

Trust

ANNETTE C. BAIER

THE TANNER LECTURES ON HUMAN VALUES

Delivered at

Princeton University

March 6-8, 1991

ANNETTE C. BAIER is Professor of Philosophy at the University of Pittsburgh. She was born and educated in New Zealand and studied philosophy at Otago University. Post-graduate study at Oxford was followed by teaching positions at the Universities of Aberdeen, Auckland, and Sydney and the Carnegie-Mellon University. She has published a volume of essays, *Postures of the Mind* (1985), and *A Progress of Sentiments: Reflections on Hume's Treatise* (1991). She sees her recent work on trust as growing not only out of her philosophical reading but also out of her experience of living in a variety of cultural settings and climates of trust.

1. TR?UST AND ITS VULNERABILITIES*

They fle from me that sometyme did me seek
With naked fote stalking in my chambre
I have seen them gentill tame and meke
That nowe are wyld and do not remembre
That sometyme they put theimself in daunger
To take bred at my hand . . .

Sir Thomas Wyatt

Most of us are tame enough to take bread at someone's hand. And we do thereby put ourselves in danger. So why do we do it? What bread is good enough to tempt us into the hands of possibly dangerous people tamers? Or do we simply prefer being gentle, tame, and meek? Trust in trustworthy people to do their more or less willing and more or less competent bit in some worthwhile cooperative enterprise whose benefits are fairly shared among all the cooperators is to most of us an obviously good thing, and not just because we get better bread that way. The only ones who might dissent from the value of trust are those "wild" loners who value their independence more than anything else, who prefer to

*The original titles of these two Tanner lectures were "The Pathologies of Trust" and "Appropriate Trust." I was greatly helped in revising the lectures by the prepared comments of Francine du Plessix Gray, Geoffrey Hawthorn, Thomas M. Scanlon, Jr., and David Shipler when I gave the lectures; by suggestions from the audience; by later discussion with Princeton University faculty and students; and by subsequent correspondence with Sarah Buss, Pamela Foa, Richard Moran, and Thomas Scanlon. The revisions were made at the Rockefeller Study Center, Bellagio, Italy, and I am very grateful to have had the opportunity to enjoy the beauty, peace, and good company to be found there. The peace was also instructive for my study of trust, since our idyllic headland was protected, during most of my stay there, by an armed guard. Italian soldiers with machine guns patrolled the grounds and guarded the entrances against perceived terrorist threats. So our easy mutual trust, within our sanctuary, had as its exterior face an apparent distrust for all outsiders. I am pleased to report that by the time I left, the perceived danger, and with it the guard, had gone.

get their bread baked by solo efforts, rather than to join with others in any sort of joint scheme. To such extreme individualists my lectures will have nothing persuasive to say. Most of us are fairly tame, and what John Locke said is true of us: "We live upon trust."¹ But we do not always live well, upon trust. Sometimes, like Elizabeth I of England, we have to report "In trust I found treason," or, less regally, betrayal, or, even less pompously, let-down.² Trust is a notoriously vulnerable good, easily wounded and not at all easily healed.

Trust is not always a good, to be preserved. There must be some worthwhile enterprise in which the trusting and trusted parties are involved, some good bread being kneaded, for trust to be a good thing. If the enterprise is evil, a producer of poisons, then the trust that improves its workings will also be evil, and decent people will want to destroy, not to protect, that form of trust. A death squad may consist of wholly trustworthy and, for a while at least, sensibly trusting coworkers. So the first thing to be checked, if our trust is to become self-conscious, is the nature of the enterprise whose workings are smoothed by merited trust.

Even when the enterprise is a benign one, it is frequently one that does not fairly distribute the jobs and benefits that are at its disposal. A reminder of the sorry sexist history of marriage as an institution aiming at providing children with proper parental care should be enough to convince us that mutual trust and mutual trustworthiness in a good cause can coexist with the oppression and exploitation of at least half the trusting and trusted partners. Business firms whose exploitation of workers is sugarcoated by a paternalistic show of concern for them and the maintenance of a cozy familial atmosphere of mutual trust are an equally good

¹ *The Correspondence of John Locke*, ed. E. S. de Beer, vol. 1 (Oxford: Clarendon Press, 1976), p. 122.

² Geoffrey Hawthorn quotes these words of Elizabeth to Parliament in 1596 in his essay "Three Ironies of Trust," in *Trust: Making and Breaking Cooperative Relations*, ed. Diego Gambetta (New York: Basil Blackwell, 1988), p. 115.

example. Trust can coexist, and has long coexisted, with contrived and perpetuated inequality. That may well explain and to some extent justify the distrust that many decent vigilant people display toward any attempt to reinstate a climate of trust as a social and moral good. Like most goods, a climate of trust is a risky thing to set one's sights on.³ What we risk are not just mutually lethal betrayals and breakdowns of trust, but exploitation that may be unnoticed for long periods because it is bland and friendly. The friendly atmosphere — the *feeling* of trust — is of course a pleasant thing, and itself a good, as long as it is not masking an evil.

Trust and distrust are feelings, but like many feelings they are what Hume called “impressions of reflexion,” feeling responses to how we take our situation to be. The relevant “situation” is our position as regards what matters to us, how well or badly things are going for us. The pleasant feeling that others are with us in our endeavors, that they will help, not hinder, us, and the unpleasantly anxious feeling that others may be plotting our downfall or simply that their intentions are inscrutable, so that we do not know what to expect, are the surface phenomena of trust and distrust. This surface is part of the real good of genuine trust, the real evil of suspicion and distrust. But beneath the surface is what that surface purports to show us: namely, others' attitudes and intentions toward us, their good (or their ill) will. The belief that their will is good is itself a good, not merely instrumentally but in itself, and the pleasure we take in that belief is no *mere* pleasure but part of an important good. Trust is one of those mental phenomena attention to which shows us the inadequacy of attempting to classify mental phenomena into the “cognitive,” the “affective,” and the “conative.” Trust, if it is any of these, is all three. It has its special “feel,” most easily acknowledged when it is missed, say, when one moves from a friendly, “safe” neighborhood to a tense, insecure

³ According to Niklas Luhmann, trust always involves some assessment and acceptance of risk, so that to call trust risky becomes pleonastic. See his essay “Familiarity, Confidence, Trust: Problems and Alternatives,” in Gambetta, *Trust*, p. 100.

one. It has its (usually implicit) belief component, belief in the trusted ones' goodwill and competence, which then grounds the willingness to be or remain within their power in a way the distrustful are not, and to give them discretionary powers in matters of concern to us. When we trust we accept vulnerability to others.⁴

A third obvious way in which trust can go wrong is when the belief-cum-feeling-cum-intention of trust is faked — when a person is only apparently trusting. False pretenses can infect a trust relationship, and it may continue apparently healthy for long periods while all the time harboring such low-grade infection. A wife may not really trust her husband farther than she can see him, but she might pretend she does, perhaps pretend to herself that she does, and close her ears to any unwelcome messengers. Alternatively, she may indeed really count on his marital fidelity, but not because she trusts him. She may rely instead on her unuttered threat advantage (when, say, she controls the money and is known to have her reliable spies, so that the husband does not dare stray). Real trustworthiness, like real trust, involves feelings, beliefs, and intentions, which sometimes can be faked. The trustworthy person will feel some concern for the trusting, and this feeling will be especially noticeable if things go wrong. She will believe that she is responsible for what she is trusted for and will intend to discharge that responsibility competently and with a good grace. A "good grace" excludes not merely resentment of the responsibility but also a too calculative weighing of the costs of untrustworthiness and the benefits of trustworthiness. Should one do what one is trusted to do only because one fears that the response to discovered untrustworthiness would be very costly to one, then that very attitude, if known, would be a good enough reason for those who had trusted one to cease trusting. They might not cease relying on one, but their reliance would no longer be on one's

⁴ Here and in what follows I develop the analysis of trust given in my "Trust and Antitrust," *Ethics* 96 (January 1986): 231–60; repr. in *Feminism and Political Theory*, ed. Cass Sunstein (Chicago: University of Chicago Press, 1990).

goodwill. Trust is an alternative to vigilance and reliance on the threat of sanctions; trustworthiness is an alternative to constant watching to see what one can and cannot get away with, to recurrent recalculations of costs and benefits. Trust is accepted vulnerability to another's power to harm one, a power inseparable from the power to look after some aspect of one's good.

Trusting the untrustworthy who parade as trustworthy ("You know you can trust me!"), or living up to what another presents as her trust in one, when that is not really trust but reliance on her evident power to punish those who fail her ("I am trusting you and don't you forget it!"), are among the most common sorts of disease in a trust relationship. Healthy trust rarely needs to declare itself, and the mere occurrence of the injunction "Trust me!" or of the reminder "I am trusting you" are danger signals. Even when such pronouncements are not insincere, they may still be false, and will be, if trust has been confused with reliance on threats.

A "Trust me!" speech act (I suppose J. L. Austin would have called it a "commissive" illocutionary act),⁵ or its gestural equivalent, will be false in a more straightforward way when the implied prediction that the truster will not be "let down" proves false, not because of any deceit but because of the trusted's false estimate of his competence to "hold up" the truster. If, during one of those exercises which I believe some psychotherapists get their patients to play, I am encouraged to let myself fall back into the arms of the fellow patient behind me, whose job is to say "Trust me!" and then to catch me, I do my bit, go limp and fall, but my weight proves too much for the appointed catcher, so that I am literally let down, then I will naturally feel angry both toward the false supporter and toward the psychotherapist who choreographed my downfall. Some of those we trust let us down through their false estimate of their willingness to support us. If my upbringing has

⁵ See J. L. Austin, *How to Do Things with Words*, William James Lectures (New York: Oxford University Press, 1962), pp. 156–58.

encouraged me to rely on male escorts for defense against attack, but, when we are attacked by angry Australian magpies, my gentleman escort instinctively ducks behind me, using me as a “living shield,” then I will blame both my escort and, more, my own silly acceptance of the myth of male protectiveness.

Thomas Scanlon has helpfully separated out the different but related moral principles that he believes should govern the conduct of anyone who says “Trust me!” to others, or who somehow communicates encouragement to trust.⁶ The first principle (Principle M) forbids manipulation of others by deliberately raising false expectations in them about how one will respond to something one wants them to do. The second (Principle D), requires one to take due care not to lead others to form reasonable but false expectations about what one will do, where they would face significant loss if they relied on such false expectations. The third (Principle L), requires one to take steps to prevent any loss that others would face through reliance on expectations about one’s future behavior, expectations that one has either intentionally or negligently (that is by infringing Principles M or D) led them to form. Principle L could demand a very great deal of us, if we really tried to live by it; it would require us to *notice* what others are coming to rely on in us, and to protect them against loss from such reliance by whatever steps were needed. The fourth principle, the fidelity principle (F), does not require us to do more than we have assured another we will do; it requires us to do precisely what we assured them we would do. (I have given Scanlon’s principles in a somewhat oversimplified form. His main aim is to show that the fidelity principle is what makes a promise binding, whether or not there is a “social practice” of promising, or a special recognized force to the words *I promise*, and his careful wording of the fidelity principle has that end in view.)⁷

⁶ Thomas Scanlon, “Promises and Practices,” *Philosophy and Public Affairs* 19 (Summer 1990): 199–226.

⁷ “Principle F: If (1) A voluntarily and intentionally leads B to expect that A will do *x* (unless B consents to A’s not doing *x*) (2) A knows that B wants to

The psychotherapist who instructs me to “Trust him!” that is, tells me to trust the weakling behind me to catch me, is manipulative, negligent, and fails to prevent the loss I incur through his manipulation, but since he himself need not have said “Trust *me!*” he need not have offended against the fidelity principle. The one who offends against that is the fellow patient who, as per instructions, says “Trust me!” even if he rightly fears that he cannot catch and support me. He offends against Scanlon’s principles D, L, and F, but not against M, since he really has no wish that I should fall into his arms — we are both merely following instructions. But do we on reflection accept Scanlon’s principles? Do we not regularly and without guilt try to manipulate each other (in advertising, for example), take little care what expectations we may be arousing (in the wild birds and squirrels that we feed, and in the charities we give to), impose losses upon one another by giving misleading indicators of our intentions (in poker, in clever bargaining, and in our military strategies), let others down, often being forgiven for so doing and even sometimes invited to repeat the performance?

