there is indeed an ordering of duties and obligations, what is the principle by which they are ordered, if it is not a consequentialist principle? The consequentialism of J. S. Mill, for example, was intended to provide, by means of the principle enjoining the promotion of the greatest happiness of the greatest number, a standard for just such an ordering. What an evaluation of consequences by means of that principle is to tell us is which binding rules in practice at least have no exceptions (or almost so; see the penultimate paragraph of chapter 5 of *Utilitarianism*) -the rules prescribing justice, for example

— and which do have a few well-defined classes of exception, such as that otherwise prohibiting lying. And the onus seems to be on the adherents of Johnson's Christian anticonsequentialism to offer us an alternative and rationally superior principle of ordering. Moreover, if the rule prescribing truthfulness is to be defended as Johnson defends it, further consequentialist questions are raised. Conformity to the rule prescribing truth-telling seems for Johnson to be a means to a further end, what Johnson calls "the comfort of life," a necessary condition for which is "that we should have full security by mutual faith." But insofar as this rule is treated only as a means to some such further end, no matter how important, the possibility of evaluating the consequences of making a few well-defined exceptions to it has been opened up. And once again we need to know why we should not move to some more general consequentialist position, such as Mill's.

One answer to this question may well be that I have only reached a point at which it seems difficult to reply to consequentialist claims, because I erred in my starting point. I began after all by considering the kinds of explicit rules that are taught to young children when they are first detected in a lie, perhaps at three or four years of age, and at once noted that often such rules allow for exceptions to the general prohibition of lying. But, it may be said, I ought to have begun with another, more fundamental exceptionless rule, one learned somewhat earlier and not by explicit instruction. This is the rule prescribing truth-telling that we all learned to follow by learning to speak our native language, whatever it is. That rule governs speech-acts of assertion. To assert is always and inescapably to assert as true, and learning that truth is required from us in assertions is therefore inseparable from learning what it is to assert. So two Danish philosophers of language, H. Johansen and Erik Stenius, suggested that "the utterance of a falsehood is really a breach of a semantic rule" (Erik Stenius, "Mood and Language Games," *Synthèse* 17, no. 3 [1967], 269), *although Stenius understood the relevant rule as one concerning what he called the language-game of reporting, while in fact it is assertion*

*in general — acts of reporting are only one species of acts of assertion — that is governed by the semantic rule “Assert p, only if p is true.” Mary Catherine Gormally has more recently characterized the relationship of lying to assertion by saying that “a lie (in language) is a cheating move in the language-game of truth telling” (“The Ethical Root of Language” in *Logic and Ethics* ed. P. Geach [Dordrecht: D. Reidel, 1991], p. 53) and by further arguing that “ ‘assertion’ . . . carries moral weight, like ‘property,’ ‘right’ and ‘obligation.’ It is a value-laden concept” (p. 65). It is, that is to say, among those “concepts which are used to describe human actions in a way which makes it appear why our actions or omissions are bad if we act in certain ways, or fail to do so” (p. 67).*

Note that the rule enjoining truth-telling in speech-acts of assertion is constitutive of language-use as such. It is a rule upon which therefore all interpreters of language-use by others cannot but rely. And it is not merely a rule of this or that particular natural language. Hence Gormally concluded that about it “one cannot be culturally relativistic” (p. 58), in this following Peter Winch, who had argued that it would be “nonsense to call the norm of truth telling a ‘social convention,’ if by that were meant that there might be a human society in which it were not generally adhered to” (“Nature and Convention,” in *Ethics and Action* [London: Routledge and Kegan Paul, 1972], pp. 62-63). And David Lewis (who has also argued that in part our commitment to truthfulness in speech is a matter of convention, since “a language \mathcal{L} is used

by a population P if and only if there prevails in P a convention of truthfulness and trust in j, sustained by an interest in communication,” “Languages and Language,” in *Philosophical Papers* [Oxford: Oxford University Press, vol. 1, 1983], p. 169) says about what he calls the “regularity of truthfulness and trust simpliciter” and characterizes as “the regularity of being truthful and trusting in whichever language is used by one’s fellows” that it “neither is a convention nor depends on convention” (p. 184). *We stand, so all these writers agree, and surely*

rightly, in the same relationship to speakers of other languages in respect of the semantic requirement of truthfulness in assertion as that in which we stand to other speakers of our own language, a relationship defined by the rules governing the use and interpretation of asserted sentences as such.

What then are these rules, if they are not conventional? Winch's answer was framed in terms of the distinction that Aristotle drew between natural and conventional justice, by saying of the precepts of natural justice that they "have the same power everywhere and do not depend for it on being accepted or rejected" (*Nicomachean Ethics* V, 1134b19-20). This characterization of the natural holds equally of the semantic rule requiring truthfulness in assertion, which, like the precepts of natural justice, cannot but be accorded universal recognition, and in the vast majority of cases obedience, by the users of all natural languages. In Aristotle's terms the generally tacit semantic rule enjoining truth-telling is to be accounted natural because recognition of it belongs to the essential nature of human beings as language users.

We notice at once that liars cannot withhold recognition from it any more than the truthful can, and this not only because even habitual liars cannot but tell the truth far more often than they lie, sustained in their truth-telling by the interest in communication that, as Lewis emphasized, they share with everyone else. But liars have in addition their own distinctive interest in general conformity to that rule. For they can only hope to lie successfully insofar as it is taken for granted by others that the rule requiring truthfulness in assertion is respected, more particularly by the liar herself or himself. The liar, as Kant put it, cannot consistently will that the maxim upon which she or he acts in lying should be, and should be understood to be, the universal rule governing truth-telling and lying. What successful lies achieve for those who utter them is an advantage with respect to information over those who are deceived. And successful liars necessarily deceive us not only about the subject matter about which they lie, but also about their own beliefs and about their intention in asserting

what they assert falsely, and indeed about their further intention to conceal this intention from us. So that even in the simplest cases of lying there is a complexity in the liar that is absent from the truthful person. Truthful persons may have much to conceal, including their own intentions not to disclose what they are concealing. But they do not misrepresent themselves to others as liars do, with regard to the relationship of their beliefs and their intentions to their assertions.

The kinds of advantage to be gained by lying are of course various and so therefore are the motives for lying. Many lies, as I noticed earlier, are protective, motivated by a fear of harm at the hands of others. Some are acts of aggression, motivated by a wish to damage others. Some are intended to maximize advantage in competitive situations. Some lies are acts of flattery and some are intended to make the speaker appear more interesting than he or she in fact is. Some lies are told by office-holders from devotion to what is taken to be the public interest and some are told both to and by office-holders to subvert that interest. But in each of these different types of case, if a lie has been successful, it may well be that the liar will have altered the relationships of power in her or his own favor, or, perhaps, on occasion in favor of someone else. Yet in so doing, whether the lie is successful or not, the liar will also have altered her or his relationship to others in general, by deliberately violating the norm presupposed in all human relationships involving assertive speech-acts. She or he will have relied upon the general human regard for truth, while failing to have regard for it. "Without truth," Kant wrote, "social intercourse and conversation become valueless" (*Eine Vorlesung Kant's fiber Ethik*, ed. P. Menzer, p. 285, trans. L. Infield, *Lectures on Ethics [Indianapolis: Hackett, 1980], p. 224*). **And the offense of the liar, thus understood, is not a matter of the harmful consequences of particular lies. To tell a lie is wrong as such, just because it is a flouting of truth, and it is an offense primarily not against those particular others to whom this particular lie has been told, but against human rationality, everyone's rationality, including the liar's own rationality. By lying she or he has**

failed not only to acknowledge truth as a good that is indispensable in rational relationships with others, but also to recognize that a failure to respect truth is a failure in respecting oneself as a rational being.

This conception of the wrongness of lying was elaborated within a moral tradition whose central theses were in crucial respects at odds with those of the tradition that I described earlier. For where the exponents of that tradition, from Clement to John Stuart Mill, had agreed on the need to exempt certain types of lie from the general prohibition of lying, the adherents of the tradition of which Kant was a late and distinguished member agreed in insisting that the rule prohibiting lying was exceptionless. Instead of looking back to Plato, its protagonists look back to Aristotle's condemnation of all lying as disgraceful and to his praise of the lover of truth who is truthful whether something further is at stake or not (*Nicomachean Ethics* IV, 1127b4-8). There are trenchant restatements of this standpoint by St. Augustine, by St. Thomas Aquinas, by the Catechism of the Council of Trent, by Pascal, and by Protestant theologians both before and after Kant. Augustine declared in the *Contra Mendacium* (31C) that "it is said to God 'Your law is truth.' And for this reason what is contrary to truth cannot be just. But who doubts that every lie is contrary to truth? Therefore no lie can be just." Aquinas argued that truth itself is a virtue, since to say what is true makes a good act and a virtue is that which makes its possessor good and renders its possessor's action good (*Summa Theologiae* IIa-IIae, 109, 1). Of the vices opposed to the virtue of truth lying is the first (110, prologue). Aquinas captured a thought central to this tradition when he distinguished between the wrong done by intentionally asserting what is false and the wrong done by intentionally deceiving someone by that false assertion. Even without an intention to deceive, the intentional assertion of what is false is wrong (110, 1 resp. and 3 ad. 6). The offense is against truth.

Some adherents of these two contrasting and generally rival traditions may in fact disagree about very little of moral substance.

For among some of those for whom lying is altogether prohibited, the definition of a lie is such as to exclude just those cases that some adherents of the other tradition treat as permissible or required lies. But it would be a mistake to conclude from these cases that the differences between the two traditions are unimportant, as Newman seems to have done. Those differences extend to three kinds of issue.

First there is the question of how a lie is to be defined. Those for whom some types of lie are permissible or even required characteristically define a lie so that an intention to deceive is an essential defining property of a lie, and the wrongness of lies is the same as that of other acts of deception, while those for whom no lies are permissible characteristically define a lie in terms of an intention to assert what is false, sometimes, like Aquinas, denying that an intention to deceive is necessary for an assertion to be a lie. A second difference concerns the nature of the offense committed by a liar. For those for whom some types of lie are permissible or even required the wrong done by a lie is understood in terms of the harm inflicted upon those social relationships that need to be sustained by mutual truth and credibility. Because of the constitutive part played by such trust in every important human relationship, that harm is never held to be entirely negligible. But evidently there are occasions on which the utterance of a particular lie will prevent some harm greater than that which its telling will cause to the social fabric. By contrast, for those for whom no lie is permissible the wrong committed by making a false assertion is understood as a type of wrong that inescapably puts in question one's standing as a rational person in relationship to other rational persons.