John Updike has a marvelous variant of the common tragedy of the letdown and its typical effects. In his story “Trust Me,” a three-year-old boy, Harold, is lovingly bullied by his father into leaping into the deep end of a public swimming pool, where the father waits to catch him. “It’ll be all right, just jump into my hands” encourages the father. The child trustingly jumps, the father misses the catch, the child goes briefly under, the father fishes him out and lands him, coughing and spluttering, on the pool side. He picks the child up, to comfort him, and is quickly

be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this, and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of some special justification, A must do *x* unless B consents to *x*’s not being done” (ibid., p. 208).

joined by his alarmed wife, Harold's mother. Let Updike go on with his story:

His mother swiftly came up to the two of them, and, with a deftness remarkable in one so angry, slapped his father on the face, loudly, next to Harold's ear. . . . His mother's anger seemed directed at him as much as his father. . . . Standing wrapped in a towel near his mother's knees while the last burning fragments of water were coughed from his lungs, Harold felt eternally disgraced. . . . He never knew what had happened . . . by the time he asked, so many years had passed that his father had forgotten. "Wasn't that a crying shame," the old man said, with his mixture of mournfulness and comedy. "Sink or swim, and you sank!" Perhaps Harold had leaped a moment before it was expected, or had proved unexpectedly heavy, and thus had slipped through his father's grasp. Unaccountably, all through his growing up he continued to trust his father; it was his mother he distrusted, her swift sure-handed anger.⁸

It is not really so unaccountable that distrust should be directed not so much at those who once or twice let one down in the most obvious way, who manipulated one or gave one what turn out to be false assurances, but rather at those who prove angrily unforgiving of the letters-down, who do not forgive those who forgive others, who show themselves to be completely reliable punishers of the ones who violate the fidelity principle and even of their forgiving victims. Harold could continue to trust his father (who after all did competently save him after first endangering him), for he had shown the child affection, and manipulated him out of a will to share the fun, flawed only by a faulty estimate of what frolicsome feats were feasible for the pair of them. Incompetence is more easily remedied than ill will, and Harold doubtless learned a little from his sorry experience. (He learned what and what not to expect from his father. Harold keeps on trusting and if need

⁸ John Updike, *Trust Me* (New York: Fawcett Crest, 1988), p. 3.

be forgiving those loved ones who let him down in the well-meaning way his father did.) Harold's mother showed concern for her child and anger at her husband, who had endangered him, along with impressive slapping competence. Was she not a faithful mother and guardian? If trust were simply belief in the dependability of a person to do some range of things, on cue, then we would have to say yes. Harold could count on his mother to attack anyone who harmed or endangered her child. Like a mother cat, or a well-programmed robot, she could be counted on to leap into action to protect her young. But trustworthiness is not just mechanical dependability, and trust is not merely confidence in a range of particular actions in a range of particular circumstances. The trustworthy can show their trustworthiness in surprising ways, and to trust is to be willing to give the trusted the benefit of the doubt when the surprise is, initially at least, unpleasant. *For to trust is to give discretionary powers to the trusted, to let the trusted decide how, on a given matter, one's welfare is best advanced, to delay the accounting for a while, to be willing to wait to see how the trusted has advanced one's welfare.*

As we sometimes but not always wisely delay gratification, so we sometimes can delay knowing or understanding just what others are doing with what matters to us. The pathologies of trust therefore have to include both the truster's bad timing of the demand for an account, and also the trusted's misuse of discretionary powers, both by too-adventurous uses of them (as perhaps Harold's father was guilty of) and also by a refusal to relax some inflexible rule, that is by a refusal to use discretion at all, by simply falling back on reliance on some stimulus-response mechanism, on some automatic pilot, be it instinctive anger or rigid principle. To say, "I can trust him to remember my birthday: he has given his bank a standing order to send the same flowers each year on that date. Short of bank collapse, I can count on it," would be to speak at least ironically, if not sourly. One frequent thing that goes wrong with a personal trust relationship is that it degenerates into

one of mutual predictability. Not merely does this make it boring (as in marriages that freeze into unimaginative, repetitive, and numbingly dependable mutual service), but it also lessens the likelihood that anyone's good is really being furthered by the dependable behavior. For, as Aristotle emphasized, judgment must continually be used when we aim at contributing to someone's well-being. Turning to automatic pilot is not often a serious possibility for those whose goal is the good of another — or even when their goal is their own good. The assurance typically given (implicitly or explicitly) by the person who invites our trust, unlike that typically given in that peculiar case of assurance, a promise or contract, is not assurance of some very specific action or set of actions, but assurance simply that the trusting one's welfare is, and will someday be seen to have been, in good hands.

An institutional example may, at this stage, be a good thing, since I do not want to suggest that it is only in the context of personal relationships that trust is a good, and its diseased variants an evil. There are interesting differences between the trust of intimates, and what is good about it, and the nature and value of more impersonal trust; each is prey to some sickness peculiar to that type, but the main dimensions of fragility are the same, and there are interdependencies between a healthy climate of impersonal trust and the likelihood of a strong trust relationship of a more personal sort.⁹ For good marriages, and for marriages whose ending is not too disastrous for the spouses themselves and for their children, there must be well-functioning background institutions, such as divorce courts, for when marriages break up, and in more normal times schools, legislatures, supreme courts, and regulatory agencies, whose decisions will affect such matters as family planning and the planning of both spouses' careers, not to mention

⁹ See my "Trusting Ex-Intimates," in *Person to Person*, ed. George Graham and Hugh LaFollette (Philadelphia: Temple University Press, 1989), pp. 269–81, for some peculiarities of trust between intimates, and its breakdown.

banks, insurance agencies, and other organizations whose policies will affect the security each spouse can have.

But let us leave the domestic scene and turn to the academic arena. Universities have boards of trustees, to whom the welfare of the institution is entrusted. And an elaborate chain of trust relationships usually goes from these trustees to president, provost, dean, chairpersons, and their appointees within departments. At each of these levels the one in whom trust is placed is not merely a rule applier but a decision maker. Rules there certainly will be, and applying them will not always be such an easy and automatic matter, but no set of university regulations will decide for university administrators everything that they have to settle in their day-to-day activities. When times are tough and cuts have to be made, no rule will tell them what to do, They have discretionary powers — their job is to think about the mission of the university, to listen to all sorts of advisers and affected parties, and then work out priorities as best they can. If they do their job competently and with an appropriately firm will to the good of the institution (seen always in light of its mission), their decisions need not be predictable; they may surprise many, disappoint some, relieve others. Some timid or tired administrators do become predictable in their decision making and to that very extent show themselves unwilling to use discretionary powers and so reveal themselves as unsuitable recipients of our trust. For example, when a cut in spending is judged necessary, a mechanical spreading of it across all administrative units is usually a sure indicator of dereliction of the duty to think about what should be done, rather than a sign of a deep, thoughtful commitment to equality (especially given the sort of units who are being accorded equal treatment). In some institutions such mindless “equal sharing” of burdens becomes the rule, a predictable administration response to any fiscal crisis. But I do not think that anyone would be tempted to give those administrators awards either for superior wisdom or for dogged integrity, let alone for boldness of vision. It is under-

standable that those who do not wish any longer to have to think about the good of their university might fall back on already-made decisions, such as those structuring an institution into various schools or faculties, and then simply say, “let them share the cuts equally,” regardless of their unequal needs, unequal opportunities, or unequal history of previous shares of burdens and benefits and of use made of the benefits. Such a ruling would be a use of discretionary powers that amounts to a refusal to use them. It would be like the policy of the relative whose birthday gift of flowers came with mindless regularity on the right day. In these cases it is not that there is no attention to the good of the person or body at which the person is theoretically aiming. But it is not the sort of attention that would be given by a trustworthy, thoughtful well-wisher. To trust is to let another think about and take action to protect and advance something the truster cares about, to let the trusted *care for* what one cares about. Thoughtless care verges on “careless care,” on plain failure to give care.

It might be thought that where we have answerable officials, who may be removed from office for poor performance of what it is their responsibility to do, no question of trust will arise. University presidents, provosts, and deans, it might be said, are not “trustees.” But it would be too swift to suppose that because we do not call them trustees, no trust and entrusting has gone on. We cannot and do not rely simply on the conditional threat of removal from office to motivate officials to discharge their responsibilities properly. We do hold them “accountable,” but the accountant’s audits, so to speak, will be infrequent. We have no choice but to entrust them with some matters, where constant checking on performance is either impractical or undesirable. Nor does the fact that there is some reliance on the threat of sanctions mean that there will be no room for trust. In our attitude to other people whom we are counting on, we typically combine trust on some matters with careful checks on others.

In dealings with those we know a little, and are willing on that basis to have business dealings with, we typically do partition the matters we have them see to into those where we check up on them and those where we do not. Some failures of such normal business dealings occur because of the fuzziness of the understanding of just where the division falls. If one stands over one's builder, watching and querying every move she makes, she may well refuse to finish the job, since what self-respecting builder would put up with such apparent lack of *any* trust in her professional skill and standards of care? Is she, the builder, not supposed to be the one with the know-how? If the client thinks he knows so much more, why did he hire the builder in the first place — why not hire unskilled workers to obey the client's commands? In some cases the trust dimension pathologically shrinks to near zero, and there is a commensurate expansion of checking and testing. Contracts are the useful contrivances we have for such cases; as much as possible is spelled out, and checks and tests are always in order. Another unhealthy condition is where there is such an exaggerated fear of insulting the other that *any* checking, even on matters where an honest mistake or miscounting is both easy and easily detected without offense, or any request for an account of the trusted's activities regarding what was entrusted is seen as tactless, dangerous to the cooperative relationship.

Rules to guide us on where to trust, where not to, where to insist on precise specification in a contract, where not, are notoriously lacking. We seem to have no choice but to trust our own trust or suspicion on these matters, to check when we harbor suspicions of some bad performance, to trust when we have no such suspicions; to spell matters out in an enforceable contract when we judge that the other bears us “no real kindness,” as Hume put it, and to leave things more casual when we judge that the mutuality and “good offices” are a little “more generous and noble.”¹⁰

¹⁰ David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1975), p. 521.

Some suspicions will be baseless and costly, some contracts regrettable and destructive of fellowship; some trustings will turn out to have been naive and unwise — that is only to be expected. But if the alternative to sometimes giving trust is the policy of trying to check everything out, to protect all one's dealings with others by formal contracts, or the empowerment of Leviathans to stand guard over all of us, then the costs of that policy, especially its opportunity costs, may sensibly persuade us to become like the child Harold in the Updike story, and to take a few letdowns in our stride.

If to trust on a given matter is to leave that matter to the trusted, to see no need, for a while, to check up on how she is doing, to assume that she is doing just fine, that her memory, competence, and good will (Hume's "kindness") are all as one expected when one entrusted the matter to her, then, some might say, only fools ever trust.¹¹ For are not locks and checks always sensible or, at the very worst, a slight waste of time and resources? Might trust itself be pathological? I resist that thought, but I can accommodate the cynics who ask it by allowing that it would usually be foolish, in one's attitude toward a given person on a given matter, not to mix trust on some matters with doubts and prudent checkups on others. It would be offensive to make a surprise visit to check up on the babysitter, but it would be only sensible, with a new, untried one, to arrange to phone during the evening and to stick to that arrangement. If one is asked to trust an ally to defend one from missile attacks and is trusted not to retaliate oneself, what checks should one want and accept? Some evidence, not just of good past performance, but that each party is now managing to do what it is trusted to do and not to do? (*Now* meaning at this minute? *Now* meaning today? The really

¹¹ For a pioneering examination of how some forms of valued trust — in particular, trust in a friend's veracity — may be seen to flout accepted canons of belief-formation, so prompt us to reexamine those canons, see Judith Baker, "Trust and Rationality," *Pacific Philosophical Quarterly* 68 (1987): 1–13.

hard questions are these, the ones involving the details, the ones concerned with the timing of the audits.)

It takes trusters and *their* functional virtues as well as trusted ones and *their* functional virtues, to keep a trust relationship healthy. And most participants will need both sets of virtues, since stable trust is usually mutual trust. Scanlon's principles forbidding manipulation by the deliberate arousing of false expectations, requiring us to take due care not to raise false expectations, to prevent loss to those who rely on our doing what we have encouraged or allowed them to rely on our doing, and to do precisely what we have assured them we will do, are all principles for those who *invite* our trust. They need perhaps to be supplemented by a few principles for the potential trusters, or at least, since trusting is rarely something we *decide* to do, by an analysis of some virtues that are displayed in our dispositions to trust, or to distrust. To the extent that the will is involved, that it makes sense to speak of considering whether or not to trust, what principles could be offered for those considering trusting (or acting as if they trusted), considering whether to continue or resume trusting, considering how much insurance they need to take out against loss contrived or negligently caused by others or by reliance on their false assurances. How much vigilance and checking is worth the cost? And which locksmiths, guards, accountants, and insurers are trustworthy? (The best cover for a burglar is the police force, or the security and insurance professions.)

But what principles can we offer? How many times should we forgive? How much tact must we show? And what principles are there for those who find themselves trusted although they never invited the trust, whose only assurances to the trusting have been their continuing to behave as expected, whose only "Trust me!" has been the failure to issue a warning, "Don't trust me!" Or for those once-trusted ones who find themselves inexplicably mistrusted? Francis Bacon tells us that "base natures, if they find themselves once suspected, will never be true," and we may pre-

sume that nobler natures will rise above suspicion, will not live down to the distrust they may encounter.¹² But what is our advice for those who are instinctively distrusted, and for those who, like Saul Bellow's self-deprecating protagonist in *More Die of Heartbreak*, suspect that "there is something about the slenderness of my face and my glance suggesting slyness. Some people don't feel at ease with me and sense that I am watching them. They suspect me of suspicion."¹³ Such metasuspicions can be as self-confirming as trust in trust, or the will to believe. Their costs are opportunity costs, while the costs of unlucky trust and metatrust can be more dramatic and less easily overlooked. What is the magic formula for the right mix of trust and suspicion, metatrust and metasuspicion?

I am myself skeptical of the chances of success of the Scanlon Kantian enterprise of trying to formulate "valid" moral principles intended to sum up what we expect of participants in a trust relationship. It is likely that we will come up with rules on how many times betrayal should be forgiven, or how distrust is properly focused after "enough" betrayals, or how long insulting distrust should be ignored? (How many rapes should a woman take, before turning against all men? After one rape, how should she focus her future distrust? How should her male acquaintances react if they all become "unfairly" distrusted by her?) I am skeptical both of the insistence that there must be such general moral rules that codify our moral beliefs, and of the assumption that we can establish their validity in a non-question-begging manner, without taking some form of trust on trust. I shall return to this point in my second lecture. So that we can appreciate the full complexity of trust situations that any principles we endorse would have to cover, I shall give two anecdotes, illustrating the vulnerabilities incurred by trust and also showing, I think, the difficulty of formulating any useful rules about how not to misplace our

¹² Francis Bacon, *Remains* (London: Robert Chiswell, 1679), p. 70.

¹³ Saul Bellow, *More Die of Heartbreak* (New York: Dell Books, 1987), pp. 47–48.

trust nor to misuse our capacities for being trustworthy. They are both stories of trust leading to unpleasant surprises.

First Anecdote: A student accepts her tutor's offer of a rental cottage in the West of Ireland during the summer to get on with her dissertation (since steady rain can be expected). The cottage is delightfully situated, delightfully primitive, delightfully isolated; the weather unexpectedly fine. Her farmer landlord comes on foot each evening to bring her milk, and buttermilk so that she can make soda bread on her peat fire. (His wife has instructed her in the finer points of how to do this, since there is no local store, and the staple diet is whole wheat bread, milk, a few locally grown vegetables, and when one is lucky, some locally caught fish.) He brings one or two of his several young children with him on his evening visits, and often stays for a cup of tea and a piece of bread. He is not a great talker, and the children are shy, so there is more companionable silence than conversation on these business-cum-social occasions. A month passes and the time for departure approaches. On the evening of the last milk delivery, the landlord arrives without any accompanying children. As usual, he is offered a cup of tea; as usual he accepts. He lingers longer than usual and makes some conversational moves, offering stories about adventurous young academic women tenants of former years. Has she found the coast and countryside worth exploring? Has she found the cottage acceptable? Yes, she replies, it has all been fine. Was the bed comfortable? Surprised and slightly alarmed at the length and direction of the conversation, she gives a curt affirmative and remarks that it is getting dark, and that he still has a long walk home. At that he rises, and grabs her by the arms, purporting to admire their fine muscles (developed after an unexpectedly athletic and unintellectual summer). She shrugs him off and asks him to leave. He informs her that he cannot possibly, at this hour, walk home by the road, alone—both his and her reputation would be ruined. There are in fact only two houses to be passed on the road to his home farm, and she finds it difficult

to believe that watchers will be waiting there, behind the lace curtains. Nor is her future reputation in the area of very great concern to her. But she humors him, agreeing that he might, as he proposes, take the uninhabited hill track home. Shrewdly sizing up her weak points, he tells her that the trouble with that plan is that, with his stiff leg (an old, badly healed fracture), it will be difficult, perhaps dangerous, for him to clamber up the hillside behind the cottage to where he can join the track across the peat bogs on the ridge. Would she take pity, and accompany him up the hill, to the track? Impelled by visions of him lying on the hill behind the cottage, injured and needing first aid, she reluctantly goes with him up the hill toward the ridge track. Pretending to lurch for needed support, he trips her up with his stiff leg, brings her down on the steep hillside, and attempts rape. Her newly developed muscles come into action fairly effectively, and he is eventually dispatched, to stumble ignominiously home. Even at the time, she cannot help finding the situation slightly comic, and she suspects that the charge “she asked for it” may in this case have some justice. Next day she herself walks to the farm (along the road) to pay the rent and the milk bill to his wife, who keeps the household accounts, and to make her good-byes to the two of them. They ask her to give their kind regards to her tutor, their regular tenant. She assures them she will. This, I suppose, is a story of one forgiving too many, as well as of misjudgment of the extent of “real kindness,” of the undefined limits of normal friendliness, especially of cross-cultural friendliness. The costs were minor, but that was unforeseeable, plain good luck.

Second Anecdote: A young faculty member (the first woman to be appointed in her department) soon after her appointment finds herself the object of amorous attention from two married colleagues, both of them old and close friends of her chairman. He, a charming bully, eventually informs her that he had known all along that it was a mistake to appoint a young, unmarried woman, that only trouble would follow. “They” had, he tells her, passed

around the photo that she had been required to submit with her job application from abroad and discussed how high the risk of such trouble was and whether it was worth taking. Outraged, but also resigned to the inevitability of a few “pioneer” dangers facing women entering professions where the professionals were unaccustomed to having women colleagues and had not yet worked out civilized conventions of coexistence, she does her best to put up with the tense and unhappy working conditions. One of her two married “admirers” considers moving away to a new position, to escape the difficult situation. This provokes his friend, her chairman, to explosions of rage. Accosting the “troublemaker” in a public hall of the university, where faculty are standing around chatting and students are passing on their way to and from classes, he gives her the news of his old friend’s possible departure, then shouts, “See what you have done! Why don’t you get yourself married and out of circulation!” Shocked, she gathers her dignity around her as best she can and immediately writes and submits to the vice-chancellor a letter of instant resignation, complaining of her chairman’s behavior. She is called in by the vice-chancellor and implored to withdraw her letter. After giving thought to the slim chances of finding another university job in midyear, she reluctantly agrees but immediately sets about applying for positions elsewhere for the following year. Her chairman, unapologetic and as far as she is informed unreprimated, continues on his charming macho way. When at the end of the year she leaves for a new job, he makes her an embarrassingly fine farewell present of a first edition of her favorite author, as if to make it difficult for her to keep her grievance alive and well. In this sorry story, the young woman felt let down, but by whom exactly? Not the two admirers, who doubtless let their wives down and made life difficult for their new colleague but, in the latter case, not by breaking any of Scanlon’s principles. Nor did the chairman betray any personal trust his new appointee had in him, since he was so obvious (if likeable) a scoundrel that she had distrusted

him from first meeting and had tried to be on her guard in dealings with him. Nor did the university break any contracts or explicit assurances that it had given. Still, it was the chairman as an officer of the university, and the university community more generally, who were at fault, and the fault was untrustworthiness of a kind that I think Scanlon's principles fail to capture. The victim in this story had trusted her welcome into the university community, taking it to be acceptance as a full colleague. The acceptance had, perhaps, been ambiguous in its quality, and she may be seen to have had "fair warning" of the sexism that was eventually shown so blatantly. But as we all know by now, it is exceedingly difficult to formulate and implement regulations to prevent all such "betrayals" of women. Was this victim deceitfully manipulated? Not exactly. Had she been negligently misled into expecting a nonsexist working environment? By whose negligence? Was there failure to prevent losses she faced because of false reliance on her colleagues' decency? One might construe the efforts to get her to withdraw her proud but imprudent resignation as in accord with, not against, the loss-prevention principle. And since no specific assurances on sexism were ever given, or asked for, the fidelity principle was not infringed. But still, surely, she did suffer letdown.

In these two anecdotes, the disappointed trust may have been a bit silly but not necessarily pathological. For what in each case was the realistic alternative to such trust as was shown? For the tenant to have maintained preemptive distrust of the Irish landlord, from start to finish, was of course a possibility. It would have been possible for the woman faculty member, from the moment of appointment, to have cynically expected the worst of the no-more-than-normally-sexist university — or to have refused a job in such a place, or to have rebuffed all friendly approaches from the men there. But living either unemployed or with sustained watchful distrust of those one sees daily and depends upon in normal daily activities is a high price to pay for avoidance of

ugly letdowns. It is not clear that giving people and administrations the benefit of the doubt, as long as it still is *doubt*, not certainty (while at the same time developing one's muscles), is not the better policy, even given the serious costs of this policy. There are few fates worse than sustained, self-protective, self-paralyzing, generalized distrust of one's human environment. The worst pathology of trust is a life-poisoning reaction to any betrayal of trust. Trust makes life "commodious," in Hobbes's sense, and without it we really are in conditions where our lives will be solitary, poor, and nasty, even if not short or "brutish" (the brutes are in some ways better at trust than we are).

Both these stories focused on trust by a woman, in one case of an individual man, in the other of an institution run by men. Do such tales have any general significance, or is their moral one for women only? Male-female relations can, I believe, serve fairly well to model a wider class of relationships of trust and distrust, where the power of the different parties is unequal, or is shifting and uncertain. The common use of the metaphor of *rape* for any sort of unfair treatment of the weaker, insufficiently vigilant, or inadequately armed, itself suggests that rape serves as a paradigm of a wider class of moral violations. Of course real rape need not involve any abuse of trust. Distrust is no protection against it. But in cases such as date rape and incestual seduction, the unsuspecting trust of the victim is part of the sorry story. The trust-increased vulnerability of the victim and the peculiar trust-dependence of the fragile good entrusted (intimacy with another, closeness that always holds a faint possibility of future mutually sought sexual intimacy) means that what the victim can suffer is not just a grave harm but the poisoning of a once-possible future or an erstwhile good. It can result in at least a temporary allergy to any such goods. These dangers make this case symbolic of other important cases of the abuse of trust. Not everyone can, like Harold and the tough protagonist of my first anecdote, bounce back from letdown, ready for more of the same, for further ad-

ventures in trusting. Bad enough betrayals of trust lead not just to loss of a particular entrusted good but to a lasting inability to partake of that sort of trust-dependent good. And if the trust-dependent goods are the most precious, then that is a severe disability.

In my second anecdote there was no question of rape or rape equivalent, but merely of insult, exploitation of inferior bargaining power, and possibly of conspiracy to maintain this inferiority. The shift in the status of women from a position of exclusion from most professions to reluctant permission to enter, the uncertainties as to whether in being allowed this entry they were also being accorded equality in other historically related matters, such as sexual freedom and sexual initiative, all these provide good models of the pathologies of trust and distrust in conditions where power is shifting, where old monopolies are being challenged, double standards slowly eroded. Problems at the international level, say, between “great” powers possessing nuclear weapons, and upstart lesser powers like Iraq who “dare” arm themselves with equally lethal weapons and dare act as ruthlessly in pursuing their own perceived interests (in particular by invading neighboring territory) as great powers are known to have done (the Soviet Union in Afghanistan, the United States in Panama), present us with a spectacle that, were it not so dangerous in its ramifications, could cause a sour smile or two. The pseudomoral indignation of the powerful when their “inferiors,” especially their recent allies, act as if they too were powerful, as if there were no double standard, is a phenomenon that we should be very familiar with, from many contexts besides the international one, in particular from frequent male reaction to female ambition or “uppityness.” When small, insignificant or medium-sized, less significant allies or former allies, such as New Zealand or Iraq, oppose the will of great powers by banning arms in unacceptable ways or arming in unacceptable ways, great powers feel outraged and even betrayed, as if some understanding about who bows the knee to whom has

been broken. And when, as in the case of Iraq, there was fear of really destructive power, of more lethal equality than was expected, the powerful were at first nonplussed. The “balance of power,” if it leaves a great power less relatively great, will reliably look dangerously upset to such a power.

Relations between those of unequal power are, one would think, what we are most practiced in, since real equality is so rare. So one might have thought that trust and trustworthiness in such relations would be the standard kind. In a way it is — the trust of a child in a loving parent is a standard example of trust. But the distrust of the adolescent is equally paradigmatic and often attended by equal distrust of the adolescent by the parent. Hume noted how upset we can get when inferiors advance upon us, and we are particularly upset when what they advance in is power to make their resentment felt.¹⁴ Growing teen-age children are obvious cases of inferiors in formidable advance. Pathologies of trust occur where there is the will to monopolize and hang on to power, to keep the underdogs under, to prevent inferiors from advancing.

Families usually work out ways of giving increasingly equal voice in decision making to growing children and have coming-of-age ceremonies. Bodies such as the United Nations, and every federal union, have worked out devices for facilitating cooperation and trust between bodies of unequal and shifting relative power—the idea of the right of each state to one vote, regardless of the relative size and power of that state, is a device for empowering the less powerful, for approximating to conditions of equality and so of mutual vulnerability and stable interdependence. Conventions concerning embassies and the treatment of diplomats serve a similar function, to give a voice to every nation however small, even in the capitals of its temporary enemies. As Hume noted, rules of good manners and “mutual deference” serve a purpose similar to the rules of justice and give rise to informal rights or

¹⁴Hume, *Treatise of Human Nature*, p. 377.

dues.¹⁵ All bills and lists of rights empower the less powerful so that they are less vulnerable to the more powerful, so that they can avoid begging for favors.¹⁶ But it takes cooperation, in particular the cooperation of the powerful, to get rights and civilities respected. When great nations give themselves “airs of superiority” and “eagerness for victory,” then even if they do not offend against the law of nations concerning the sacredness of the persons of ambassadors, they insult and offend lesser powers and may in extreme cases provoke in the insulted nation that anger which Grotius likened to the bite of a desperate, ferocious dying beast.¹⁷ It is difficult for nations who are not treated with due respect, as nations, to feel bound by the law of nations (such as the law to refrain from the use of poisoned arms).¹⁸ If the more powerful members of the United Nations veto or disregard its censure when *they* are censured, yet organize military coalitions to give teeth to the censure that they initiate, then the substitution of might by right (or their coalescence) will be only pretense. Until the smaller nations can trust the larger nations to respect the judgment of an international body when it goes against them, as much as when it supports them, no real empowering of the weaker, and no real disempowering of the dangerously strong, will have been effected by any international “one nation, one vote” rule. It is not, then, surprising if smaller nations try to empower themselves by other, more destructive, means. Until there is, on the part of the stronger, exemplary obedience to the rule of international law, respect for

¹⁵ David Hume, *Enquiries*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978), p. 261.

¹⁶ I have argued that rights are a way of avoiding having to beg or be begged from in “Claims, Rights, Responsibilities,” in *Prospects for a Common Morality*, ed. J. P. Reeder and G. Outka (Princeton: Princeton University Press, forthcoming).

¹⁷ Hume, *Enquiries*, p. 261; Hugo Grotius, *The Rights of War and Peace*, chap. 25, sec. 4.

¹⁸ I discuss the special problems that arise when insulted or especially aggrieved peoples and nations resort to terrorism in “Violent Demonstrations,” published in *Violence, Terrorism, and Justice*, R. G. Frey and Christopher W. Morris (Cambridge: Cambridge University Press, 1991), pp. 33–58.

United Nations censure, and for the authority of the world court, there will be no good reason for weaker nations to trust stronger nations. Once the stronger have abused the trust of the weaker, then the burden of proof is on them patiently to demonstrate their goodwill, attempt to show new trustworthiness, should they thereafter want to recover anyone's trust.