A third set of issues concerns the kind of justificatory argument advanced within each tradition. Those who hold that some types of lie are permissible advance justifications that cite the effects of different types of lie, even when those who advance them are not consequentialists in general. Those who hold that all lies are forbidden advance justifications citing the nature of the act of lying. And at

this point the self-definition of each of these two rival traditions makes something plain that has been insufficiently remarked within either tradition. There are, so I argued, two distinct grounds for our concerns about truth-telling and lying: one deriving from the invariant semantic rule governing the utterance of assertions and one from our varying evaluations of the motives for and the effects of the utterance of different types of lie. Reflection upon the first of these focuses attention upon lying as an offense against truth, as an error-engendering misuse of assertion, while reflection upon the second focuses attention upon lying as an offense against credibility and trust, as having effects that tend to be destructive of relationships between persons. And each of the two rival moral traditions that I identified has developed a line of argument well designed to uphold the claims upon our allegiance of its formulation of what it takes to be *the* moral rule concerning truth-telling and lying.

In this case at least two moral traditions seem to be one too many. In answer to such questions as “What should be our socially established rule about truth-telling and lying?” “What should we teach our children?” “And how should we justify rationally what we teach them?” we are presented with two incompatible and rival **types** of rule and two incompatible and rival types of justificatory argument. At the same time we cannot but recognize the compelling and insightful character of central considerations advanced from each side. The problem is therefore not simply that of finding sufficient reasons for choosing to align ourselves with one standpoint or the other. It is rather that we need, if at all possible, to find some rationally justifiable framework within which the concerns articulated within both traditions can be integrated in such a way as to provide a single set of answers to those questions.

This then, in outline at least, is the problem. In what direction should we turn in search of a solution? One obvious suggestion would be first to examine the practice of one or more other cultures with a somewhat different moral tradition concerning truthfulness: for example,

Confucianism with its conception of appropriate speech and of the virtue of *hsin*. And in the larger enquiry of which these lectures are a part this will be a necessary undertaking. But an important preliminary is to understand a good deal better just what it is that we ourselves need here and now and why. What is the moral condition of the culture now dominant in North America in respect of truth-telling and lying?

II

Three features of that culture are relevant and notable: the nature and extent of disagreement about what the rule concerning lying should be, the frequency of lying of a variety of kinds, and the nature of the underlying dilemmas that make that disagreement and that frequency intelligible, at least in part. Consider each of these in turn.

Discussion, sometimes in depth, with a number of different American groups in the last ten years has convinced me that the only shared near universal agreement is on the form that any acceptable rule concerning lying and truth telling should take. That form is “Never tell a lie” — this part of the rule is generally enunciated firmly and clearly, especially to children — “except when” — here the voice begins to drop — and there then follows a list of types of exception, culminating with an “etc.” That list includes most often “when by lying one will save an innocent human life,” almost as often “when by lying one will avoid offending someone,” and quite tolerably often “when by lying one will secure advantage in one’s career or to one’s financial prospects.” At one end of a spectrum there are those Americans who hold that one ought *never* to tell a lie; at the other those who regard themselves as free to misrepresent their own past or the truth about others in trivial anecdotal gossip as readily as on occasions when something important is at stake.

We have then a first set of wide-ranging disagreements not only about what excepting clauses should be included in the list, but also about how these should be understood to apply. To what classes of person may we avoid giving offense by telling a lie? Is untruthful gossip only permissible when it could not damage anyone, or are there people whose reputations need not matter to us? May I secure advantage to my career only by lying about what I, but not others, take to be irrelevant considerations or may I misrepresent what everyone would agree to be relevant? A range of different answers to questions such as these is expressed not only by what people say about lying, but also by how and when they lie. That this is so makes the facts about the incidence of certain types of lying a little less surprising than they might otherwise be.

What then are those facts? Bella DePaulo, a University of Virginia psychologist who studied lying by having her subjects keep a diary recording the lies that they told, concluded from her study that "People tell about two lies a day, or at least that is how many they will admit to" (*New York Times*, February 12, 1985, p. 17). James Patterson and Peter Kim, whose expertise is in research for advertising, reported in 1991 that 91 percent of Americans lie regularly, that only 45 percent refrain from lying on occasion because they think it wrong, and that those who do lie lie most to friends and relatives (*The Day America Told the Truth* [New York: Prentice-Hall]). They also found that a distinction was made between more and less serious lies and that 36 percent admitted to serious lies. Dan McCabe of Rutgers University found that 57 percent of business students would admit to having cheated on an examination at least once (*Harpers Index*, September 1991), while in an earlier *Psychology Today* study the percentage of students who admitted to being willing to cheat on examinations or other test assignments, if they judged that they could get away with it, was 67 percent (James Hassett, *Psychology Today*, November 1981).

Unsurprisingly, those who lie commonly also believe that others lie to them. So Patterson and Kim found that 31 percent of their subjects

believed that they had at some time been lied to by their physicians, 34 percent believed this of their accountants, and 42 percent of their lawyers. American lawyers are of course professionally divided about lying; some have held that a defense lawyer who knows that a client is committing perjury in court has a duty to use that perjury to secure acquittal, if she or he can; others have denied that this is so. And such divisions occur in a number of professions. But the division both in private life and in the professions is not just one between different individuals. It is also one within many individuals. The extent to which it is within and not only between individuals can be gauged by the extent of the unhappiness about their own lying that significant proportions of those who nonetheless regularly lie evince. They evince that unhappiness in a variety of ways. To a significant extent they report that they feel uncomfortable when they lie. They betray their anxiety, when they are put to the question about their lies, by systematically failing polygraph tests, in this being quite unlike those Eastern Europeans cited by Richard Helms, “who could defeat the polygraph at any time,” because they had spent their lives “lying about one thing or another and therefore become so good at it” (investigation of the Assassination of President John F. Kennedy, vol. 4, pp. 98-99, 118, cited in John Ranelagh, *The Agency: The Rise and Decline of the C.I.A.* [New York: Simon and Schuster, 1986], pp. 568-69). These are the same people described by Erazim Kohak as having developed under Communist regimes an inability to admit to the differences between illusion and reality. “A factory manager, seeing the collapse around him, yet reporting inflated production figures to assure premiums for his factory, could not believe, but neither could he just lie. Instead he would refuse to acknowledge the distinction.” And so after communism this refusal persists. “Though there is no one to deceive, deception has become a habit” (“Ashes, Ashes . . . Central Europe after Forty Years,” *Daedalus* 121, no. 2 [1992], 203; for systematic understanding of the function of lying in the Soviet Union itself, the indispensable works are by Alexander Zinoviev, both the novel *Yawning Heights* [New York:

Random House, 1979) and *Homo Sovieticus* [London: Gollancz, 1985]). But this is not at all how contemporary Americans are.

They seem to recognize what they are doing, while lying, and are often far from satisfied with their own justifications for lying. This unhappiness is perhaps one cause of those oscillations and inconsistencies in responding to discovered lies that mark so much of American life, directing our attention to further dimensions of those divisions about lying, both between groups and individuals and within groups and individuals, on which I have already remarked. Those oscillations and inconsistencies are most obvious in political life. The lies of Richard Nixon and Oliver North incurred instant and extreme obloquy, the lies of a Lyndon Johnson about Vietnam or of a James Baker about relationships with the government of China much less (on Lyndon Johnson, see the *Chicago Tribune*, October 20, 1991, p. 4-i; on James Baker, see Hodding Carter III, "Viewpoint," *Wall Street Journal*, January 25, 1990, p. A15). Arthur Schlesinger, Jr., who had proposed to the Kennedy administration that "lies should be told by subordinate officials," so that they and not the president would take the blame, if discovered (Peter Wyden, *Bay of Pigs: The Untold Story* [New York: Simon and Schuster, 1979], p. 161), has since been among the most vehement denouncers of lies told by subordinate officials to protect presidents. And public blame for lying is in general unevenly and haphazardly distributed. What does such unevenness and inconsistency reveal concerning the range of disagreement about lying over and above disagreements as to what types of items should be excepted from the general prohibition?

They are of two related kinds. There is first a set of disagreements about which types of lie are to be treated as more serious and which as less serious offenses, and within each category how different types of lie are to be ranked. If I lie to the police about the whereabouts of my friend, who has fled from the scene of an unreported automobile accident, is this better or worse than lying to my friend about my part

in wrecking his car? If I lie to my wife about having lost my job, is this better or worse than lying to my employer in order to keep that job? Yet it is not only that we do not agree on the gravity of the offenses committed by different kinds of liar. It is also that we do not agree upon how to respond to different kinds of lie, when someone's lies are discovered and we are the offended party. If a lie concerns some relatively trivial matter, should we just ignore it or is this to treat lying as acceptable? If a lie is a serious breach of trust, should we break off all relationships with the liar? Ought we to make the fact of such lying public in order to warn others? Should a lie of a certain gravity disqualify a liar from public office or from friendship? And if we ourselves are discovered in a lie, what do we have to do to merit forgiveness? There seems to be no consensus on how these questions are to be answered.

Not all North Americans belong to the dominant culture that is in such a peculiar condition in respect to lying and truth-telling. Orthodox Jews, conservative Roman Catholics, some Southern Baptists, and devoted Confucian Chinese families provide examples of minorities that advance systematic and unambiguous answers to these and to kindred questions. But outside such minorities — minorities that are deviant with respect to the dominant North American culture, but nondeviant with respect to the larger history of humankind — the lack of consensus upon these issues is a sign of a remarkable absence. The dominant culture fails to provide any generally accepted and agreed-upon public rule about truth-telling and lying, by appeal to which we could in relevant instances call each other to account. Why is this so? What do we need to understand about North Americans belonging to the dominant culture, if this absence and the divisions and disagreements that accompany it are to be intelligible?

The salient moral fact about such modern Americans is, so I want to suggest, this. They are brought up to give their allegiance to two distinct sets of norms. One of these enjoins each individual to pursue her or his own happiness, to learn how to be successful in competing

against others for position, power, and affluence, to consume and to enjoy consumption, and to resist any invasion of her or his rights. The other set instructs individuals to have regard for the welfare of others and for the general good, to respect the rights of others, to meet the needs of those who are especially deprived, and even to be prepared on some particular occasions to sacrifice one's own immediate happiness for the sake of the happiness of particular others. On many occasions of course these two sets of norms are not in conflict. But on others, and some of those among the more significant in individual lives, Americans not only discover that such norms make rival and incompatible demands for their allegiance, but they also find that they possess no third, higher-order set of norms that would enable them to make a rationally justifiable choice between those conflicting demands.