The dangers of trust I have so far sketched range from the most obvious, trusting the untrustworthy, to the less obvious, bad judgment as to what matters to check up on and when, what matters to entrust and what to keep under one's own control, bad judgment as to when to give those who have once proved untrustworthy a second or a third or an *n*th chance. Willingness to use discretionary powers is part of what one trusts the trusted to do, and discretion is also a vital part of what the truster needs, discretion in judging when trust is worth its risks, and whether, after some of the risks have eventuated, the best response is indignant complaint and unforgiving withdrawal of trust or whether apologies and new starts are acceptable. Discretionary judgment is also called for once the betrayed have opted for sustained distrust and are about to up and leave, judgment as to whether farewell ceremonies are or are not tolerable. There is a presumption in their favor. As Niklas Luhmann has emphasized, the arts both of tactful invitation and of tactful withdrawal are among the civilizing arts. It is not just that, for all the injured party knows, she might one day want to return and so had better not burn too many boats in her self-righteous storming off. (Or burn too many oil wells in her vindictive retreat.) It is also that there are always innocent parties who are endangered by the expression of (possibly justified) resentment. For example, the Irish landlord's wife of my first anecdote, that patient, long-suffering, and neighborly woman who had instructed her husband's "victim" in bread making, was surely owed a normal farewell, whatever the misdemeanors of her husband. And the college tutor was owed the continued goodwill of her Irish landlord and lady. Had the insulted young faculty

member of the second anecdote actually walked out midyear, her students would have been abandoned mid-course. *They* would have been let down. Trust comes in webs, not in single strands, and disrupting one strand often rips apart whole webs. Sometimes we judge that this has to be done, despite the cost to “innocent” victims. And all of us, as ones caught up in such webs of trust, know that sometimes the abrupt cessations of friendly and mutually trusting relationships mysteriously inflicted upon us by some people can be responses to the offenses of others in the same web. We come to learn to share each other’s penal burdens, and burdens of suspicion. There are times, for all of us, when “they fle from me that sometyme did me seek.”

Discretion is needed here too in judging how long we should put up nonresentfully with unexplained withdrawals of trust, or with sudden failures to meet our trust, even when we believe that the ones whose behavior changed probably had good reason to change, given the letdowns they themselves suffered from third parties. If my faithful mailman, who has always delivered my letters to my door, rather than leaving them in the mailbox at the end of the long drive I share with two neighbors whose houses are nearer to the road, stops this service without warning, it may be because one of my neighbors let her wolfhound run free, and the mailman, who comes on foot, was badly bitten while returning down the drive.¹⁹ If I suspect this possibility, I will of course “forgive” the mailman. But if it turns out that not merely the mailman, on foot, but every other delivery person, bitten or un-bitten, knowing about the dog or not, after as well as before the dog’s owners move away, take to leaving deliveries for me at the end of the drive, although for a while they still deliver to the other

¹⁹ I may appear to have a bit of a hang-up about the fidelity of mail persons (see “Trust and Antitrust,” p. 239, and “Trusting Ex-Intimates,” p. 278), but in fact the international mail service, along with the International Red Cross, is one of the few stable cooperative schemes that function across national boundaries, even in wartime, so faithful mail persons and trustworthy ambulance drivers are quite proper moral paradigms.

houses who share the drive, then it will look as if it is my undemandingness, rather than any other factor, that explains the deterioration of service. It will probably be only a matter of time before the service deteriorates for my neighbors too. As Bacon reminds us: "He that injures one threatens a hundred," so we must take care that our individual willingness to forgive does not put others in danger.²⁰ Overwillingness to excuse untrustworthiness, as well as undue distrust, may not merely deprive me of a good, but may destroy a minisystem, a little network of mutually beneficial expectations. Uncomplaining or automatically forgiving long-suffering invites its own continuation. Demanding one's rights belligerently is certainly one way to destroy trust, but never standing up for them, or not bothering to find out if they are being ignored, is an equally effective destroyer of a network of trust.

Unforgiving rigidity and, at the other extreme, easygoing willingness to keep on forgiving, are both dysfunctional weaknesses, if our goal is to maintain and repair a network of beneficial trust, one composed of normally faulty human persons. Both an unwillingness to be part of such a web, given the real risks of let-down, and a naive optimism in entering it are usually threats to its continued good health. But reliable guidelines on how to judge the risks of trust, how wisely to decide whom to take bread from, and whom to offer it to, are very difficult to find. We may be betrayed not only by those whom we trust but also by overreliance on any of the mixed bag of precepts that our moral tradition offers us, or indeed by any refinements of them that we might concoct. To forgive seventy times seven for the same wrong from the same person, or even from persons of some one group (if only we could be confident how to group people in trust-relevant ways), would be treachery to one's fellows, who could also become victims. To look too hard before one leaps into any cooperative scheme can ensnare one into lonely paralysis, and sometimes spoil the game

²⁰ Bacon, *Remains*, p. 63.

for others. (Nothing venture, nothing win.) To trust any such moral maxim, it seems, would be to be deceived. Quintus Cicero writes: “All things are full of deceit, snares, and treachery. As Epicharmus said, ‘The bone and sinew of wisdom is “Never trust rashly.” ’ ”²¹ But how are we to tell rash trust from wise trust, sensible ventures from silly adventures? There are, as far as I have yet discovered, no useful rules to tell us when to trust or even when we should have trusted. (“Never trust rashly” is an utterly useless rule, if the ghost of Epicharmus will forgive my saying so.) If our Kantian rational capacity to be law-abiders, to apply guiding rules, cannot give us much help here (in the absence of suitably trustworthy rules), and if even our spontaneous mistrust can prove fairly unreliable, then what capacity of ours can we trust, to distinguish rash from appropriate trust? I have already appealed at many points to our powers of judgment, those very powers that we expect those whom we trust to exercise. The truster too must possess them, in order to recognize their presence in others, those to whom she entrusts the care of what matters to her. How can we recognize and develop such skills of judgment? In what conditions are they likely to be shown? My second lecture will address that difficult question.

II. SUSTAINING TRUST

The water is wide, I cannot get o'er
 And neither have I wings to fly
 Give me a boat, that can carry two,
 And both will row, my love and I

Elizabethan song

Thus two men pull the oars of a boat by common convention for common interest, without any promise or contract: Thus gold and silver are made the measures of exchange; Thus

²¹ Quintus Cicero, *Commentariolum Petitionis*, 39-40.

speech and words and language are fixed by human convention and agreement.

David Hume, *Enquiry Concerning
the Principles of Morals*

My first lecture concentrated on the perils of trusting, and the difficulty of finding any rules to help us avoid leaky boats and unreliable rowers, and to navigate our way through life so as to avoid these perils without thereby sacrificing the goods that trust promises, and sometimes delivers. We have to rely on our own judgment and on our ability to improve our judgment if we are to enter and sustain mutually beneficial relationships of trust-involving cooperation. Trust is acceptance of vulnerability to harm that others could inflict but which we judge they will not in fact inflict. Must we develop great psychological insight into the motives of our fellow persons, in order to judge whom to trust? In particular, must we become adept at discerning when our fellow persons' motives are egoistic, when altruistic?

This seems to be the assumption that is made by Bernard Williams in his lead article to the helpful and informative volume entitled *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta. Williams asks what could motivate people to cooperate, where to "cooperate" is taken to involve some of the cooperators letting some other cooperator(s) have the immediate control of some of the actions necessary for the intended cooperative outcome — in my terms, letting someone "take care" of something that matters, leaving it to them. (Israel leaving its defense against Iraq to the United States for a while is a perfect example.) Why on earth, William asks, would rational persons (or nations) make themselves thus vulnerably dependent on others, or take on the burdens of proving dependable when others' good is left in their control? His answer, in short, is that they have no good reason to do this, or at least none we can count on." *Vertrauen ist*

²² Bernard Williams, "Trust Considered: Formal Structures and Social Reality," in Gambetta, *Trusts*, pp. 12–13; the volume is the outcome of a series of seminars at

gut, Sicherheit noch besser, as I learned in Berlin to say.²³ The would-be trusted will dependably sacrifice the good of the trusters to their own perceived good, if it comes to a choice. Neither any macro-motivation, egoistic or altruistic, nor either type of micro-motivation can be relied on to make people into trustworthy cooperators. Egoistic macro-motivation would be shown if, for example, because of fear of some Hobbesian sovereign, or as members of a chain gang, we felt we must cooperate with whatever partners we had been allocated. Altruistic macro-motivation would be shown, Williams believes, if we felt we must keep all our promises. It would be “micro” if, say, we took only promises by and to selected persons, say, Nietzsche’s sovereign free spirits, to be really binding. (As Nietzsche said, such promise-worthy ones are ones whom we fear to let down.) Then, when anyone was given a promise, she would have to judge if the promisor saw her as a free spirit, respect-worthy enough to have it kept. Micro-motivation is sensitive to particular context and to our knowledge of the particular person whom we would have to trust in order to cooperate on a given matter, whereas macro-motivation is general, directed at people of a given, easily recognizable sort, in some recurrent, recognizable type of social context. All possible motivations to cooperate are taken, plausibly enough, to be sensitive to perceived costs, but the micro-motivations are, in addition, sensitive to any special information or special hunches we may have about the given individual who wants us to trust him, and to par-

Kings College, Cambridge, and the other contributors, mostly social scientists, refer to Williams’s essay frequently.

²³ “Trust is good, being sure even better.” My study of trust, begun in “Good Men’s Women: Hume on Chastity and Trust,” *Hume Studies* 5 (April 1979): 1–19, and continued in “Trust and Anti-Trust,” “Trusting Ex-Intimates,” and “Trust and Distrust of Moral Philosophers” (forthcoming in a collection of essays on applied ethics, ed. Earl Winkler), was pursued further while I was a fellow at the Wissenschaftskolleg zu Berlin, in 1988–89. There I gave a lecture on trust and had helpful discussions following it. I was also privileged to discuss the topic with Dieter Claessens, author of *Familie und Wertsystem* (Berlin: Duncker and Humboldt, 1979) and *Instinkt, Psyche, Geltung: Zu Legitimation menschlichen Verhaltens* (Cologne: Westdeutscher Verlag, 1970).

ticular feelings, such as love or fear, that we may have toward him. If we let ourselves become “dependent,” in Williams’s sense, on loved ones, or selected loved ones, by trusting some things that matter to us to their control, that will usually be micro-motivation. If we distrust the secondhand-car salesman just because that is what he is, that will be macro-motivation; whereas if we distrust him on the basis of his especially shifty eyes, that will be micro-motivation. Williams grants that the distinction may not be a sharp one. (We may distrust anyone with eyes like that.)

But in any case it turns out to matter little for Williams which sort of motivation is macro, and which micro, which egoistic, which altruistic, since he concludes that, outside intimate friendship, no one sort can be depended on to do the trick of giving rational persons a motive to become “dependent” in the way a trusting cooperator must, or trustworthy in the way a trusted cooperator should. If anything is to be a “solution” to the “problem” of how societywide trust-involving cooperation is to be motivated, it will, he concludes, be some judicious mixture of motivations. The most promising mixture, he judges, will be a combination of “egoistic micro- and non-egoistic macro-motivations,” in “the combination of egoism and a few constraining moral principles,” such as “keep promises” (or, presumably, “live up to any assurances you have given”). Williams hints that there may be richer non-egoistic macro-motivation than that found in such attempts at minimally principled action; there might be what he says might “very recklessly be called Durkheim’s solution,” but he warns that there are well-known and grave problems about how to prevent encouragement of such non-egoistic macro-motivation from degenerating into “ineffective humbug,” especially when people are “constantly and professedly expressing egoistic micro-motivation in much of their life.”²⁴

After this gloomy opening consideration of the “structural” problems that face any reliance of trust to oil the wheels of our

²⁴ Williams, “Trust Considered,” p. 13.

societywide joint ventures, it is not surprising that many later contributors to the Cambridge symposium are pessimistic about our chances of creating and maintaining socially extensive trust. Geoffrey Hawthorn, in his fine essay “Three Ironies of Trust,” seems equally pessimistic about less extensive trust, more or less limiting its altruistically motivated variants to perfect and perfectly asexual friendship — “the only wholly no-fault relation there is.”²⁵ This verdict of Hawthorn’s is given after he approvingly quotes Shakespeare’s verdict in his 138th sonnet, that love’s best habit is in seeming trust, and

Therefore I lie with her, and she with me
And in our faults by lies we flattered be.

Friendship, provided it can avoid or forgive (or even selectively welcome?) lies, can be no-fault in its acceptance of the risks of trust.²⁶ But lies are not so very easily completely avoided in friendship, and the unwelcome ones are not at all easily forgiven. “Just as there can be treachery in a kiss, so there can be betrayal in an honourable man.”²⁷ The trust of members of an “aristocracy,” originally linked by ties of friendship and by a shared code of honor, may enjoy a brief flourishing but will not last long. (Hawthorn’s first two “ironies” of trust are that socially extensive trust cannot be created except in some sort of “aristocracy”; and that having been so created, it will dependably be eventually undermined by the aristocrats who for a while sustained it.)

I think that the distinguished Cambridge group who studied trust together accepted Bernard Williams’s philosophical analysis of it too trustingly, and one of the questionable aspects of it, high-

²⁵ Geoffrey Hawthorn, “Three Ironies of Trust,” in Gambetta, *Trust*, p. 113.

²⁶ I have discussed the sorts of reasons why deceit might on occasion be acceptable in “Why Honesty Is a Hard Virtue,” in *Identity, Character, and Morality: Essays in Moral Psychology*, ed. O. Flanagan and A. Rorty (Cambridge: MIT Press, 1990), pp. 259–82.

²⁷ Hawthorn, “Three Ironies,” p. 115.

lighted by Geoffrey Hawthorn's use of it in application to the illuminating case studies he presents (of the breakdown of agricultural workers' trust in the Indian Congress party in Northern India, and of "devious" creation in post-World War II Korea of a successful economic system out of extensive mistrust, which was itself employed to police the system), is its sharp contrast between egoistic and non-egoistic motivation. It does seem a little extreme, does it not, to find parental solicitude and perfect friendship between honorable males the only cases of human non-egoistic micro-motivation? And more than a little odd to see commitment to moral principles to be typically motivated by wholly non-egoistic considerations? Any woman whose morality has been painstakingly (and often painfully) overhauled to eliminate all appearance of condoning exploitation knows better than to think that reflective moral commitments lack some element of enlightened egoism.²⁸ Nor is "keep promises" a very suitable example of a moral rule that would have to be adhered to out of non-egoistic motivation, if the force of promises lies either, as Hume thought, in the accepted threat of the penalty of loss of reputation in the event of one's nonperformance (or, as Atiyah thinks, in such participation as promises have in the general force of mutually advantageous exchange.²⁹ Williams is looking for some moral rule that, if observed, would inhibit the more powerful and less dependent parties to a cooperative venture from defecting and so defrauding the more dependent members — a moral constraint that, if effective, would prevent, say, bank officials from doing what Charles Keating and his collaborators did. (This would need to be accompanied by some rule to prevent frightened depositors from withdrawing their savings from other banks, after such a

²⁸ It could be that, in our culture, it is men who typically see mature morality as reflective of *altruistic* constraints on more spontaneously egoistic motivation, whereas women see it as reflective of more *egoistic* constraints on the altruistic caring motivation they have been trained from childhood to feel and display.

²⁹ Hume, *Treatise of Human Arature*, p. 522; P. S. Atiyah, *Promises, Morals, and the Law* (Oxford: Clarendon Press, 1981).

scandal. But wily would-be defrauders can fairly easily avoid giving promises, should they be burdened by compunctions about breaking them. They may simply point to the promises that others, who insure them, are giving. There are very many more or less trusted powerful persons who have given no formal promises to those dependent on them. It is the *inept* politician who makes rash and unnecessary promises, rather than herding his sheep by other, less dangerous means. A promise can boomerang on the promisor — that is the whole point of what Hume called this “artificial contrivance for the convenience and advantage of society” — and it takes the extensive mutual trust of the members of a society to keep this contrivance well serviced, with its coercive teeth in good shape.³⁰ If its implicit threats are not credible, if we reinstate defaulting promisors after a brief and profitable formal penance in the bankruptcy courts, then we have ourselves to blame if promises lose their public force and if politicians and bank managers do not even bother to avoid giving lying promises. “Read my lips” has come to mean “Truster, beware.”

Our actual motivation, in situations where trust comes into play, is not very helpfully seen as a mixture of egoistic and non-egoistic, unless we can be fairly sure which strands are egoistic, which altruistic. But many of our motives resist easy classification in these terms. Is parental concern egoistic or non-egoistic? It is treated by Hawthorn as a rare exceptional instance of non-egoistic motivation, but others, such as Richard Epstein, take concern for the continuers of one’s own selfish genes, plausibly enough, to exhibit a variant of egoism.³¹ Is our pleasure in each other’s company, and our preference for a life that gives us opportunities to

³⁰ Thomas Scanlon, “Promises and Practices,” *Philosophy and Public Affairs* 19 (Summer 1990): 199–226, tells a story about an attempt to get an agreement for the return of a boomerang in exchange for the return of a spear. His example is more apt for his purpose than he may have realized; Hume, *Treatise of Human Nature*, p. 525.

³¹ Hawthorn, “Three Ironies,” p. 112; Richard A. Epstein, “The Varieties of Self-Interest,” *Social Philosophy and Policy* 8 (Autumn 1990): 102–20.

get some such pleasure, egoistic or nonegoistic? Is it egoistic to wish to have the respect of others? Is our will to sustain friendships to be decreed egoistic to the extent that our concern for our friends is for ones who are “second selves” to us? Is the desire for revenge, even when we must bring the temple down on ourselves as well as our enemies, egoistic? Is patriotism a clear case of extended egoism, or is “selfless patriotism” a possibility? The ego’s boundaries are less clearly marked than are most nations’ boundaries, but in the absence of clear boundaries we cannot be sure when our concern is for ourselves alone, when for others. Where does a person stop? Is the skin the ego’s barbed-wire boundary? Is the hair a no-man’s land?³² The clothes mere protection of the border? The reputation an everyman’s land? Even in the area where motivations do seem to be fairly easily classified as “egoistic,” such as personal ambition or vanity about one’s appearance, the “ego” that is thereby tended is a fairly fluidly bounded thing, and apt, like Absalom’s hair, to get entangled with others and caught in the outer fringes of other living things. To trace “the line or hyperplane” that Robert Nozick takes to bound a person’s space is no easy matter, and we need to be clear about when and for what purposes we think that effort worth making.³³ Some important variants of trust do take the form of alliances and other forms of willingness to let others close enough to us to be able easily to invade our “space,” to “violate” us, in the trusting confidence that they will not in fact do this. But what counts as “intrusive,” as coming too close in, varies enormously both from context to context and from culture to culture. One allows one’s physician forms of intrusiveness that one might deny even to a lover. One expects conversational partners not to get so close that

³² Judith Jarvis Thomson’s example of “trespass” in *The Realm of Rights* (Cambridge: Harvard University Press, 1990) is a kiss on the bent bare neck of an attractive fellow library user, a total stranger. Would caressing his hair have been as good an example? (Do we grant our hairdressers liberties with our person?)

³³ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 57.

we get their spit along with their spiel, but as Allan Gibbard notes, this varies considerably from culture to culture.³⁴ (It seems to depend on how far from the equator the conversation takes place. While one might have expected those in colder climates, like emperor penguins, to huddle for warmth, and so to tolerate closeness to nonintimates, in fact it is the warmer climes that have developed human customs allowing the greatest conversational moving in.)

Important though it is for everyone's sake that we do have standards, even if shifting ones, of what counts as invasion of ourselves, and that we be able more or less to count on one another to observe those standards, there are equally important variants of trust that take the form of toleration of absence and distance. To let others on whom one depends be far from one can be as risky as letting them get very close. Roughly speaking, intimates are trusted away from us, strangers are trusted close up.³⁵ But that *is* very rough, since it takes trust to let the strange surgeon go off, taking one's loved one with him, out of one's sight and into the operating theater for emergency surgery (and surgeons, I have found, are on the whole reluctant to let the patient's relatives go with them in there), and there are occasions when it takes more trust to let an intimate up close than to let a stranger. If one has a splinter down one's fingernail, or something in one's eye, one's ham-handed beloved may be the last person one wants to allow to try to remove it. This example suggests that the special risk of close quarters is vulnerability to some particularly painful and easily inflicted forms of harm. Knifing, strangling, and smothering are primitive and "easy" ways of killing selected victims, pro-

³⁴ Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990), pp. 69–70,

³⁵ For some time I took it that infant trust was displayed in willingness to be held close, and to accept food from others ("Trust and Anti-Trust," p. 241). But sociologists and psychologists more plausibly take infant trust to be displayed by willingness to let the parent *out* of sight, trusting that they will return. See E. H. Erikson, *Childhood and Society* (New York: Norton, 1950), and Claessens, *Familie und Wertsystem*, pp. 109–10.

vided one has entry to the victim's presence. Torturing is equally easy if one is close. The forms of harm that can be inflicted from afar are of course equally serious, and can be equally painful, but we do seem to have a special horror of those typically inflicted at close quarters, especially if inflicted by those who not only have been let past our guards, but who themselves were supposed to be our guards. (Think, for example, of Indira Gandhi's assassination.) When we trust another to come close, we let down our guard against these primitive terrors and evils. If betrayed in such a way by such a trusted one, we typically know exactly what is happening but are powerless to stop it. Those we trust away from us have powers to harm us in other ways, and what we then renounce is our ability to keep track of, let alone to control, what they are doing with what matters to us. Up-to-date knowledge is not always control, but it is needed for most forms of control. Those we trust utterly (if any) can come as close as they please behind our defenses and go as far as they please out of our vigilant gaze, without our becoming anxious. The two apparent dimensions of trust, renunciation of guard or defense and renunciation of intelligence, do really seem to be two, neither reducible to the other.³⁶ And our motivation for either type of renunciation seems better characterized as concern for some more or less shared good than as either egoistic or altruistic.

I have been talking of our acceptance of vulnerability to intimates and strangers, but of course a large part of our lives is spent with people who are in neither of those relationships to us, with colleagues, business acquaintances, neighbors, fellow committee members, who to some degree do get behind our guards, and whom we certainly do not keep very close track of. Earlier I said that micro-motivation would be the sort usually involved when

³⁶ I toyed with but rejected the idea that the special vulnerability of closeness comes with a special kind of *being known*, warts, body odors, and all, so that absolute trust would be willingness both to be vulnerably known by the trusted and to be vulnerably ignorant about the trusted. But the torturer's power need not require special personal knowledge of the victim, so I rejected that attempted unification.

intimates are trusted on all the matters, distant and close, where we do trust them. But the contrasting terms *micro-motivation* and *macro-motivation* do not serve to sort our motives any better than *egoistic* and *non-egoistic*, and Williams himself indicates that he has some qualms about this classificatory contrast. He allows that an adequate account would involve “a good deal of elaboration and qualification.”³⁷ As he sketches it, macro-motivation is shown whenever an agent “regularly” performs acts of cooperation of a certain labeled sort, such as fidelity to his promises or, presumably, obedience to the laws of his country or to armed police persons’ orders. These are all cooperative acts and ones that may be regularly displayed, generating patterns of conduct. Micro-motivation, by contrast, will, he writes, be shown in “friendly relations towards a given person,” or, I suppose, in special fear of a particularly brutal-looking police officer.³⁸ But in our friendliest relations with those closest to us we fall into habits of trusting cooperation as regular as those we show with officials and strangers. Up to a certain age, sometimes all too young an age, we trust relatives such as uncles not to make what we later learn to call sexual advances. (Then one may switch to micro-motivation, sizing up one’s uncles one by one.) One gets into bed as usual with one’s spouse, trusting that he has not suddenly succumbed to any brain disease that would turn him into a mad aggressor (as in fact happened to a couple I knew) just as mindlessly as one sits down in the empty seat on the bus, trusting that the person beside whom one seats oneself does not have a flick knife at the ready. (In some cities one does learn to look before one sits, to try to employ “micro-motivation.” Our attitudes and actions in our dealings with persons standing in all degrees of closeness and distance from us fall into pretty regular patterns of habitual behavior and so to some degree show what Williams and others call macro-motivation. We trustingly surrender our passport to whatever person occupies

³⁷ Williams, “Trust Considered,” p. 9.

³⁸ *Ibid.*, p. 10.

the appropriate-looking booth when we cross a country's borders — it does not occur to us to try to check to make sure that this is not some terrorist masquerading as an immigration official, and no more does it occur to us to check to make sure that the dark sleeping shape in the marriage bed is our spouse, rather than, say, some possessor of the ring of Gyges, or some devil in sudden possession of our spouse's body. We take many appearances on trust, and we would go mad if we did not and could not. We trust uniforms, badges, and framed certificates on professionals' walls, all of them fairly easily faked.

It is toward normal appearances, familiar uniforms, and badges of office, as much as to the individual people who are playing their allotted parts in the normal scene, that we should perhaps utter Cicero's solemn incantation "uti ne propter fidemve tuam captus fraudatusve sim! [That I not be deceived and defrauded through you and my confidence in you!]."³⁹ So what does the wise person and the wise society do to guard against false pretenders without giving up on trust? If there are none except unhelpful rules to guide the would-be wise, what help is there? Fortunately our intelligence is not artificial and so we do not have to depend on algorithms, or even on our own skill in coming up with them. We have powers of judgment, and we can use them not merely in case-by-case decisions whether to trust a "Trust me!" invitation (which in any case is rarely a matter of voluntary decision) but in the design and overhaul of institutions, schemes of cooperation, procedures of certification, and procedures for inspecting and repairing all such procedures and schemes. We can invent institutions and procedures as well as algorithms, and we have powers of observation that can tell us how well or badly a given institution or set of interlocked ones are working, whether they are encouraging or discouraging the trust they need for their own best workings. It takes very little observational acumen to realize that a savings-bank system is not working when its elderly

³⁹ Cicero, *De Officiis* 3.70.

clients cut their wrists in their baths on losing the modest savings that they had looked forward to donating to a home for abused children.⁴⁰ A system that punishes trust in bank officials as brutally as that declares itself an obscene failure. Where judgment is tested and given scope is not in the recognition of failures but in the redesign of the system.

Cicero quotes Ennius “Nulla sancta societas nec fides regni est [No fellowship is sacred, no faith kept, where rulership is concerned].”⁴¹ It is power, the opportunity to acquire power after power, and to sustain monopoly of power, that is a proven corrupter of trustworthiness, and so of networks of trust. (I am assuming, of course, that money is one form of power.) What we easily can come to see to be a twin truth to that is that the meekness, servility, and undemandingness of the relatively powerless are equally responsible for this corruption. Domination and obedience, self-promotion and self-abasement, whatever the motives that prompt them, work together to corrupt schemes of cooperation. The practical judgment some of us sometimes, we hope, display is needed to find a way to empower and embolden the relatively powerless and to disempower and humble the dangerously powerful.

Alliances of the relatively powerless alter power distribution in an obvious manner, which blocks monopoly of power. The invention of the concept of a universal right was a piece of practical genius that empowered the less powerful, and there are other such advances to encourage us. Unions of different peoples to recognize and to try to enforce rights — in, for example, the World Court and the United Nations — obviously benefit the weaker, but to see what they offer the stronger takes some breadth of vision, some attention to the fact that eminent power in nations has never lasted more than a few centuries, so roles can be expected eventually to be reversed. The present relative powerlessness of once-mighty

⁴⁰ News item, CBS Evening News, November 29, 1990.

⁴¹ Cicero, *De Officiis* 1.26.

Britain and Austria are instruction in this important lesson. Not just the breakup of empires but the outcome of wars can effect the disempowerment of the once mighty, but this is a dangerous method. Hugo Grotius, discussing that most unequal of situations, the making of a peace treaty and the redrawing of boundaries after a war in which one nation has brought another to surrender, wisely writes: “As in making peace it scarcely ever happens that either party will acknowledge the injustice of his cause, or of his claims, such a construction must be given as will equalize the pretensions of each side, which may be accomplished either by restoring the disputed possessions to their former situation, or by leaving them in that state to which the war has reduced them.”⁴² The victor may by this rule get some territorial spoils of war, but the main aim is to “equalize the pretensions on each side” so that both victor and vanquished can take their place again as functioning members of a community of nations, respecting each other’s rights. Humiliating the vanquished is not among the arts of peace. (And disarming them usually is humiliating them.)