This moral situation is not of course confined to North America. It characterizes in varying degree all the cultures of advanced modernity. It was first articulated in philosophical terms in the late nineteenth century by Henry Sidgwick in *The Methods of Ethics*, a text that in its foreshadowing of the subsequent history both of morality and of moral philosophy deserves to be accorded the status of a prophetic book. Sidgwick had taken it to be a discovery of that distinctively modern moral philosophy that first emerged in seventeenth- and eighteenth-century England that there is not one single governing authority in moral matters, the role to which "Reason" is assigned in most Greek moral philosophy, but two distinct authorities, "Universal Reason and Egoistic Reason" (*Outlines of the History of Ethics for English Readers* [London: Macmillan, 1886], p. 198). The first of these prescribes how it is reasonable to act if the general good and happiness is to be achieved, the second how it is reasonable to act if my own good and happiness is to be achieved. Sidgwick took it to be his own philosophical discovery, after an extended study of the claims of Kantian, utilitarian, and intuitionist moral philosophy, that when the injunctions of these two kinds of practical reason conflict, there is no rational method for deciding

between their claims or for reconciling them (“Concluding Chapter” of *The Methods of Ethics*, 7th ed. [London: Macmillan, 1907]).

Sidgwick’s own treatment of what he spoke of as the duty to veracity consists chiefly of an examination of those convictions that belong to what he took to be “the morality of Common Sense” (*The Methods of Ethics III*, chapter 7). About veracity he concluded that among persons of common sense “there is no real agreement as to how far we are bound to impart true beliefs to others” (p. 317), perhaps because such persons seem unable “to decide clearly whether truth-speaking is absolutely a duty, needing no further justification: or whether it is merely a general right of each man to have truth spoken to him by his fellows, which right however may be forfeited or suspended under certain circumstances” (p. 315). Summarizing common-sense beliefs about truthfulness, Sidgwick declares that it is commonly held that lawyers may be justified in saying what they know to be false, if so instructed by their clients, that it is held by most persons that benevolently intended lies to invalids are justifiable, and, perhaps more surprisingly, that no one “shrinks from telling fictions to children on matters upon which it is thought well that they should not know the truth” (p. 316).

Common sense offers us no principle by which we may decide systematically in these or other cases. We have no alternative to “weighing the gain of any particular deception against the imperilment of mutual confidence involved in all violation of truth” (p. 316). The metaphor of weighing invites Sidgwick’s readers to ask: what are the scales? And it turns out that, for the reasons that I have already cited, Sidgwick can in the end only offer us two alternative sets of scales, which will provide us with different measures of weight, that of Universal Reason, appealing impersonally to the standard of the greatest happiness of the greatest number, and that of Egoistic Reason, by whose standard my happiness outweighs that of everyone else. Beyond these there is no third and higher

standard of practical reason to decide on each particular occasion which of these two rivals it is to whose verdict we should attend.

Sidgwick's philosophical analysis confirms what the reports by sociologists and social psychologists on contemporary North American moral culture already suggested, that no formulation of a rule concerning truth-telling and lying and no account of the virtue of truthfulness will meet our contemporary needs, unless they overcome that moral dualism that seems* to debar so many from the possibility of ordering within a single rational scheme their selfregarding reasons for action and those reasons that have regard either for particular others or for the general good. So it is not just that we need to integrate the insights and concerns of the two rival moral traditions concerning truth-telling and lying. We have to impose a further condition, that this integration provide a rational ordering of the relevant types of reason for action. The satisfaction of these two major conditions requires of course more and other than the provision of a more adequate philosophical theory. What is needed is the identification of some mode of institutionalized social practice within which generally established norms and reflective habits of judgment and action could sustain a coherent and rationally justifiable allegiance to a rule concerning truth-telling and lying in a way and to a degree very different from the present dominant culture. And this is a large undertaking. But a more adequate philosophical theory would be at least a first step. How then should we proceed in attempting to develop such a theory?

We might begin by asking whether there is not more for us to learn from the most distinguished modern philosophical representatives of the two rival traditions, J. S. Mill and Immanuel Kant, than Sidgwick supposed. Sidgwick after all concerned himself with lying and truth-telling only incidentally and his treatment of both Kant and Mill was restricted in scope. We not only have the benefit of what can be learned from later interpreters and more adequate editions, but we are able to bring to our reading of Kant and Mill

questions that go beyond Sidgwick's, in part because of what we have learned from Sidgwick. So in order to move forward, we should first turn back, noting as we do not only that truthfulness was a topic of continuing philosophical concern for both Kant and Mill and but also that both Kant and Mill cared deeply about truthfulness. I might have begun this enquiry with either thinker, but Mill is perhaps somewhat closer to us, not just chronologically but in his hopes and fears for the culture. So it is to Mill that I turn first.

III

In the second chapter of *Utilitarianism* Mill attempted to dispel misunderstandings of the Greatest Happiness principle by defending it against a variety of accusations. Against the accusation that utilitarianism reduces morality to expediency Mill set out his account of truthfulness, arguing that

inasmuch as the cultivation in ourselves of a sensitive feeling on the subject of veracity is one of the most useful, and the enfeeblement of that feeling one of the most hurtful, things to which our conduct can be instrumental; and inasmuch as any, even unintentional, deviation from truth, does that much towards weakening the trustworthiness of human assertion, which is not only the principal support of all present social well-being, but the insufficiency of which does more than any one thing that can be named to keep back civilization, virtue, everything on which human happiness on the largest scale depends; we feel that the violation, for a present advantage, of a rule of such transcendent expediency, is not expedient, and that he who for the sake of a convenience to himself or to some other individual, does what depends on him to deprive mankind of the good, and inflict upon them the evil, involved in the greater or less reliance which they

can place in each other's word, acts the part of one of their worst enemies. Yet that even this rule, sacred as it is, admits of possible exceptions, is acknowledged by all moralists; the chief of which is when the withholding of some fact (as of information from a malefactor, or of bad news from a person dangerously ill) would save an individual (especially an individual other than oneself) from **great** and unmerited evil, and when the withholding can only be effected by denial..

This is in some respects a very plain statement. Mill is evidently a rule-utilitarian, prepared to allow only a very few types of exception to the prohibition of lying. He mentions only one such and he is careful to affirm a stringent prohibition on all merely convenient lies. And certainly if contemporary Americans were systematically to obey Mill's rule, ours would be a very different society. Mill did elsewhere consider the type of case in which the cost to some individual of telling the truth on a matter in which it is important not to lie is serious, perhaps mortal danger to herself or himself, and asserted that no general rule governs such cases, independently of circumstances (letter of February 9, 1867, to Henry S. Brandreth, in *Collected Works of John Stuart Mill, vol. 16: The Later Letters of John Stuart Mill 1849-1873*, ed. F. E. Mineka [Toronto: University of Toronto Press, 1972], p. 1234). But the tone as well as the content of all Mill's remarks about lying place him, not too surprisingly, particularly if we remember how influenced he was by Coleridge, in the same moral tradition as Milton and Dr. Johnson.

As to the logical structure of the justification of the rule that Mill formulates, matters at first sight appear equally straightforward. The premises are: first, that lying always weakens to some greater or lesser extent trustworthiness: second, that trustworthiness is the indispensable support of that upon which "present well-being" and "civilization" and human happiness in general depend; and, third, that right action is action that promotes the general happiness, the greatest happiness of the greatest number. Therefore lying is

(almost always) wrong. But questions arise about what Mill meant in affirming the second and third premises of this argument.

When Mill asserted in support of the second premise that both “present social well-being” and “civilization” depend on trustworthiness, he might be thought by a casual reader to be advancing no more than a strongly worded version of a commonly reiterated warning that lying undermines credibility and that credibility is needed to sustain the social fabric. Yet experience goes to show that the social fabric generally survives a good deal more lying than Mill would have allowed. As Harry Frankfurt has remarked, “The actual quantity of lying is enormous after all, and yet social life goes on. That people often lie hardly renders it impossible to benefit from being with them. It only means that we have to be careful” (“The Faintest Passion,” Presidential Address to the Eastern Division of the American Philosophical Association, 1991, Proceedings and Addresses of the A.P.A. 66, no. 3 [November 1992], 6). So that if this is all that Mill meant, his second premise is false and his argument fails. But this is not what Mill meant. For, when Mill used the word “civilization,” he did not use it lightly. The words Mill uses when he speaks elsewhere of those whom he took to be uncivilized are “barbarians” or “savages,” and barbarians need the rule of a benevolent despot, not the doctrines of *On Liberty* (*On Liberty*, chapter 1) or the moral rules that are the counterparts of those doctrines. Among those not yet civilized Mill took lying to be endemic. In the essay “On Nature” (*Collected Works*, vol. 10: *Essays on Ethics, Religion and Society*, ed. J. M. Robson [Toronto: University of Toronto Press, 1969])

Mill considered whether it was right to think of truthfulness as natural to human beings, since “in the absence of motives to the contrary, speech usually conforms to, or at least does not intentionally deviate from, fact,” but against this he cites what he takes to be the case, that “savages are always liars” (p. 395). Moreover, the same holds of the inhabitants of “the whole East and the greater part of Europe” and even in England

it is only a small minority — “the higher classes,” as he says elsewhere — who make it a point of honor to respect truth for truth’s sake.