On the home front, the makers of the American Constitution did a fairly good job of designing a system to replace colonial domination and misrule, a system intended to distribute power in such a way that its abuses would be minimized. It obviously is not a perfect system, but at least it shows what can be done in the way of collective creative design (which in this case was minimally helped by guiding rules, since such old rules for constitution-framing as existed were deemed to have been discredited by the bad record of the nations whose constitutions they supposedly blessed). And we can, if it pleases us to do so, try to extract from this last example of relative success a sort of reusable “recipe” for designing lasting schemes of cooperation not just in governing nations but in other spheres—empowerment of the more vulnerable, equal respect, balance of power, provision for amendment, a place for the hearing of grievances, all give us ideas that

⁴² Grotius, *Rights of War and Peace*, chap. 20, sec. 11.

we could try incorporating into rules for the design of other stable schemes of trust-involving cooperation so that all trust would come closer to being mutual trust and so also to being mutual vulnerability.

It is, however, not always exactly *stability* that we want. *Was* the United States constitutional scheme in fact stable? Did it survive its own Thirteenth Amendment, abolishing slavery, or its Fifteenth Amendment, ruling out denial of the right to vote on grounds of “race, color, or previous condition of servitude”? Was the Constitution reconstructed along with the South, rather than simply “amended” by these changes in the basic structures of the social scheme of cooperation? Did the Nineteenth Amendment, which more than doubled the number of participants in the political scheme of government by extending suffrage to women, merely swell the electoral rolls but not alter the basic scheme? Fortunately we do not need to develop hard-and-fast criteria for the identity of a constitution over time. This one changed fairly radically in some of its basic provisions, and it took one civil war and quite a bit of civil disorder from the fighters for women’s suffrage to effect these changes, and it required further civil disorder to make the Fifteenth Amendment really take effect. *Stability* seems not quite the right word for this intermittently tempestuous progress. But Article 5 of the Constitution, the procedure for amendment, did its work, and the First Amendment continues to do its intended job of testing the ground for further amendments. Hence I think we can say that this is a splendid example of the creative design of a self-transforming but not altogether self-repudiating scheme of government, of making the notoriously untrustworthy, the governors, relatively trustworthy. We could amend Nietzsche’s pronouncement and say that some good things provide for their own continuous selective self-overcoming.

It might be queried whether this fairly successful scheme is in fact a scheme of cooperation that requires mutual *trust* for its smooth working. Need the judicial arm trust the executive arm,

or either of them trust the legislature? Do the voters trust their representatives, and do they need to? Are they, the voters, trusted with anything significant by anyone or any agency? It is true that the scheme's proven strength lies to a significant degree not merely in the balance of judicial, executive, and legislative power but in the checks it provides for, so that trusting, in the sense of handing over to the discretion of the trusted, is strictly limited. Amendments such as the twenty-second, limiting presidential terms to two, show proper fear of the abuses that prolonged power would make tempting to a president. Vesting one's interests in governing has been effectively blocked for presidents but not for senators and judges. Particularly with the judicial branch, whose highest officials do not have the check of the prospect of standing for reelection or reappointment to keep them honest, *trust* seems the proper word for what is given. We entrust the interpretation of our laws and our Constitution to these people, for their lifetime, subject only to their "good behaviour" (Article 3), which, for all the Constitution specifies, is to be taken to mean no more than "noncriminal behavior." We surely do have special standards of behavior for Supreme Court judges, but there is no police force to detect putative failures to live up to these standards, no special court to decide on guilt or innocence. There are of course the normal procedures for indictment, but in general we trust our judges to monitor their own trustworthiness. This is a tremendously important case of public metatrust.

Although the Supreme Court is perhaps the most obvious place where, to get the expected cooperative behavior from those on whom others are dependent, we rely not on procedures of investigation and threat of sanctions but simply on these officeholders' trustworthiness, their sense of our sense of the central importance of that particular bit of our total network of social and public trust, and it is not the only case where the Constitution entrusts vital matters, rather than arranging for regular vigilant checks on performance, along with penalty for bad performance. (And even

then, we have to trust the checkers and the penalizers.) Legislators are vulnerable to defeat at the next election for performance that appears unacceptable, but this check does not prevent, and has not prevented, the existence of ongoing schemes of bribery by special-interest groups who can successfully offer their bribes to one after another corrupt congressman. The vulnerability of the bribe accepters to exposure and defeat at the polls does not entail any vulnerability for the bribers — they need not care if their tame congressman is the same one as before, or an equally bribable replacement for him, as long as he is in their control. Is it foolish to entrust our interests to legislators who, as a group, we know from experience are all too ready to be more “influenced” by powerful special-interest groups than by those whose interests they are elected to represent? If the threat of exposure and defeat at the polls does not effectively monitor the good behavior of our lawmakers, and pretty obviously it does not, should we invent some more effective invigilation and penalizing devices? Not necessarily, since then we might have to invigilate the invigilators, and hold penalties over their heads. We trust our lawmakers to develop their own mutual monitoring devices, as we trust other professions (lawyers, physicians) to develop and administer codes of professional ethics. Of course we can and often do exercise our First Amendment rights to protest, to cry “Shame!” and to demand some action to end abuses. But the action that is then best is not necessarily a new device of detection, judgment, and punishment. An adjustment to the network of trust, rather than the withdrawal of trust and its replacement by closer inspection and threats of penalties, might be the more efficient, as well as the “nicer,” solution. Difficult though it may be to invent the right such adjustment, it is no more difficult than inventing a punitive system that avoids inviting abuses of the trust that is inevitably placed at some point in the penalizers, or the policers of the police.

The point of my brief amateur excursion into constitutional law was to illustrate the fact that even in the case where faith is

most likely not to be kept, that is, in government, there are some success stories of the construction and preservation of networks of trust, or rather of judicious mixtures of trust and vigilance. So even if we have no recipe for how to show good judgment as to where to trust, where to invigilate, even if our understanding of human motives and their typology is not very great, we have some instructive precedents, cases where fairly good judgment has been shown. From such precedents we can hazard some informed guesses about the conditions in which trust will be warranted. Ennius's own pronouncement, that rulers and those hoping to be rulers are signally untrustworthy, is itself such an informed rule of thumb. It is the *position* of ruler that is seen to create the danger, not just the personal character of the one who occupies that position, or the motives behind her conduct. The American success story is of the design of positions of power that are not of too much, too long, and so too corrupting power, and of the orderly transfer of power to some whom the earlier powerful ones may have helped rather than hindered in their rise to power. Those who occupy such positions can still be reasonably expected to hold some social bonds sacred and to keep some faiths. The design of a position is itself the design of pressures and influences on any occupier of it. Any president has the watchful press as constant companion and has Congress with its powers (such as war-making powers) to work with — to pressure and be pressured by — and has the National Archives and its historians to remind the president of future verdicts. So even if we lack any useful rules for individuals on when to give and when to withhold trust, we are not entirely without guidelines on how to design roles for individuals that will help them avoid the worst forms of untrustworthiness, or of oppressively burdensome trust, or of overly vulnerable trusting. But it will be the historians, the constitutional lawyers, the international lawyers, the administrative scientists, the economists, the sociologists, the anthropologists, the ethologists, and the psychologists who will have the information that will inform any trust-

worthy rules of thumb of this sort — philosophers in their arm-chairs need plenty of books on their desks that are not purely philosophical and they need colleagues in other disciplines correcting their thoughts if they are to get very far in formulating such indirect guidelines for trusting — guidelines, that is, on the design of social roles that provide the circumstances of appropriate trust.

In the Gambetta volume there is a splendid and reassuringly upbeat essay, “The Biological Evolution of Trust and Cooperation,” by Patrick Bateson. What is cheering about it, after Bernard Williams’s gloomy estimates of our slim chances of achieving or maintaining extensive social trust, is not merely Bateson’s brisk dismissal of the usefulness of the concept of altruism, and its implied contrast with egoism but his bypassing altogether of the question of what *motives* might produce trusting and trustworthy behavior.⁴³ By inspiring stories of the way lichens, sticklebacks, and dunnocks propagate their kind by fairly successful schemes of cooperation, he directs our attention away from the ill-formed question, “What egoistic or altruistic motives might lead agents to behave in cooperative and trust-preserving ways?” to the better question, “In what conditions has cooperative behavior, and in particular trusting cooperative behavior, been known to occur and to have been sustained or replicated?” Now of course in a species like ours, which, unlike the algae and fungi that compose lichens, does have motives and unlike the stickleback and dunnock does reflect upon its motives and consciously adopts and tries to stick by general policies of action, motives will of course be among the actual conditions that will determine what degree and type of cooperative behavior will occur. But they may not be easy for us

⁴³ Patrick Bateson, “The Biological Evolution of Trust and Cooperation,” in Gambetta, *Trust*, p. 19. In “Trust and Anti-Trust” I proposed a test for appropriate trust, namely that the trust relationship be able to withstand mutual knowledge of each party’s motives for trusting and/or trustworthy behavior. Not only was this “test,” as recognized then, unusable, but it also supposed a degree of understanding of human motives that we do not have. Of course *mysterious* motives might still be reassuring to the trust partner, or disillusioning, but it now seems to me that my erstwhile “test” for trust did take self-understanding too much on trust.

to discern and analyze. Other ways of characterizing the conditions for cooperation may be more reliably accessible to us than motive specifications.

Bernard Williams's first "reality condition" is what he calls "Hume's Axiom," that only motivations motivate.⁴⁴ (This is about as helpful as our current insight into determining motives is likely to become.) Now Hume famously described how some of our motives can be shaped or reshaped by the institutions of our society. The "interested passion," he believed, is given what he called "an oblique direction" by conventions establishing property rights, contractual rights, and rights to govern, as well as by accompanying customs such as record keeping and reputation spreading. By now, it is merely an intellectual exercise or game (one that social philosophers seem addicted to) to try to figure out what might have been the "right angle" of self-interest, from which it has, under the pressure of millennia of social evolution, shifted to its present "oblique angle." What we now have is a form of "interest" that is permeated by social and cultural factors — by language, by the inherited network of social roles, by the family as a social unit (with all its advantages and problems, as far as trust is concerned), by our ever-faster and more extensive means of communication of information (about people, about their transactions, about climates of trust), by all our varied devices for assuring and reassuring each other, for insuring against various sorts of loss, and for protecting our varied investments. And our less interested affections, such as benevolence, are equally twisted by cultural forces, such as tax exemptions, and religions that teach that by benevolent acts we pile up for ourselves treasures in heaven.

Hume's Axiom, when understood with due Humean modesty about our limited insight into human motives, does little to limit,

⁴⁴ Williams, "Trust Considered," p. 9. The term *reality condition* is used by Hawthorn, not Williams, but alludes to Williams's subtitle "Formal Structures and Social Reality."

and nothing to classify, the range of our actual motivation. Take, for example, our wish to have others “second” our decisions, policies, attitudes, to get them to agree with us. Whatever its origins, this drive is now shaped by our conversational conventions, our inherited cultural values, perhaps even by fashionable contractualist moral theories. Hume writes of his own moral theory that “it seems a happiness in the present theory that it enters not into that vulgar dispute concerning the *degrees* of benevolence or self-love that prevail in human nature, a dispute which is never likely to have any issue.”⁴⁵ To understand trust, we too will do well to avoid that vulgar unsettleable dispute. To understand it in a way that can help us give and withhold it more wisely, we will need to hazard a well-informed guess about what, in given conditions, we need and do not need to know, as well as what we do and do not need to guard; or rather, about the *relative* importance of certain sorts of knowledge and certain sorts of guards.

I have argued that we do not need to wait until we have expert insight into human motivation, and can recognize “altruistic” motivation should we encounter it, before we can design schemes of cooperation that will encourage both trustworthiness and trust, and can judge the comparative success of different schemes. What motivation does the U.S. Constitution assume in politicians and citizens? Not anything much more detailed than a concern for life, liberty, and the pursuit of happiness, along with a due estimate of the dangers of the form of happiness of those hungry for political power. Must we suppose that any participant in a reasonably well functioning scheme, some human analogue to the way parent birds cooperate to care for their young, will have given some implicit assurances to the other participants who are trusting her to do her bit, and so will be subject to the moral principles that Scanlon articulates, and in particular subject to his “fidelity” principle? Fidelity is certainly the virtue of those who do not let down others when they have encouraged them to trust them, as

⁴⁵ Hume, *Enquiries*, p. 270.

most holders of political office have. Is fidelity adequately analyzed as doing precisely what one assured another one would do, if and when the other wanted that assurance and is relying on it? That understanding of it would account very well for the popularity that the term *fidelity* enjoys in the insurance business, where explicit assurances are sought, given, and relied on. Scanlon's principle is intended to articulate a moral principle that does not presuppose the existence of any particular social practice, such as that which Hume believed gave the words *I promise* their special force, or, presumably, the practice of having presidents take oaths of office, and of dividing our social labor in such a way that some people and some companies have the job of selling others insurance. Hume spoke of "fidelity to promises" as if that were just one special case of fidelity, and he speaks also of "fidelity to the marriage bed" as another such special case. (The marriage ceremony, as he understood it, was unlike typical promises in that it was fixed in content, not flexible to varying individual wills and wishes.) In both these cases explicit assurances are wanted and given, and fidelity consists in living up to them. Both these cases, Hume believed, require the background presence of a social convention giving the words *I promise . . .*, *I take this man . . .*, their binding force, what J. L. Austin called their illocutionary force. Scanlon believes that there is some virtue appropriately called "fidelity" which does not depend on any social custom, but which could be and should be displayed by Robinson Crusoe and Friday, by any two people who can somehow communicate their intentions to one another.

I agree with him, and, I think, with Hume too, that there must be some sort of trustworthiness or fidelity that is possible and desirable independently of the existence of what Hume called "social artifices," such as those instituting private property, or contract, or insurance policies.⁴⁶ Hume took sustained friendship to

⁴⁶ Hume, *Treatise of Human Nature*, p. 603, lists fidelity among the "natural" virtues, that is, ones not consisting in obedience to the rules of a social artifice or practice.

require such a virtue. He is also committed to the virtue of “truth,” which I take to be not merely veracity, but a more general trustworthiness. (The words *truth* and *troth* share common etymological roots.) No social custom, such as that giving the words *I promise* their special force, could ever be established, on the Humean story, unless those hypothetical parties to a Humean convention, who see the need for and the possibility of binding promises, are “true to their word,” that is, to their acceptance of the cooperative scheme inventing contracts.⁴⁷ So *some* general moral virtue, which one might try to spell out in a moral principle, is certainly presupposed both by Hume’s “fidelity to promises” and by “fidelity to the marriage bed.” Has Scanlon captured that virtue in his Principle F? Or has he captured it in that principle, along with his other principles requiring nonmanipulation, due care, and loss prevention? Do the circumstances of sustainable and so appropriate trust include our willingness to conform our conduct to Scanlon’s principles?