Habitual lying is, Mill believed, a consequence of “the natural state of those who were both uneducated and subjected.” It is “a vice of slaves.” (For one source of Mill’s beliefs on this matter, see James Mill, *The History of British India*, 4th ed. [London: J. Madden, 1848], Book II, chapter 7, p. 467: “The Hindus are full of dissimulation and falsehood, the universal concomitants of oppression”). And it was, on his view, greatly to the credit of the contemporary English working class that, although they lied, they were ashamed of it (speech of July 8, 1865, during the Westminster Election, in *Collected Works, vol. 28: Public and Parliamentary Speeches*, ed. J. M. Robson and B. L. Kinzer [Toronto:

University of Toronto Press, 1988), pp. 35-36). A central political and educational problem then is that of how to transform those hitherto uneducated and subjected into the condition of that minority that does already respect truth for truth’s sake. For a repudiation of lying is, on this view, an inseparable part of the rise of any social group from a condition of subjection and lack of education to one of liberty and a cultivated intelligence, both of them necessary for happiness. When Mill speaks approvingly of those who respect truth for truth’s sake, he is of course not contrasting them with those who respect truth for the sake of their own or general happiness. It is true that only happiness is, on Mill’s view, desired for its own sake, but virtue is desired for its own sake precisely because it is, or rather has become, a part of happiness (Utilitarianism, chapter 4). Virtue is originally valued only as a means, but then, as a result of experience of the life of virtue, it comes to be valued also as an end. We may therefore safely infer that truthfulness, as a virtue, is itself, on Mill’s view, originally valued only as a means, but then also as an end. And the life of civilization is a life in which truthfulness has come to be so valued. So that when Mill, in the second premise of his argument in Utilitarianism, claims that a trustworthiness uncorrupted by lying is indispensable

not just for happiness and well-being, but for those conjoined with civilization, his use of the word “civilization” should convey to us a conception of the general happiness to be aimed at in England in the mid-nineteenth century, one that is not adequately communicated by the philosophical treatment of happiness in Utilitarianism.

What then is an adequate conception of happiness — I mean not in the abstract and general terms of Utilitarianism, but in terms of those political, social, and personal goals that Mill set for himself and for others in his own time and place? And how, on Mill’s view, can we come to have such a conception and communicate it effectively to others? Mill’s answer to this second question was that such a conception could be acquired only by extended intellectual, moral, and emotional enquiry and education. Such enquiry and education involves continuous conversation and debate with others, debate of a kind in which Mill himself had participated, both within utilitarian circles and in controversies between utilitarians and their critics. Exclusion from such debate is deeply injurious to moral education and “participation in political business” is “one of the means of national education,” helping to draw human beings out of “the narrow bounds of individual and family selfishness” that otherwise make them stupid, ill-informed, and selfish (“Thoughts on Parliamentary Reform,” 1859, in *Collected Works, vol. 19: Essays on Politics and Society*, ed. J. M. Robson [Toronto: University of Toronto Press, 1977], p. 322).

How is that education to be contrived? Mill took himself to have learned from Coleridge the importance of providing state support for an educated class, one that would in each locality provide moral and intellectual leadership and instruction (“Coleridge,” *London and Westminster Review* [1840]). Such an educated class, so Mill argued, had to have a special place in and influence upon both public debate and the activities of government, for one person *is not* as good as another” (“Thoughts on Parliamentary Reform,” p. 323). But our constitutional and electoral arrangements, while securing the influence of the better

educated, ought to be such that they become a means for general moral education, in order to remedy “the mental and moral condition of the English working classes” (p. 327). Hence Mill’s disapproval of the secret ballot, which he took to promote a cowardly concealment of one’s true views, and which he thought able to produce its intended effect “only at the cost of much lying” (p. 337). It is then one of the tasks of moral education to construct forms of institutional debate in and through which, among other things, those who participate in them can be sustained in their truth-telling and transformed, if need be, from liars into truthful persons. Exclusion of those not yet thus educated from processes of political debate and decision debars them damagingly — damagingly for others as well as for themselves — from such education, but inclusion in those processes of debate must be such that they learn from those better educated. And the better educated themselves still need to learn from such debate. For those who do not participate in debate can only have untested opinions, whether about happiness or anything else, and not genuine knowledge.

That this is so was made clear by Mill in *On Liberty*, where he asserts that “no opinion deserves the name of knowledge” that has not emerged from “an active controversy with opponents” and where he treats the Socratic mode of dialectic and even the medieval disputation as models for a type of institutionalized controversy much needed in his own time, but no longer provided. Without such controversy there can therefore be no knowledge concerning that happiness that is the end of right action. Infringements of liberty of thought and discussion are to be condemned precisely because liberty is necessary, if such forms of debate are to arrive at truth. But debate will also presumably require protection from violation by those types of act that Mill takes to be “fit objects of moral reprobation, and in grave cases, of moral retribution and punishment,” a class that includes acts of “falsehood or duplicity” in dealing with others. So the rule requiring truthfulness will be among those rules to which conformity is necessary as a means for securing the kind of controversy in debate and enquiry from which

there can emerge a true and adequate account of human happiness as an end and of the part to be played by truthfulness in any life answering to that account. What Mill called the “trustworthiness of human assertion” will have to be, on his view, if I have construed it rightly, first recognized as a necessary means to, and then as an essential constituent of, both my own happiness and the general happiness.

What I have identified as the second premise in Mill’s argument for the justification of lying is then something more and other than a general claim that the social fabric is somehow endangered by lying. It is the much stronger, and also the much more interesting, claim that what Mill meant by civilization, a type of social order constituted as a project of moral education through political and moral conversation and debate, requires a stringent and very widely respected rule prohibiting (almost all) lying. A civilized social order is one collectively and cooperatively concerned to understand the truth about human beings and nature, and the violation of truthfulness is injurious to the project of such a social order for the same reason and in the same way that a violation of truthfulness in reporting data is injurious to the sciences. Truthfulness in both cases is not just a useful and necessary means to, but is constitutive of the ends pursued.

In saying this I may have gone a little, although only a little, beyond what Mill himself actually asserts. But, if this is the direction in which Mill’s argument points us, we need to go even further. Mill in his statement of the rule about lying in *Utilitarianism*

identified lying as an offense against trustworthiness. But the argument that I have developed out of his writings requires us not only to identify it also as an offense against truth, but also to understand the relationship between these two aspects of truthfulness in a particular way. It is not trustworthiness in general that is crucial to our well-being as actual or aspiring members of a civilized social order, characterized as Mill characterized it, but the peculiar kind

of trustworthiness that is required of those who are participants in a particular kind of social enterprise, who are collectively and cooperatively engaged in seeking through shared enquiry the truth about their present condition and their future good, as an essential part of the project of moving from their present condition toward the achievement of that good. Truth is *the* good internal to rational enquiry and the kind of trustworthiness required from each other by those who participate in enquiry includes an unfailing regard for truth and for truthfulness. So it is with those who are engaged cooperatively in the investigations of the natural sciences or the researches of historians or anthropologists. And insofar as the moral life is a life of communal enquiry — to say this is not to deny that it is also a number of other things — the kind of trust that those who engage in it have to repose in each other must therefore include mutual trust in respect of a shared regard for a norm of truth that has to be exceptionless, for the same reasons that the norm governing truth-telling in scientific and other research communities has to be exceptionless. But in reaching this conclusion I have, by following a line of argument developed by Mill, arrived at conclusions that are obviously at odds with Mill's own.

In the passage from Utilitarianism from which I began Mill identified at least two kinds of exception to the rule prohibiting lying, and he justified those exceptions by suggesting that on certain types of occasions the consequences of telling particular lies for the happiness or unhappiness of particular individuals were such as to outweigh any detriment to the general good. But how can this be reconciled with the claims that I have just made for an exceptionless rule, one necessary for us to arrive at an adequate conception of happiness? A first response may well be that it cannot be so reconciled, and that, if the line of argument that I have developed out of certain of Mill's texts is really there, then there are to be found in Mill strains of thought that are in serious tension with each other, something that a number of commentators have discerned. But a second response might run as follows.

Of the two kinds of exception allowed by Mill in Utilitarianism one is a matter of the withholding of information from those who would be harmed by it, the other of the prevention of serious harm intended by malefactors. About the former we should note that there are ways of withholding information other than lying and that, if we take systematic precautions in advance, as it is our duty to do, lying generally becomes unnecessary. If it does seem to have become necessary, this is perhaps to be taken as evidence of our own or someone else's lack of wit, ingenuity, and foresight, itself an important kind of moral failure. So we can perhaps agree with Mill about the need on rare types of occasions to withhold information, without agreeing that this provides any good reason for rejecting the authority of an exceptionless rule. Moreover, we thereby signify that those whom we are thus protecting, whoever they may be, still remain our partners in the enterprises of the moral life, and therefore persons to whom we may not tell lies. The symbolic importance of upholding this rule universally without exceptions as to persons is not to be underestimated.

What then are we to say about the other class of exception, the type of lie told in order to avert grave harm intended by malefactors? The exceptionless rule requiring truthfulness, just because the moral life is one for which truth is a supreme value, binds the members of the moral community in general as rational persons, just as the analogous rule binds the members of the scientific community in particular. It is a norm defining the relationship of the members of those types of community with each other. But what if someone constitutes herself or himself a deliberate enemy of moral community and not just of particular persons, as someone, for example, bent on murder does? In such situations does the same rule bind us? If so, why? If not, why not? These questions were already raised for us by Samuel Johnson. But the most important, as well as the most notorious, discussion of how to answer them is of course by Kant.

IV

At first sight and on a conventional reading no two moral philosophers are more sharply at odds concerning truth-telling and lying than are Mill and Kant. Mill held that some lies are not only morally permitted, but morally required, while Kant held that all lying is prohibited. In Utilitarianism at least Mill's justifications, both of his formulation of the rule generally prohibiting lying and of his statement of the types of exception to that rule, are consequentialist, while Kant rejects consequentialist justifications and grounds the rule prohibiting lying in the rational nature of human beings. But perhaps this opposition is not as unqualified as conventional readings have made it. I have already suggested that, when Mill reflected on the requirements that must be met, if political and social relationships were to become rational, he moved much closer to an unqualified condemnation of untruthfulness than, on a conventional reading, we might have expected. And, since Mill's concerns about rationality bring him very close to what were also central concerns of Kant, it is worth asking whether there may not be respects in which their undeniably antagonistic views may nonetheless be understood as contributing to a common enterprise. Yet if we are to do so in a way that also does justice to their disagreements, we should begin our discussion of Kant in those areas in which that difference is most evident.

I have distinguished two rival moral traditions with respect to truth-telling and lying, one for which a lie is primarily an offense against trust and one for which it is primarily an offense against truth. For adherents of the former tradition unjustified deception is what offends against trust and unjustified lies are a species of unjustified deception. For such persons it therefore generally makes no significant moral difference whether or not a deception is carried out by means of a lie or otherwise. If it is a justified deception, then

that it was carried out by lying will not make it any less justified. If it was an unjustified deception, it will be none the worse for having been carried out by a lie. But for adherents of the rival tradition no lie can ever be justified, although some deceptions may be.

Hence the importance within this rival tradition of anecdotal teaching about the moral praiseworthiness of the ingenuity of those who succeed in some justified act of deception without committing the wrong of lying. A signal example is that of St. Athanasius. Persecutors, dispatched by the emperor Julian, were pursuing him up the Nile. They came on him traveling downstream, failed to recognize him, and enquired of him: "Is Athanasius close at hand?" He replied: "He is not far from here." The persecutors hurried on and Athanasius thus successfully evaded them without telling a lie (see F. A. M. Forbes, *St. Athanasius* [London: R. and T. Washbourne, 1919], p. 102). Whether one thinks this a pointless anecdote or not reveals something fundamental about one's attitude to lying. Kant's attitude appears in an anecdote that he told about himself.

When in 1794 Kant was required by King Friedrich Wilhelm II, shortly before the latter's death, to refrain from any distortion or depreciation of Christianity, he knew that if he made public anything further of his thoughts on religion, as he had hoped to do, he would be held guilty of just such distortion or depreciation, perhaps with baneful consequences. He therefore responded by making a declaration "as your Majesty's faithful subject, **that I shall in future completely desist from all public lectures or papers concerning religion, be it natural or revealed.**" The Prussian censors and, if it was reported to him, the king himself would have understood Kant to be saying that he would never so publish. But that is not of course what Kant had in fact declared. As he later pointed out, his pledge to desist was made only "as your Majesty's faithful subject," a status that Kant would lose when this particular king died. "This phrase," wrote Kant in recounting the story (in the preface to *The*

Quarrel between the Faculties), after the king's death in 1797, ". . . was chosen by me most carefully, so that I should not be deprived of my freedom... *forever*, but only so long as His Majesty was alive" and Kant knew that the death of Friedrich Wilhelm II was expected imminently. So Kant succeeded in misleading the Prussian censors without lying, something that he thought it morally important to do.

Kant therefore places himself among those who hold that my duty is to assert only what is true and that the mistaken inferences that others may draw from what I say or what I do are, in some cases at least, not my responsibility, but theirs. Those others, if they discover that, in such cases, what I said or did was well designed to mislead, as it was in Kant's own case, will certainly in the future treat me, and possibly others, as less trustworthy. But it is not this possible consequence of injury to trust that matters; what matters is the avoidance of the assertion of falsity.

In what then does the wrongness of the intentional assertion of what is false consist? I have claimed that what has been fundamental for those who have understood lying as an offense against truth is the semantic rule requiring the assertion only of what is true; the need for conformity to this rule is learned by everyone who learns a natural language. The fact that all language-users in the vast majority of instances cannot but conform to this universal rule, and cannot but interpret others as conforming to it, is what makes effective lying possible. A liar therefore deliberately violates that rule, while at the same time willing that others should unsuspectingly adhere to it. And so no liar can coherently will that the maxim upon which she or he acts should be universally acted upon by others. It is thus at first sight a very short step — almost no step at all — from the semantic rule to Kant's first formulation of the categorical imperative: "Act only according to that maxim whereby you can at the same time will that it should become a universal law."

We may easily be tempted by this to suppose that it is because universalizability of the maxim determining the liar's action, thus

understood, leads the liar on toward self-contradiction that lying is prohibited for any rational person. But this would be a mistake. It cannot be universalizability as such or by itself that is sufficient for the prohibition of lying. Why not? Consider two important types of case. The first is that of someone who has judged on empirical grounds that social life is, one way or another, a war of each against all, who takes pride in her or his own craft in using force and fraud and whose determining maxim for many actions is “Let each exert herself or himself to overcome others, by whatever means are available, including lying, and may the strongest win!” The second is of a person whose empirical judgments about social life and about her or his own capacities are the same, but whose determining maxim is “Let all who are strong take pride in refusing to do anything as mean-spirited as lying in their war against others, let the weak do as they wish, and, if those who are both strong and truthful go down to defeat, so be it!”

The first of these two persons — and I have known both of them — is on occasion a liar, the second always truthful, and both are able to act according to maxims that they are prepared to universalize and are able to universalize without any incoherence. But we would of course be in error if we were to suppose that they provide counterexamples to Kant’s thesis. For their maxims fail to be genuinely Kantian maxims in at least two respects. First, their maxims embody what their authors take to be lessons, both about social life and about themselves, that had to be learned empirically.

But Kant held that the prohibition on lying could not be such. In the “Fragments of a Moral Catechism” Kant put into the mouth of the teacher the words: “The rule and direction for knowing how you go about sharing happiness, without also becoming unworthy of it, lies entirely in your reason. This amounts to saying that you do not have to learn this rule of conduct by experience or from other people’s instruction; your own reason teaches and even tells you what you have to do” (*Metaphysic of Morals*, “Methodology of Ethics,” section 1,

481). Kant then chooses as his illustration for this point the prohibition against lying in a situation “in which you can get yourself or a friend a great advantage by an artfully thought out lie (and without hurting anybody else either)” and he speaks of the unconditional constraint of this prohibition as “this necessity, laid upon a human being directly by her or his reason . . . The two nonKantian maxims that I have described, by reason of their empirical content, have no such necessity.

Second — a closely related point — those two nonKantian maxims are willed qua strong, cunning, resourceful, or proud person, whereas authentically Kantian maxims have to be willed qua rational person. As such, they have to be imposed, or rather on Kant’s view impose themselves, independently not only of consequences, but of the agent’s merely contingent circumstances. Hence these two nonKantian maxims, although certainly universalizable without inconsistency, cannot play the part that maxims have to play for Kant. And this makes it clear that the first formula of the categorical imperative, as presented in the *Grundlegung*, cannot stand by itself. What is needed by way of further interpretation is provided by the second and third formulas. It is for this reason that the question of the rational justification of the derivation of maxims with particular content from Kant’s first formula for the categorical imperative by itself may not have quite the significance that both some critics of Kant, including myself, and some Kantian, NeoKantian, and QuasiKantian defenders of Kant have sometimes supposed.

It has indeed been a commonplace, ever since Hegel’s critique of Kant, that there are problems about precisely how actionguiding maxims with particular content are to be derived from the categorical imperative in its first formulation. Onora O’Neill (in one way in *Acting on Principle* [New York: Columbia University Press, 1975] and in another in “Consistency in Action,” in *Constructions of Reason* [Cambridge: Cambridge University Press, 1989]) and Christine Korsgaard (“Kant’s Formula of Universal Law,” *Pacific Philosophical Quarterly* 66 [1985]) have made a number of different

compelling suggestions here. And more recently Barbara Herman has concluded that, although on one interpretation of that formulation of the categorical imperative -in terms of what Onora O'Neill has called contradiction in the will (Acting on Principle, chapter 5, pp. 82-93) — *it* excludes maxims that ought not to be excluded, and on another — in terms of what O'Neill has called contradiction in conception (chapter 5, pp. 63-81) — *it* fails to exclude what ought to be excluded, a joint use of these two formulations, supplemented by subsequent deliberation of a highly specific kind, *can* generate in a rationally justifiable way the needed kind of practical conclusion (*The Practice of Moral Judgment* [Cambridge, Mass.: Harvard University Press, 1993], chapter 7; it should be noted that O'Neill's own view both of the relationship between the different formulas of the categorical imperative and of how principles relate to particular types of case is not the same as either Herman's or Korsgaard's).

Each of these detailed and elegant reconstructions of Kant's forms of argument is instructive and insightful in bringing out the richness of Kant's resources. Each inevitably goes beyond the letter of the text in its interpretation — and even at points in ways that are incompatible with Kant's own positions, since he held the three formulas of the categorical imperative to be equivalent — but none of them illegitimately. Yet before they can be evaluated as adequate or inadequate what needs to be remarked is the striking contrast between their detailed interpretative subtleties and disputed questions and Kant's representation of the straightforward apprehension of the necessity of true moral judgments by plain moral persons. This was of course a problem for Kant himself before it was a problem for Kantians, the problem of how to capture what Kant called “the happy simplicity” of “the ordinary understanding” of plain persons (*Grundlegung*, first section, 405) in adequate philosophical terms without distortion. So that it might after all be best to begin not with the necessarily problematic and philosophically sophisticated issues about derivation raised by Kant's recent interpreters, but with the

relatively straightforward moral conclusions, which, on Kant's view, plain persons are able to reach from their own inner rational resources, and to enquire what light those conclusions throw upon the premises from which they are taken to be derived rather than vice versa.

In the case of lying it will turn out, so I shall argue, that Kant's moral conclusion — or rather what Kant takes to be the moral conclusion of “the ordinary understanding” — brings out the importance for the Kantian moral standpoint of the fact that the first formulation of the categorical imperative cannot stand by itself, but needs to be interpreted and supplemented by the second and third formulations — and in this at least I follow Christine Korsgaard — and that there is therefore a more complex relationship between the categorical imperative prohibiting lying and the semantic rule prohibiting false assertions than at first appeared. What then are the important features of Kant's conclusions about lying? They turn out to be just those features that outraged Benjamin Constant. Constant had argued that obedience to a moral principle unconditionally enjoining everyone to speak the truth and unmodified by other principles would make all social life impossible. “We have the proof of this,” he said, “in the consequences drawn from this principle by a German philosopher, who goes so far as to assert that it would be a crime to lie to a murderer who enquired whether our friend, whom he was pursuing, had not taken refuge in our house” (*Reactions politiques* [Paris, 1797], chapter 8, quoted in *Un droit de mentir? Constant ou Kant*, by F. Boituzat [Paris: PUF, 1993]).

This example may have been a commonplace in eighteenth-century discussions of lying. Samuel Johnson, as I noted in the first of these lectures, had already discussed it and Johann David Michaelis, professor of theology at Göttingen until his death in 1791, anticipated Kant's conclusions with regard to it. Later on, Newman was to make use of it. Kant's response to Constant's report of his position was at once to acknowledge that he really had said this, although he could not remember where (*Kant's Critique of*

Practical Reason and Other Works on the Theory of Ethics, trans. T. K. Abbott [London: Longmans Green, 1873], p. 361). But he focused his attention upon Constant's statement of Constant's own rival view.