In my first lecture I expressed doubts about whether we do sincerely accept moral principles as demanding as Scanlon’s first three and about whether Principle F captures the most basic sort of trustworthiness and fidelity that we hope for from one another. Indeed I expressed doubts that *any* principle we can spell out would capture the full complexity of this core virtue. I suggested that any attempt we might make to show the validity of a principle like Scanlon’s Principle F — or Hobbes’s fifteenth “law of nature” (to give mediators of peace safe conduct) or Hugo Grotius’s version of that law — would turn out to be question begging, since we would find ourselves having to take some form of trust on trust in order to “validate” it. I shall now defend that claim.

Scanlon’s version of moral justification, invoked to apply to his fidelity principle, is spelled out in “Contractualism and Utili-

⁴⁷ I discuss the details of Hume’s account in chapter 10 of my *A Progress of Sentiments: Reflections on Hume’s Treatise* (Cambridge: Harvard University Press, 1991); I discuss the Humean virtue of truth in chapter 12 and also in “Why Honesty Is a Hard Virtue.”

tarianism.” It is the disarmingly simple procedural requirement that we be able to justify to our fellows the rule or policy on which we are acting, where to justify a rule is to show that “no one could reasonably reject [it] as a basis for informed, unforced agreement,”⁴⁸ (Scanlon seems not to mean anything very limiting by “basis for . . . agreement” but simply to be requiring that the principles not be reasonably rejected, however “basic” or less than basic they be relative to other principles. A “basis for agreement” is simply something to agree on.) So what we have is a variant both of what we could call Hume’s Motive, the wish or need to have our fellows second our practical proposals, and of the contractarian thought experiment where basic moral principles are selected by their being agreed on by reasonable people, or rather, for Scanlon, by their not being rejected by reasonable people, who are aware of all the relevant costs and all the relevant benefits (including opportunity costs and benefits) of what they are accepting or rejecting.

Let us imagine Scanlon proposing his fidelity principle as a basis of agreement to all those who have this Humean motivation and who are capable of accepting or rejecting principles, and purporting to do so with reason. Suppose there are some holdouts — call them Friends — who object to a principle that would deprive people of the freedom to do what they judged best when the time came to act, a principle that would require them to do what they had given others to believe that they would do.⁴⁹ The only assurance it is proper to give, these dissidents suggest, is the assurance that we will on every occasion do what we then deem best. Scanlon, let us suppose, tries to reassure them that, even given their

⁴⁸ T. M. Scanlon, “Contractualism and Utilitarianism,” in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (New York: Cambridge University Press, 1982), p. 110.

⁴⁹ Alternatively we could call them Tongans. My Friends have only their name in common with Quakers, members of the Society of Friends, who will not take oaths but will give promises and assurances. Real Tongans, supposedly, once did manage without solemn assurance giving. See Fred Korn and S. R. Decktor Korn, “Where People Don’t Promise,” *Ethics* 93 (April 1983): 445–50.

convictions, they can in reason accept his principle, which does not require that more particular and more limiting assurances ever *be* given, but simply that, *should* they have been given, they be honored. Our Friends, unconvinced, reply that it is not so easy to avoid giving some sort of often-inadvertent assurance to others on specific matters (or to be sure when the second, third, fourth, fifth, and sixth clauses of Principle F do and do not hold good in the real world), so they would in fact find themselves having, on moral grounds, to seem to break this principle — say, when someone has said to them, “You are, I hope, going to Meeting tonight?” and they have replied, “Yes, of course,” but then find that some urgent call on their charitable instincts arises to prevent their going.⁵⁰ At this point the difficulty of knowing what does and does not count as an assurance arises, since, by hypothesis, we are not relying on the use of a special form of words, *I assure you*, as the criterion. Suppose it is agreed that an assurance has in this instance been given. Scanlon can reply that his principle imposes only a prima facie duty, not an all-things-considered one, and that it explicitly allows the assurer to get off the hook if the assured gives consent to his change of plan, a consent very likely to be given to the charitably straying, non-meeting-going Friend, at least by an equally charitable person.

But suppose that our objector, unpacified, patiently replies that the point is that no Friend should want or ask for specific assurances from a Friend, that the freedom to change one’s mind for good reason is of such high moral value that it should not be tampered with; that Friends *trust each other, without special assurance*, to do on each occasion what anyone should always do, namely, what she then judges best. Principle F, from the Friends’ standpoint, offers a standing temptation to take too much thought for tomorrow, to crave a form of security that ought not to be sought and that is the very antithesis of trust. Such cravings for security, they might reasonably say, destroy a climate of trust,

⁵⁰ See note 7.

endanger true Friendship, and are unacceptable curbs on our freedom of morally good will. (*Sicherheit ist gut, Vertrauen noch besser?*) Scanlon may reasonably be skeptical that the Friend really means what she says. “Come now,” he might say, “you can’t disapprove of *all* specific assurances and commitments, you must allow *some* covenants. And if it is the Bible you are appealing to, is not the commandment not to take thought for the morrow accompanied and explained by the assurance that a heavenly father is taking thought for one?” Our Friend may reply, “Yes, indeed, but assurances are a divine creator’s prerogative, not for use by mere fallible and incompletely informed creatures, for whom the liberty to change their mind in the light of new information is a *vital* liberty.”

Scanlon has another reply for the Friend who is skeptical about the wisdom of promulgating his proposed principle of fidelity.⁵¹ He can point out that any who have the sort of conscientious objections to giving assurance that the Friend purports to have can always explicitly disown assurance giving on their own part. They can simply add “but don’t take that as an assurance” to any expression of intention they give to another, just as we often say “I’ll try to be there, but I can’t promise” to deter our fellows from regrettable reliance on our expressed intentions. Could not the Friend, if she accepts the due-care principle, accept the duty to add “don’t take that as an assurance” to any expression of intention that she utters, and so keep her prized freedom to change her mind without letting her exercises of it unduly endanger others? Yes, the Friend might generously accept the onus of disclaiming, if Friends must live in a society where there is disagreement on the value of assurance giving and where the demanding due-care and loss-prevention principles are generally accepted. Constantly making disclaimers will be a nuisance, but not a very great one. She will, however, find herself presented with frequent occasions of temptation when others try to give her assurances. Of course she need not accept

⁵¹ I am grateful to Scanlon for actually making this reply.

and rely on them, but it will ask great strength of will for her to ignore other people's insistent "You can count on me to do X." Scanlon's Principle F is restricted to assurances given to those who want to receive them. If the Friend does *not* want the assurances she receives, then by Scanlon's principle, the assurer will not be bound to live up to the assurances, and so the Friend should not expect that these unwanted assurances will have any binding force. But how does the Friend know whether the assurer was aware whether or not his assurances were wanted? By hypothesis, assuring is not a *social* practice, so we cannot assume that it requires either formal acceptance by the assured, after it has been offered, or a formal request for it, before it is given. Uncertainty is bound to occur frequently. If it is to be avoided, then we will have to impose another verbal inconvenience on our Friends — that of explicit rejection of any assurances offered them. Some gesture meaning "I do not want and do not accept your assurance" will have to become accepted currency. But now we seem to be driven to social practices after all. The question now becomes: given that there is disagreement about the value of assurance giving, should we make it into a fairly formal social practice, or should we institute a social practice (say, wearing *F* on one's clothing if one is a Friend) for indicating that one neither gives nor receives assurances? It seems to me that life goes better if we assume that the onus to bother with formal indicators is on the would-be assurer and the would-be assured, not on others. Scanlon's own wish to free us from unnecessary institutional clutter, to keep the rule-governed social practices to the needed minimum, seems to me to argue for rather than against restricting binding assurances to fairly formally given ones, that is, to argue in favor of treating assuring as a social practice.

But Scanlon may not believe that the Friend really sees no value in assurance giving, really will be seriously inconvenienced by the need to ward off unwanted assurances and to prevent others thinking that she is giving them wanted assurances. He then will

have to decide whether to believe that the Friend speaks sincerely and really *does* believe that, even if we should accept divine assurances, we should give none but very general and fairly empty ones ourselves (and the divine assurance in question is itself about as nonspecific as it could be). Perhaps the Friend deceives herself or, worse, deceives him — she may be merely using her apparently pious and moralistic talk as a mask for a ruthless reformist will to topple a capitalist economy by undermining faith in the value of its sacred moral currency, the assurance, the quasi-contract. (For if too many Friends are around, assurances may indeed come to lose their force.) Might the Friend be a communist agent, who not only does not believe what she has said about divine assurances but does not take any such moral discourse seriously in the first place? Should Scanlon trust her? Unless one trusts one's fellow discussant to be engaged, as she purports to be, in the same enterprise as oneself, in this case that of seeking agreement on morally acceptable principles that can serve as a basis for action and interaction, then the whole justificatory discourse becomes a farce or a contest of wits.

We know that the exchange of moral slogans at the international level can be such a contest, can be mere attempted manipulation and intimidation. Between mutually distrustful parties to a parley, claims about moral acceptability or unacceptability are rightly discounted, not even dignified with the label *hypocrisy*, so little expectation is there that anyone would take them seriously. Until we can trust those with whom we are talking to be doing with words what the form of their words suggests (proposing, counterproposing, raising serious objections, seriously considering the merits of a proposal), no justificatory discourse can be sustained, no principles ratified or vetoed. But if we can trust each other to mean what we seem to say in such a context, then the attempt at validating a basic fidelity principle by enquiring if anyone rejects it, will be superfluous, at least in that context. Some more basic sort of fidelity must already be implicitly recognized

and exhibited in our speech behavior, if our putative acceptances and rejections of principles are to carry any force. If it is not, then nothing that we say about our acceptance or rejection of proposed principles will be worth the wind it is written on. (Similar points, I think, hold good for the validation, and the application, of other candidate “basic” trust-facilitating principles, such as Hobbes’s or Grotius’s laws. The “reason” that shows us the force of such theorems of peace must itself use words, and use them sincerely, nonarrogantly, and nonmanipulatively.)

At this juncture the persistent seeker after validated (universally acceptable) articulated moral principles, might respond that Scanlon’s principle of fidelity, and any other trust-protecting moral laws that we can spell out, do indeed apply to speech itself, not merely to speech acts using the words *I promise*, *I solemnly swear*, *I assure you*, *I guarantee*. Some assurances will be taken to be *tacit*. Kant claimed in his lectures on ethics that any breach of the duty of veracity, the duty to say only what one believes, was also a breaking of a “pact” and so a breach of the duty to keep promises, since any act of speaking itself constitutes an implicit assurance or promise of one’s sincerity made to one’s hearers.⁵² So the principle of fidelity would then indeed apply to any verbal act of acceptance of the principle of fidelity. This, one might say, is desirable reflexivity in a norm, not undesirable question begging in its validation. This response would be, I think, a fair one, but it offers little comfort to Friends or to those who are perplexed about whether there might not be some truth in the Friend’s position as I have given it. It would give no help at all to those trying to decide whether or not to trust those who claim to be mediators of peace, or to decide when to respond to apparent trust with the expected trustworthiness rather than disregarding it or betraying it for the sake of higher values or more important trusts. “Trust those who

⁵² See Immanuel Kant, *Lectures on Ethics*, trans. Louis Infield (New York: Harper and Row, Harper Torchbooks, 1963), p. 228, for Kant’s reported reference to lying as breaking our pact with fellow speakers.

sincerely assure you that they accept the fidelity principle” and “Meet the trust of the trustworthy” can be added to “Don’t trust rashly,” and the rest of the useless rules for giving trust. As an intellectual exercise, it may be satisfying to find principles that apply to their own endorsement, but that scarcely gives those principles a basic role, let alone selects for us which principles, if any, we have reason to endorse, and whose endorsement to take seriously. My hypothetical Friend might endorse “Don’t give assurances” and be perplexed at first when asked, “Are you assuring me that you believe you should not give assurances?” but able, after a minute’s thought, to explain patiently, “No, you should not *need* assurances that I am speaking sincerely, and I am not offering you any. If you distrust me, we should stop the conversation, or talk through some intermediary whom we both trust.” The Friends’ rule can satisfy the reflexivity test as easily as the fidelity principle and, like it, can also be used by manipulative speakers as a snare for the unwary.⁵³

Would there be any point in trying to spell out a more basic principle than Principle F, what we might call Principle T, that hypothetically regulates all our talk when it is trustworthy? What J. L. Austin called the “sincerity condition” of any illocutionary act seems as good as we can hope to get here — the deceptively simple requirement that one actually is doing what one purports to be doing in speech — to be joking, if that is what one seems to be doing; really expressing one’s regrets, if that is what one’s words in their context make it seem one is doing; affirming (possibly by double negatives), if one’s words suggest that; really giving an assurance, if that is what one leads others to think; endorsing, if one purports to be endorsing; translating others’ remarks accurately, if that is what one purports to be doing; and so

⁵³As my sympathetic treatment in *A Progress of Sentiments* of what I take to be a reflexivity test employed by Hume in epistemology and ethics makes clear, I should be the last person to deny the normative relevance of reflexivity tests. They can be relevant, however, without being sufficient.

on.⁵⁴ Austin's is a quite general nondeceit rule, one, I think, that Kant would have been happy with. But its helpfulness as a guide, or even as a standard of criticism, is limited. It insists that we keep the verbal appearances in line with the reality of the speaker's actual intentions. But neither appearances nor reality are here so evident, even to those in the best position to get the evidence. Do rhetorical questions, such as this one, *appear* as questions despite their not being sincere askings, and so have to count as insincere? Do words like *Yeah, yeah*, meant ironically, count as insincere, or do they *appear* as what they are, irony? When Sydney Morgenbesser uttered them with an interrogative intonation from the back of the lecture hall in which J. L. Austin, the inventor of sincerity conditions, had just put forward the thesis that, while we could affirm with double negatives, we cannot deny with a double affirmative, was Morgenbesser affirming, denying, querying, joking, or all of these? "Let your yea be yea, and your nay nay" is *not* what sincerity conditions amount to, given the complex speech acts that we sometimes sincerely want to perform. (Equivocation is a form of speech, and has its role to play.)