Constant's view was that "to tell the truth is a duty only towards a person who has a right to truth" and that therefore to someone who by reason of her or his malevolent intentions has no such right it is not wrong to lie (as quoted by Kant in *On a Pretended Right to Lie from Benevolent Motives*, in Abbott). Against Constant, Kant contended that "truthfulness in assertions that cannot be avoided is a human being's formal duty to everybody, whatever the disadvantage that may ensue to oneself or to another." Someone who has unjustly compelled me to make a statement is not, on Kant's view, the one wronged by my lie. So the question of whether or not such a one has or has not the right to truth is irrelevant. If I lie, "I do wrong to humanity in general in the most essential point of duty . There need be no injury to any particular person, but rather humanity itself is wronged. And, as becomes clear if we turn to Kant's other writings, it is important that veracity is something that we owe to ourselves quite as much as to others. By lying the liar in wronging humanity wrongs herself or himself.

"The greatest violation of a duty to oneself considered only as a moral being (the humanity in one's person) is the opposite of veracity: lying . And Kant proceeds to define lying by quoting

Sallust and then makes a distinction between external and internal lying. "The former," he says, "renders a man despicable in the eyes of others, the latter" — Kant means by an internal lie a lie told to oneself, a piece of self-deception — "in his own eyes which is much worse and violates human dignity in his own person

"Someone who does not believe what he says to another (even if it be a person existing only in idea) has even less worth than if he were a mere thing; a thing has utility, another can make some use of it, since it is really a thing. But to communicate one's thoughts to someone

by words which (intentionally) contain the opposite of what one thinks is an end directly contrary to the natural purposiveness of one's capacity to communicate one's thoughts. In so doing one renounces one's personality and, as a liar, manifests oneself as a mere deceptive appearance of a human being, not as a genuine human being" (*Metaphysic of Morals*, Part II, first part of the Elements of Ethics, 9).

On Kant's view then no injury other than the lie itself need have been brought about, either to oneself or to another, for a lie to be a wrong and a wrong of this magnitude. From what fundamental positions do these striking, and to some affronting, conclusions flow? To answer this question we need to remind ourselves of some of Kant's basic theses. One is that to lead a life in accordance with the maxims of morality, moved by a prudent understanding that conformity to the moral law can serve "the incentive of self-love and its inclinations," is to have a bad moral character. So, if we were to refrain from lying only or even in part because "truthfulness, if adopted as a basic principle, delivers us from the anxiety of making our lies agree with one another and of not being entangled by their serpent coils" (*Religion within the Limits of Reason Alone*, Book I), we would no more have genuinely obeyed the categorical imperative that prohibits lying than if we had lied. But now what of that categorical imperative? If it is to provide a premise that affords sufficient reason for the conclusion that no one ought ever under any circumstances to lie, it cannot be understood only as the categorical imperative of the first formulation. It must, for reasons that I have already indicated, be understood so that the second and third of Kant's formulations supplement and interpret the first. This conclusion, as I noticed earlier, agrees with that reached by Christine Korsgaard ("The Right to Lie: Kant on Dealing with Evil," *Philosophy and Public Affairs* 15, no. 4 [1986]).

She has argued that the different formulations give different answers to the question of whether if, by lying, someone may prevent a would-be murderer from implementing her or his intentions, that

person may do so. The Formula of Universal Law “seems to say that this lie is permissible,” but the Formula of Humanity “says that coercion and deception are the most fundamental forms of wrongdoing. In a Kingdom of Ends coercive and deceptive methods can never be used” (p. 337). We must then, it seems, understand the Formula of Humanity and the conception of the Kingdom of Ends as narrowing the restrictions imposed by the universalizability requirement, so that Kant’s rigorist conclusion is indeed warranted by the premises from which he derives it. But, of course, if this is so, then a problem arises for all those who stand with Benjamin Constant or with the John Stuart Mill of *Utilitarianism* or who for other reasons reject that conclusion. *For if that conclusion is warranted by the premises, then those who reject the conclusion are committed to rejecting at least one of the premises. So we need to enquire further about both conclusion and premises.*

To this way of going about things it may be objected that Kant did not in fact hold with any great seriousness the conclusion that lies ought *never* under any circumstances to be told, except as what H. J. Paton called a “temporary indiscretion,” which Paton ascribed to “bad temper in his old age” (“An Alleged Right to Lie: A Problem in Kantian Ethics,” *Kant-Studien* 15 [1954]). Sallie Sedgwick, who repudiates Paton’s characterization of what he took to be Kant’s lapse, has argued nonetheless that Kant is misunderstood if we suppose that Kant’s rigorist conclusion really follows from his premises. She points out that earlier in the *Vorlesung* Kant had held that, if I am compelled to make a statement of which improper use will then be made, I can be justified in telling a white lie (see on this *Eine Vorlesung Kant’s über Ethik*, ed. P. Menzer, pp. 288-89), trans. L. Infield, *Lectures on Ethics* [Indianapolis: Hackett, 1980], p. 228). And she contends that there was in fact no change after the *Vorlesung* in the spirit of Kant’s views, but only in the letter (see “On Lying and the Role of Content in Kant’s Ethics,” *Kantstudien* 82, no. 1 [1991]). She is, however, surely mistaken about the spirit of Kant’s later views. Kant took care to reject in explicit terms the thesis, which has been defended as in the

spirit of Kant's view not only by Sedgwick, but also by some earlier commentators, that he should have treated the prohibition on lying only as a fundamental principle, not one immediately determining action, but one that needs to be interpreted and qualified through mediating principles in its application to particular cases. When in his response to Constant Kant addressed this very issue, he concluded that "all practical principles of justice — such as the prohibition of lying — must contain strict truths, and the principles here called middle principles can only contain the closer definition of their application to actual cases. . . and never exceptions from them. . . ." For this reason as well as in the light of the texts cited earlier I cannot agree with Sedgwick and I also conclude that the *Vorlesung* should not be used as reliable evidence for Kant's developed views.

Sedgwick has, however, by the insightful way in which she has pressed her case brought out features of Kant's position that it would be wrong to ignore, features that suggest possible underlying unresolved tensions within Kant's thought. But the significance of those tensions will only appear once we have a more adequate view of Kant's position and therefore of the possible grounds for rejecting it. Consider another of Kant's basic theses, that "it is our common duty as human beings to elevate ourselves" to an ideal of moral perfection, the idea of a human being whose life would in every way satisfy the requirements of a wholly good God; and that for the achievement of "this archetype of the moral disposition in all its purity" "the idea itself, which reason presents to us for our zealous emulation, can give us power" (*Religion within the Limits of Reason Alone*, Book II, section 1, A).

What these two basic theses of Kant's make evident is that, on his view, morality requires a systematic disciplining of and freeing ourselves from responsiveness to our own inclinations. It is not that we shall not as moral beings continue to have inclinations and to be recurrently responsive to them. It is that we have to become the kind of person for whom the incentive to action supplied by

inclinations is always subordinated to the incentive of rational willing in the pursuit of moral perfection. Kantian rationality therefore involves a particular and radical kind of asceticism in respect of the passions, an asceticism directed toward the perfecting of the self. This is an extraordinary task, one that, as Kant understood, confronts even greater obstacles than those recognized by his predecessors in this moral asceticism, the Stoics (*Religion within the Limits of Reason Alone*, Book II). And the recognition of this task and those obstacles is one of the distinctive features of Kant's standpoint. What should that recognition involve in our relationships with others?

Kant's answer is illuminated by his discussion of friendship in the *Metaphysic of Morals*. Kant takes it that friendship of a certain kind "is an ideal in which a morally good will unites both parties in sympathy and shared well-being" and that aiming at such friendship is an honorable duty proposed by reason. We do need friends, but it is important that there are limitations upon the possibilities of friendship and some of them are imposed by the constraints of a morally good will. Kant distinguishes at least two kinds of friendship. He praises what he calls moral friendship, a relationship in which each friend is able to reveal her or his otherwise unspoken thoughts and opinions to the other without fear that her or his secrets will be revealed. He defines moral friendship as "the complete confidence of two persons in the mutual openness of their private judgments and sensations, as far as such openness can subsist with mutual respect for one another" (*Metaphysic of Morals*, Part II, second part of the Elements of Ethics, 47). But this of course differs in key respects from friendship as it had been traditionally understood from a variety of standpoints.

Such friendship characteristically involved not just moral, but also what Kant calls pragmatic friendship, of which he says that it burdens itself with the aims and purposes of other human beings. Because it is "a great burden to feel oneself tied to the destiny of others and laden with alien responsibilities," pragmatic friendship is a moral liability.

“Friendship therefore cannot be a bond aimed at mutual advantage, but must be purely moral” (46), a friendship of equal respect as well as of mutual confidence. And equal respect is actually incompatible with a friendship based on advantage. For “if one accepts a benefit from the other, then he can probably count on an equality in their love, but not in their respect; for he sees himself as plainly a step lower, inasmuch as he is obligated and yet not reciprocally able to obligate.”

This is the point at which it is salutary to recall that in the example that elicited Constant’s attack upon Kant the murderer’s intended victim whom one may not protect by lying is a friend. That the life to be saved is that of one’s friend gives one no reason at all, according to Kant, to lie. A friend with a morally good will would not of course will it otherwise, both because she or he would herself or himself do no other in a like situation and also presumably because it would be a burden to accept the benefit conferred by this lie from the other. We should be grateful to Kant for making so clear to us what is entailed by his fundamental theses, but, as I noted earlier, not every follower of Kant has been grateful. Because, like so many nonKantians, they have found Kant’s conclusion on this particular issue morally repugnant, they have hoped to show that it does not follow from Kant’s universal premises. But I earlier suggested reasons for holding that on this point they are mistaken. All that has now been added is an acknowledgment that what Kant takes to be the universally binding principles of reason can of course provide no grounds for an exception in favor of one’s friends.

Someone might respond by suggesting that, since Kant unhesitatingly recognizes a duty to help those in dire need, any difficulty in accepting Kant’s conclusions can be met by carefully qualified statements, first of the duty not to tell lies and second of that to help those in dire need, so that questions of which duty is to have priority in particular types of situations can be answered by making it permissible to lie in some types of situations. But this notion of priority is quite alien to Kant himself where matters of perfect duties

are concerned. Kant does indeed recognize that “two grounds of obligation can be conjoined in a subject,” so that a conflict may *seem* to arise, but if so, one of the grounds is not in fact a duty (*Metaphysic of Morals*, Introduction, 224). Where perfect duties such as that of truthfulness are concerned, each can give way to no other ground of obligation. And about this there seems to be something importantly right, both from a Kantian and from some nonKantian points of view, including my own (see Alan Donagan, “Consistency in Rationalist Moral Systems,” in *Moral Dilemmas*, ed. C. W. Gowans [Oxford: Oxford University Press, 1987], and my own “Moral Dilemmas” in *Philosophy and Phenomenological Research* 50, supplement [Fall 1990]).

For, as I suggested in the first lecture, it is difficult to make sense of the notion of weighing the value of refraining from lying by reason of truthfulness against that of saving an innocent human life. Within Kant’s own moral and philosophical scheme there is evidently no room for any conception of the scales on which such weighing might take place. But, quite apart from Kant’s scheme, it is difficult to translate the metaphor of weighing in any appropriate and relevant way into an account of a rationally justifiable criterion for deciding between the claims of what are taken to be in certain types of situation rival values. And if there is no such criterion, then what the metaphor of weighing would disguise would be arbitrary choices between values and between duties, notions equally unacceptable to Kant. It seems to follow that no revision of Kant’s moral scheme of the kind suggested is possible without abandoning too much that is crucial to Kant. So that there is further confirmation of the thesis that anyone who holds to the substance of Kant’s view in general is committed to Kant’s particular conclusions respecting that remarkable triad, the pursuing murderer, the pursued friend, and the intervening person of rational principle.

It is important to emphasize that although, on Kant’s view, the intervening person of rational principle may not lie to the pursuing murderer, there are on Kant’s and indeed on any reasonable

view a number of other things for her or him to do, or at least to attempt. She or he may and presumably must attempt to distract the would-be murderer's attention, to trip up, knock down, or otherwise hinder the murderer, to remain silent, so that the murderer is deprived of needed information, to irritate the murderer into turning his aggression against her or him instead, and so on. But, if these all prove ineffective, that ineffectiveness, on Kant's view, furnishes no reason for violating fundamental principles.

In this of course Kant is reiterating, as I noticed earlier, a longheld Christian view, not the only Christian view certainly, but the view of, among others, Augustine, Aquinas, and Pascal. Moreover, his moral standpoint agrees in its conclusions with those of a number of twentieth-century practitioners of nonviolence whose admirable moral intransigence has earned them hard-won respect. So is there after all good reason to dissent from Kant's conclusions?

I intend to assert that there is, but, before I do so, I want to accept from Kant a constraint upon any acceptable answer to this question. It is this: any principle that warrants us in lying in certain circumstances, as to a would-be murderer, must be either one and the same principle that forbids us to lie in every other case or at the very least a principle that cannot generate possible inconsistency with that primary principle. The permitted or required lie must not be understood as an *ad hoc* exception, since, for reasons that Kant makes admirably clear, there cannot be such exceptions to genuine moral rules. And the principle that permits or requires a lie must not be some independent principle, potentially in conflict with the principle forbidding lying, since, for reasons that Kant also makes clear, our moral principles must be a consistent set, consistent to this degree that they do not, in any situation that has occurred or will occur or may occur, prescribe incompatible actions, so that one or the other has to be modified in an *ad hoc* way. The best way of excluding both of these inadmissible modes of permitting or requiring a lie is to have sufficient grounds

for holding that one and the same principle *both* generally and indeed almost always prohibits lying *and* yet requires it on certain **normally rare types** of occasion. Is there any such principle and what might it be?

V

I begin by considering two objections that may be made to Kant's position, objections with which, on the view that I shall be proposing, any acceptable account of lying and truth-telling must come to terms. Both are objections directed not only against Kant's position, but more generally against any position that entails the same conclusions about the legitimacy and justifiability of only nonviolent and nonlying resistance to the evil of intended murder. The first of these objections is that, willingly or unwillingly, the consistent Kantian can rarely escape being a moral free-rider. The social and civic orders within which the vast majority of human beings live out their lives are sustained by systematic uses of coercion and lying that Kantians, pacifists, and others may disown and condemn, but the benefits of which they cannot escape. Indeed, if such Kantians or pacifists are to discharge adequately certain responsibilities within their own society, they may find themselves forced to recognize this. One notable example concerns the government of Pennsylvania by members of the Society of Friends in the early eighteenth century. Themselves morally committed to nonviolence and to the abhorrence of all violence, they could not protect those for whose safety they were responsible without providing a military defense against American Indian incursions. And so they hired others to fight in their place. Failure to do so would have been a dereliction of political duty, but by doing so they became moral free-riders, relying upon others to do what they themselves could thereby avoid doing. I use this example not at all to stigmatize eighteenth-century

Pennsylvanian members of the Society of Friends. My point is rather that, if they, among the most conscientious and admirable of human beings by any reasonable standards, could not evade this outcome, no one else espousing such principles is likely to be able to do so.

A second objection is of a very different kind. It is that there are some particular cases — I speak here of particular cases and not of types of case, although to present the particular cannot but be to present it as being of a certain type — about which your judgment or mine may be such that, if those judgments are incompatible with the universal and general principles that you or I have hitherto held, then it is the universal and general principles, as up to now formulated, that we shall have to reject or at least revise. We have very few philosophical discussions of the status of such particular judgments (there is one in *Den Etske Fordring* by K. E. Løgstrup [Copenhagen: Gyldendal, 1956]; see also more recently Michael DePaul, *Balance and Refinement* [London: Routledge and Kegan Paul, 1993]) and here I shall put questions about that status on one side. But I take it that the experience of being constrained in one's moral judgment by the features of a particular case, prior to and independently of any subsequent universalizability, is not that uncommon. Which then are the two particular cases to which I appeal?

The first is of a Dutch housewife in the period in which the Netherlands were ruled by the military police power of Nazi Germany. Just before her Jewish neighbor was arrested and sent to a death camp, she had taken that neighbor's child into her own home and promised to take parental responsibility for that child. Confronted by a Nazi official who asked her whether or not all the children living in her home were her own she lied. The second example is of a somewhat different kind and does not concern a lie, although I hope that its relevance to the issue of lying will become clear. It is that of a Massachusetts single mother not so long ago, the life of whose infant child was immediately threatened by a violent and estranged man, a

former lover, physically much stronger than she, whose threats to the life of her child were without doubt seriously intended. Her response was to snatch up a gun and kill him. A question that became of focal importance at her trial was: what else could and should she have done? The two examples are importantly different. But in both cases I find, as do at least some others, that I cannot withhold the judgment that, had either of these women done other than she in fact did, she would have failed in her duty to the child whose maternal protector she was.

Is this inability perhaps no more than evidence that those of us who exhibit it are in the grip of moral superstition? Among the ways in which this accusation might be rebutted would be the identification of some well-founded principle or set of principles that is able to provide justification for those particular judgments. The formulation of such principles has to begin from a very different starting point from that from which Kant set out. Instead of first asking "By what principles am I, as a rational person, bound?" we have first to ask "By what principles are we, as actually or potentially rational persons, bound in our relationships?" We begin, that is, from within the social relationships in which we find ourselves, the institutionalized relationships of established social practices, through which we discover, and through which alone we can achieve, the goods internal to those practices, the goods that give point and purpose to those relationships. But we also begin as rational persons within those relationships, understanding them as always open to criticism, to possible modification or revision in the light of criticism, and even in the end to possible rejection, if they turn out not to be open to worthwhile modification or revision. Yet that ability to criticize is itself something characteristically acquired in and developed out of the experience of such relationships. It too, when it is rationally effective, appeals to already recognized or recognizable norms of criticism. Moreover, we cannot but acknowledge in those relationships a variety of types of inescapable dependence upon some of those others to whom we are related; we have to rely on some of these types of dependence

to foster our initial autonomy and to sustain it later on. Autonomy thus achieved does not then consist in total independence from and of the sentiments, judgments, and actions of others, but in an ability to distinguish those areas in which one ought to be independent and those where one ought to acknowledge dependence. To be in this way autonomous in one's relationships is a necessary condition for achieving many, although not perhaps all, of those key goods without which our relationships no longer have point and purpose.

Why within our relationships, if they are thus understood, is truthfulness important and why ought lying to be prohibited? For at least three mutually reinforcing kinds of reason. First, without consistent truthfulness by others and by ourselves we cannot hope to learn what we need to learn. We need to be told truthfully about our own intellectual and moral deficiencies. We also need to be able to speak truthfully to others about that in them and that for which they are responsible that is or may be damaging to our relationship with them. A lack of *ressentiment* and the possession of tact, patience, and charity are of course also required, if this kind of truthfulness is to be effective. And if it is not effective, it loses its point. So the exercise of truthfulness in this area is not independent of the exercise of other virtues. **But of course our** own character and that of others is not the only subject matter about which we need to learn and about which therefore truthfulness is required. What does single out the subject matter of character is that it is here that we generally find the strongest motives for lying, so that it is here that truthfulness as an ingrained and not to be overcome habit is most needed.

Second, we also need truthfulness, if we are to be able to put our social relationships to the question in the ways and to the degree that rationality requires of us, and this for two different reasons. If we are to have integrity as critics of the established patterns of relationship in which we are involved, then our criticisms of those patterns will have to be truthful. And if we are to deserve the trust of others and

to be able to trust those others, during periods in which we or they or both are engaged in sometimes painful and disturbing criticism of our ongoing relationships and of the social practices that provide their context, then we shall have to be able to rely on a shared prohibition against lying and all other relevant forms of deceitfulness.

Third, truthfulness is a virtue without which the corrupting power of phantasy cannot be held in check. Phantasy is of course indispensable and ubiquitous in human life. We are only able to be a good deal of what we are and to do a good deal of what we do because we are able to imagine ourselves as thus being and thus doing. Myths, dramas, and novels, and also such peculiarly modern fictions as the reports of corporations, the programs of political parties, and the confessional disclosures of televisual interviews can only function as they do because of the modes in which we all in different degrees and different ways imagine both our own lives and the lives of others. And myths, dramas, and novels are of course sometimes powerful in conveying truths. But the same power of phantasy can be and often is used to disguise and to distort our activities and our relationships and has the effect of deforming them, and psychoanalysis should by now have taught us the extent of this power. What psychoanalysis itself, at least in some versions, has also attempted to instruct us in is one particular discipline of truth-telling. And we need a corresponding discipline in our everyday lives and relationships, if we are to see those lives and relationships as they are rather than as they are misrepresented as being under the influence of a range of often unacknowledged hopes and fears.