Recent philosophers such as J. L. Austin, Paul Grice, and Stanley Cavell have, in Hume's phrase, "exerted their genius" in the philosophical enterprise of trying to make our implicit norms of speech more explicit, and their work helps us to appreciate the complexity of speech norms.⁵⁵ In my view, John Locke got as close as anyone to summing up the essence of our norms for speech when he wrote of the importance of teaching children to show, in their speaking, what he termed "civility"; "that decency and gracefulness of Looks, Voice, Words, Motions, Gestures and of the whole outward Demeanor, which takes in Company and makes

⁵⁴ Austin, *How to Do Things with Words*, esp. lectures 2 and 4 (pp. 12–24, 39–52).

⁵⁵ See *ibid.*; Paul Grice, *Studies in the Way of Words* (Cambridge: Harvard University Press, 1989); Stanley Cavell, *Must We Mean What We Say?* (New York: Scribner, 1969; repr., Cambridge: Cambridge University Press, 1976).

those with whom we may converse easy and well pleased.”⁵⁶ Locke discusses the norm of looking into the face of the one to whom we are speaking, taking it to be a basic rule of speech, more important to impart to learners, he thinks, than all the rules of grammar.⁵⁷ (Shifty eyes are a primitive warning of untrustworthiness, in speech as in other matters.) Speech is our cooperative and trust-facilitated activity par excellence, and speech acts are successful only if they “take in Company,” if they get across to our conversational partners. J. L. Austin’s concept of illocutionary “uptake,” and Grice’s concept of the complex intentions involved in meaningful talk, may be seen to have Lockean roots. We *do* cooperate in speaking, even in our uses of speech to wound and insult. We do trust each other to play more or less according to our unformulated and not fully formulatable (and perhaps shifting) norms of “decency” in speech. And trustworthy speech, like a more or less trustworthy mail service, is one of the tough pervasive webs of trust that we can, if we are clever, use to strengthen and proliferate other life-enhancing webs of trust.

I have spent some time discussing Scanlon’s principles, not because I disagree with his claim that there is a more primitive sort of fidelity and trustworthiness than that involved in conformity to the rules of any social practice, but because of my doubts whether this important virtue is adequately captured in his principle or principles. But had they not come very close to analyzing the content of what we take trustworthiness to be, they would not have engrossed my interest so much. My disagreement is in part with his principles themselves, in part with his belief that principles will capture the sort of trustworthiness we look for in one another, and in part about the places where our speech practices do come in to enable us to trust each other in specific matters. Rules must

⁵⁶ John Locke, *Some Thoughts Concerning Education* (Menston, England: Scholar Press, 1970), sec. 143.

⁵⁷ Intentional eye contact, interestingly, itself exhibits that self and mutual reference of intentions that Grice finds essential to meaningful speech.

be not just supplemented by good judgment but also based upon it if they are to serve us well, and our existent social practices of giving verbal assurances, disclaiming that this is what we are doing, endorsing principles, refusing to endorse them, and accepting and rejecting social practices, all build on our primitive cooperative practices, such as the nonverbal expression, mating, and childrearing, which at a general level we share with the stickleback and the dunnocks; on special versions of nonverbal expression, such as the resigned shoulder shrug, the acquiescent bowing of the head, the supplicatory begging gesture, which we share with apes; and on special human gestures such as encouraging smiles, questioning eyebrow raisings, various forms of eye contact and avoidance of it, and voluntary mutual disablements such as the handshake. We do put trust in most of these gestures, and both the languages we learn and teach and the social practices we learn and teach rest upon our trust in them. They are so natural and habitual to us that it takes a Darwin to get us to notice them and the extent of our reliance on them.⁵⁸ Our more explicit and self-conscious trustings rest on these more primitive ones.

We do not, typically, worry about our fellow's motives for smiles, shrugs, greetings, glances, handshakes, and we do not need to know their motives in order to trust such gestures to be what they seem to be. Their currency is strong and established. Trust as a good is, as Albert Hirschman has pointed out, one of those goods that, like language and like all habits, increase their strength, wear thick not thin, by constant use.⁵⁹ In conditions where there

⁵⁸ I have given them some notice in "Getting in Touch with Our Own Feelings," *Topoi* 6 (September 1987): 89–97, and "Why Honesty Is a Hard Virtue."

⁵⁹This is true also of trust as a commodity, something that makes our living more commodious (and that we have reason to try to produce, get by barter, or even buy and sell, as when the purchaser of a business buys its "good will"). See Partha Dasgupta, "Trust as a Commodity," in Gambetta, *Trust*, pp. 49–72. Hirschman says of trust (among other "moral resources"), "These are resources whose supply may well increase rather than decrease through use; . . . they do not remain intact if they stay unused; like the ability to speak a foreign language or to play the piano, these moral resources are likely to become depleted and to atrophy if not used"

is little or no mutual trust, such as in my thought experiment involving an assurance craving and later a suspicious Scanlon in relation to an assurance-refusing Friend (who might be an impostor), it is hard to see how trust could get started except with the help of a third party, trusted by both the others. Only if trust is already there in some form can we increase it by using what is there to contrive conditions in which it can spread to new areas. Good parents do this when they use the trust that the child already has in them, and in their eyes and gestures, to teach trusting and trustworthy habits of speech, which then become involved in so many other cooperative practices where trust is present.

This is also what we typically do with that trusted nonverbal seal of trust, the handshake. Only because we feel safe in the hand grip of the other can we use this sealer of the bargain, this sacred secular symbol of reciprocal trust in reciprocal services. And why do we trust it? The Romans, I am told, had an arm shake rather than a handshake — they grasped each other by the elbow, thereby immobilizing each other's strong right arm. So the handshake is a remnant of a mutually disempowering gesture — and still does disempower the hand. The secular oath implicitly expressed by it is something along the lines of this: "Should I prove faithless, then may my right hand lose its cunning, as it has at this moment in your hand's grip." But of course the real cunning of the hand, and of the brain behind it, is to use itself thus to limit its own freedom to strike a mean blow, to see when it serves our own long-term goals to make ourselves vulnerable, to disempower ourselves, and to empower others. Arm wrestling continues to be a flourishing bar-room sport, and we know that some have stronger arms, wrists, and hands than others. What, then, can be our motives for ever being willing to shake hands, let alone to use this gesture, this remnant of a contest, to solemnize our mutual assur-

(A. O. Hirschman, "Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse," *American Economic Review Proceedings* 74 119841: 88–96, quoted in Dasgupta, "Trust as a Commodity," p. 56 n. 5).

ances, on those occasions when it is not friends or Friend that we are dealing with, when assurance-giving is in order.⁶⁰ Should we shake hands only with altruists? The one whose hand is offered has indeed, by extending it, shown me that he has no spiked mitt concealed in his palm, but how do I know that his hand will not crush mine, as a prelude to my more total disempowerment? And how do I know that my outstretched hand will be accepted, not treated with suspicion? I do not know, I take it on trust, without any hypotheses about the egoism or altruism of the person to whom I extend my hand. On occasion I may indeed distrust, refuse an offered hand, and try to retreat out of range, rather than advancing toward the would-be handshaker, but not because I suspect him of self-interest. On occasion a formal mutual bow is a safer mutual assurance than a handshake — especially when one of the parties has obviously more power than the other. The handshake is offered between more or less equals ignorant of the details of each others' motivation; the bow or the curtsy (which also momentarily disempowers the potentially aggressive torso, or the striking knee) between unequals who are equally ignorant on this score. The handshake (and to a lesser degree these other deferential self-disempowering gestures) is a mini-case, but a significant one, of our exercise of an ability to *change* conditions in ways that make limited trust a bit less limited, and to do so without great insight into motives. It is also a case of the use of trust to let trust grow, the use of a natural social practice to build up more contrived social practices.

The forms of trust that through their use strengthen and extend trust are usually those, like the handshake and the reliance on more detached third parties as intermediaries, which are reciprocal and between those who have arranged, if need be contrived, some sort of rough equality of power and of vulnerability (or at least some movement in that direction), so as to avoid the force of

⁶⁰For a vindication of the rationality of offering assurances, and getting others to accept them, see David Gauthier, "Assure and Threaten" (Manuscript).

Ennius's dictum. (When a third party is called in, as intermediary, or mediator, the original parties are equally vulnerable to the danger that the "middle term," the third party, is a wolf in sheep's clothing, furthering her own interests, or an agent for a fourth interested party.) In relationships where some governing of the weaker by the stronger is unavoidable — the situation in child care particularly — there are devices whereby the strong can selectively disempower themselves, to be free of the corruptions of power. Solemn vows, the institution of godparenting as contrived supervision of parents by "trustees" of the family (a variant on reliance on a third less "interested" party), and similar arrangements for schools and other places where innocent trust is very easily betrayed unless such measures are taken to disempower the persons in position of immediate parental authority, are all cases, like contract, where by words we fetter ourselves, in this case against the siren voice of manipulative power. As in the handshake we let our hands lose their cunning for a moment, in order that their cunning not be used in ultimately regrettable trust-destroying ways, so in vows and in the entrusting of guardians, trustees, arbitrators, and commissions with some power over us, we use our wily tongues to arrange that they themselves will cleave, as it were, to the roofs of our mouths, whenever they are seen to be used over those in our charge in bullying or overly manipulative ways. Our effort here is to try to see to it that we not be deceived, nor let others be let down, by our unmerited self-trust in our own powers of well-meaning agency. We all know that we should not trust ourselves in positions of predominant power — our corruptibility there is proven and amply documented. So, clever species that we are, we have evolved ways of giving our provenly untrustworthy conduct in such conditions oblique and better directions. We have the handshake and larger versions of it, such as the American Constitution; we have vows of renunciation of various sorts of obviously dangerous power; we have various sorts of acts for empowering more trustworthy or less

temptation-prone others sufficiently for them to supervise us and to halt our heedless use of dangerous degrees of power. (Perhaps we need an international version of godparenting, nations advising other nations on their conduct toward weaker nations.) In such conditions we paralyze ourselves a little, the better to move toward trust-enhancing and trust-increasing forms of cooperation.

But there are other conditions in which our trustworthiness is equally well proven, and so where cautious self-trust is appropriate. We do, after all, in fact pass on to new generations the enabling, empowering, and eventually equalizing arts of speech. We give what Hume called “the new brood” the power to say no to us, to disprove our theories, mock our pretensions, question our values, disempower the father figures, write our obituaries. We encourage *these* “inferiors” to advance on us, even to outdo us. As Patrick Bateson pointed to the facts of successful cooperation between parent birds in the propagation of their kind (whatever speculations we might entertain about bird motives or proto-motives), so we can point to the fact that we propagate our powers of speech, produce new speakers, ready to claim and redefine their rights. Wittgenstein’s question, “Why do we bring up our children?” itself enough to bring to a full stop the philosopher who had been intent on sorting our motives into egoistic and non-egoistic, can be adapted to “Why do we empower them with speech?” We *do* this, and we can trust ourselves to continue doing it. So we have at least some clues about the conditions in which mutual trust is appropriate, and in which even the more powerful can prove trustworthy. When the more powerful welcome and facilitate the growth of power of the less powerful, and delay and deplore the decline of others’ nonthreatening power (say, of the aged), then they usually can trust and be trusted by those less powerful ones.

The modest interim conclusions I draw from my exploration of the circumstances of sustainable and appropriate trust are that the more appropriate questions are: Whom should I trust on a given

matter, myself or someone else? If another, who better than whom, and on the basis of what sort of credentials? If I cannot or do not need to know the details of the other's motives for working with me, in order to judge her trustworthiness, what *would* it be good to know that I have a reasonable chance of being able to find out without unreasonable effort? Given that I am in shifting power relations with those on whom I depend, what sorts of power must I get, or relinquish, in order to work with them to ensure that the positions that some occupy (and that I may someday occupy) are not positions of trust-threatening powerlessness or powerfulness? What changes are needed in our social practices to create better conditions for mutual trust? Granted that we are going to let some others past our guard (or, in bad conditions, let them help form our guard), which others? and for the sake of what goods? Granted that we are not going to keep track of the current doings of everyone whose actions matter to us (let alone to control them), what matters do and do not call for our personal surveillance? What up-to-the-minute knowledge can we do better without? What guards can we do better without? When is barring the throat our best defensive strategy? Given that we will grant some discretionary powers, on what matters should we do so, and to whom in preference to whom? When, if ever, should we ask for an accounting from those we have trusted — when and for how long should we have faith that all will be well?

Philosophers are typically better at questions than at answers, and in any case all these questions have to be asked and answered in concrete circumstances, case by case. Their best answers may turn out to confirm some general thesis, such as that trust is appropriately placed in those who, for whatever motives, welcome the equalization of power, who assist the less powerful and renounce eminence of power, who, when they ask us to delay the accounting of their use of discretionary power, do so for reasons that we will eventually see to have been good (as good parents correctly tell their young children, "One day you will understand,

but not yet. And it is *so that you will one day understand, and be my equal, that I ask you now to trust me*"). But even if such a thesis is correct, the difficult judgment is not the plausibility of such a general thesis, given a range of particular cases of sustained and broken trust, but the acceptance or rejection of the judgments made about the individual cases themselves. The appropriateness of trust, and of sustaining trust, and of supporting institutions that call for trust, is judged case by individual case, not just when trust is given or withheld, but retrospectively, when all the accounts, or enough accounts, are in. If to trust is to be willing to delay the accounting, then, when trust is successfully sustained, some accounts are bound to be outstanding. And as for sustainable accounts of trust, we may have to wait equally indefinitely before we get them.