I remarked in the first of these lectures that the successful liar exercises a certain kind of illegitimate power over those who are deceived. That illegitimate power deprives those others who are deceived of their autonomy in their relationships with the liar. And so the relationship itself is deformed, becoming one of sometimes multiplying illusions. It is therefore evident that in any relationship in which the goods of rational persons are to be achieved, the truthfulness of those

participating in that relationship will be of crucial importance. And this will have to be a truthfulness that extends beyond the persons involved in that particular or any other particular relationship, and this for a good, almost Kantian reason. The truthfulness required has to embody a respect for the rationality of all persons who are or could be involved in all actual or potential relationships. It is a truthfulness that is as necessary for integrity in our relationships with strangers as with friends and, if this integrity is lacking in our relationships with strangers, it will as a matter of fact also be at least endangered and often enough corrupted in our relationships with friends.

om this moral point of view that I have been sketching the evil of lying then consists in its capacity for corrupting and destroying the integrity of rational relationships. To understand this is to be able to relate the evil of lying to other evils. For it is one salient characteristic of evils in general that they are destructive of rational relationships. Those persons who are outside our particular set of relationships constitute no threat to those relationships simply by their being outside, by their being strangers. And to suppose that they are is always itself a corrupting phantasy. But, if and when they aggress against those who are bound to each other in some particular relationship, then it is always someone's responsibility to do whatever is necessary, so far as they can, to defend that relationship against that aggression. Whose responsibility this is will depend upon the character of the relationship. What their responsibility requires them to do will depend upon the nature of the aggression. Consider in this light the cases of the Massachusetts mother and of the Dutch housewife that occasioned my statement of this point of view.

I remarked earlier that moral development within institutionalized relationship involves growth from an acknowledged dependence toward rational autonomy. Part of what rational autonomy requires is a recognition of the dependence of others upon us, especially of the dependence of children and most of all of our own children. That

recognition is a recognition of duties and both the Massachusetts mother and the Dutch housewife are, on the view that I am taking, examples of those who did what duty required of them. Theirs were relationships in which each had assumed responsibility for the life and well-being of the dependent child, and in each of which therefore that child was entitled to trust the mother to do what was necessary for its effective protection. In the case of the Massachusetts mother this clearly required disabling the aggressor and, if the only way open to her of disabling the aggressor was by killing him, as it seems in fact to have been, killing him. Had she failed to do this, she would have failed in her duty to her child. And if, by killing the Nazi official, the Dutch housewife could have taken the only effective course of action open to her to protect the child in her care, then it would have been her duty to kill that official. But for anyone in such a situation two questions always have to be answered and will in fact have been answered by whatever action is taken. Will this proposed action effectively protect whoever or whatever needs to be protected? And does this proposed action go beyond what is needed in harming the aggressor? The latter question matters because, insofar as I become a doer of harm beyond what is needed, I pass from being a defender of those unjustly attacked to being myself an unjust aggressor.

To the Dutch housewife it must have been evident that, even were she able to kill the Nazi official, the consequence would have been a reign of murderous terror directed against the entire community, including the children whom she was pledged to protect. Moreover, killing the Nazi official would have done unnecessary harm, provided only that she was able instead to lie convincingly. In this type of case the normally illegitimate power exercised by the successful liar becomes legitimate, first because and insofar as it provides a defense against the prior illegitimate exercise of power by the aggressor, and second because by lying she avoids other more harmful uses of power. I take it therefore that the Dutch housewife's lie and all

other lies of just the same kind were and are justified. But what is this kind and how is the rule that justifies them to be formulated?

It would be misleading to state it as though its form was “Never tell a lie, except when.. .” For this would suggest that we were first formulating a rule and only later, as a second thought, introducing an exception. But this is a mistake. The rule that we need is one designed to protect truthfulness in relationships, and the justified lies told to frustrate aggressors serve one and the same purpose and are justified in one and the same way as that part of the rule that enjoins truthfulness in relationships. The Massachusetts mother and the Dutch housewife upheld in their exceptional circumstances just what the normal rational truthful person upholds in her or his everyday life. The rule is therefore better stated as “Uphold truthfulness in all your actions by being unqualifiedly truthful in all your relationships and by lying to aggressors only in order to protect those truthful relationships against aggressors, and even then only when lying is the least harm that can afford an effective defense against aggression.” This rule is one to be followed, whatever the consequences, and it is a rule for all rational persons, as persons in relationships.

About this rule two things need to be said. First, although it is evidently inconsistent with Kant’s fundamental principles, and moreover is justified by arguments that Kant could not but have rejected, it is nonetheless deeply indebted to Kantian insights and arguments. Its justification by appeal to particular examples, its teleological perspective, and its conception of persons-in-social relationship as the fundamental units of the moral life do all put it at odds with Kant’s standpoint. But in its acknowledgment of the fundamental character of respect for rationality, in its rejection of consequentialism, and in some features at least of its conception of autonomy it recognizably draws upon Kantian resources.

Second, it is a rule that is not merely consistent with but supportive of Mill's conception of truthfulness as crucial to social and moral enquiry and therefore to any social order whose relationships are systematically open to such enquiry. And it is indeed in some of its aspects a rule whose formulation is as clearly indebted to Mill as it is to Kant. It is one of the strengths of this rule that it integrates central features of Mill's view with central features of Kant's. One outcome of my examination of Mill's views in the first of these lectures was a suggestion that Mill over large areas of social life upheld what was in effect a rule requiring unqualified truth-telling. Yet it is also evident that Mill was deeply committed to the view that certain kinds of threat to human welfare not only permit but may require the telling of lies. My account of what those kinds of threat are does not entirely coincide with Mill's account, but it is in agreement with all or almost all of Mill's social and political concerns, so far as those involve lying and truth-telling. Most importantly, it enables us to understand better just why the moral and political life must be, just as Mill held, a life of practical enquiry. For if it is in and through our social relationships that we achieve goods and recognize the authority of rules, and if that achievement and recognition requires, as it does, shared activities of criticism, in which we ask how the goods of this and that relationship can be better ordered, so that they can become the goods of a whole human life and the goods directing communal activity, then systematic enquiry becomes one central thread of the moral life. And one ground for our concerns about truthfulness is the need for truthfulness in enquiry, just as it was for Mill. Nonetheless — it scarcely needs saying — this account that I have given remains deeply at odds with Mill's consequentialism.

I began these lectures by identifying two distinct sources for the universal human concern over the harms and dangers of lying, one concerned primarily with truth and one concerned primarily with trust. What reflection upon Mill and Kant has led me toward is a conception of truthfulness as informing and required by rational

human relationships, a conception that does seem to go some way toward integrating concerns about truth and concerns about trust. For to understand the rules prescribing unqualified truthfulness as governing relationships, rather than individuals apart from their relationships, is also to understand how the concern for truth and the concern for trust can become complementary. Central to my trust in you as spouse or friend or colleague, as someone to whom I stand in a relationship of commitments, including commitment to moral enquiry, is my confidence that on any matter relevant either to our relationship or to those other relationships to which each of us is committed I will never be told by you anything other than what you believe to be true. And you know that I know that you know that what I will have discovered if I discover you in an untruth, or vice versa, is that you have to a greater or lesser degree defected from our relationship. Lies then become understood, as they should be, as small or large betrayals and the virtues of integrity and fidelity are understood to be at stake in all those situations in which the virtue of truthfulness is at stake. The disturbance characteristically caused by the discovery of such a lie is well described by Frankfurt as due to its also being a discovery that one “cannot rely upon” one’s “own settled feelings of trust” (“The Faintest Passion,” p. 7). But where Frankfurt is specifically concerned with lies told to one by those whom one had taken to be one’s friends, I am suggesting that all violations of well-founded rules concerning truth-telling in established social relationships deserve very much the same response.

It remains true of course that this account will be unacceptable to anyone who is either, unlike Mill, a consistent utilitarian or, like Kant, a consistent Kantian. And moral philosophers in general these days tend to be either utilitarian or Kantian. How then should further conversation proceed?

VI

Enquiry needs to go in more than one direction. The first is that at which we have already made a beginning by considering and evaluating rival answers to questions about the permissibility of lying. And here of course we still need further consideration and further evaluation. A step beyond this would be to set those questions in a somewhat wider context, that of the ethics of conversation and discourse in general. For medieval writers and for their modern heirs up to and including Kant, lying was after all only one species of forbidden speech. Aquinas analyzed and condemned a whole range of types of malicious and abusive speech. And Kant wrote in the same tradition, when in the *Vorlesung* he discussed not only the wrongs done by liars, but also the wrongs done by those who slander, scoff, and mock.

I suggested earlier that we may be able to identify in Kant's thought certain underlying, unresolved tensions. One of these is that between his general suspicion of teleology in ethics and his occasional appeals to teleology, as when he speaks of the liar, in a passage from which I quoted earlier from *The Metaphysic of Morals*, *as doing wrong by pursuing "an end directly contrary to the natural purposiveness of one's capacity to communicate one's thoughts."* *One hypothesis that needs to be investigated is that the principles presupposed by Kant's contributions to an ethics of conversation and discourse are inescapably teleological and are so in a way that the framework of Kantian ethics cannot accommodate except by ad hoc patchwork.* Were this hypothesis to be vindicated, we should have found in Kant, as we have found in Mill, some degree of inconsistency. And a further interesting question would then be that of whether a framework of thought and practice afforded by a conception of the moral life as that of rational persons in relationship, pursuing the goods of their relationships, in activity and in conversation — developed much more fully than I have been able to develop it here —

might not more adequately accommodate both what we have to learn about truthfulness from Mill and also what we have to learn from Kant than either Mill's own utilitarianism or Kant's own apriorism can.

These then are some directions in which I would want to carry enquiry further. But we also need to become self-conscious about the moral requirements of enquiry itself. When Kantians, utilitarians, their various critics, and the proponents of a range of alternative and rival positions, such as my own, have completed the task of stating their reasons for holding their own views and for rejecting those of their opponents, we all confront the question of what moral basis it is on which enquiry can best be carried further, in a way that will ensure a reasonable outcome and that will not be question-begging. Any adequate answer will have to specify both what the functions are of truthfulness, trust, and truth in the work of cooperative enquiry itself and what the relevance of the conception of truthfulness, trust, and truth required by such enquiry is to the moral life in general.

Here my initial hypothesis would be that it is only in terms of a developed conception of the moral life as itself a life of practical enquiry that the relevance of moral enquiry to the moral life can be adequately understood. But for the present this too can only be presented as a hypothesis. It is with hypotheses and questions that I end rather than with theses and answers. I end therefore, not with an ending, but somewhere still in the middle. Yet that is after all not an uncharacteristic place for philosophers to end up.

