

INTEGRITY: ITS CAUSES AND CURES

David Luban*

Integrity is a good thing, isn't it? In ordinary parlance, we sometimes use it as a near synonym for honesty, but the word means much more than honesty alone. It means wholeness or unity of person, an inner consistency between deed and principle. "Integrity" shares etymology with other unity-words—integer, integral, integrate, integration. All derive from the Latin *integrare*, to make whole. And the person of integrity is the person whose conduct and principles operate in happy harmony.

Our psyches always seek that happy harmony. When our conduct and principles clash with each other, the result, social psychology teaches us, is *cognitive dissonance*. And dissonance theory hypothesizes that one of our fundamental psychic mechanisms is the drive to reduce dissonance.

You can reduce dissonance between conduct and principles in two ways. The high road, if you choose to take it, requires you to conform your conduct to your principles. That occasionally demands agonizing, sacrificial choices: to resign your job, for example, when continuing to do what your client asks requires you to cheat and shred and cover up. Think of the Enron lawyers. This is a lot to ask of people, particularly when those around you send the message that the actions you object to are nothing more than what grown-ups do to keep a competitive edge in a dog-eat-dog world.¹ In the business world, gaining a competitive edge is universally recognized as good rather than bad, and if it conflicts with Sunday-school morality, those around you will send mixed signals about which you're supposed to

* Frederick Haas Professor of Law and Philosophy, Georgetown University Law Center. For comments on an earlier draft, I would like to thank Sam Fraidin, Neal Katyal, Deborah Rhode, Mike Seidman, and Dan Tarullo.

1. Robert Jackall, in his superb study of moral line-drawing in corporate settings, tells the story of an accountant who got himself fired from a big company for making waves about an internal slush fund. The accountant told Jackall that he was "frightened of losing . . . my self-respect . . . a fear of falling down in a place where you have stuck a flag in the ground and said: 'This is where I stand.'" Robert Jackall, *Moral Mazes: The World of Corporate Managers* 109 (1988). But his colleagues had no pity for him. In their opinion, he broke the rules of corporate life and ignored good excuses for not acting. To them, slush fund was simply business as usual, and the accountant's moralism was abstract and maybe even hypocritical, because nobody rises as high in a corporation as the accountant had without having dirtied his hands. Worst of all, the accountant's moral code made everyone else uncomfortable. "The guy's an evangelist." *Id.* at 109-11.

obey. They will say both, deny the conflict, and leave you to draw your own conclusions about what their denial is supposed to mean. It's hard to maintain the courage of your convictions when your convictions are at war with each other, and those around you say they back one side, but behave as though they back the other.

Taking the high road to integrity may prove unappealing as an intellectual matter as well, because we recognize that hanging on to principles regardless of situations often bespeaks a kind of rigid inflexibility rather than a virtuous soul. Training in the common law teaches lawyers that the meaning of principles gets determined only in their application to specific cases, so that someone who insists on taking general principles literally can be fairly accused of misunderstanding the basic realist insight that rules have no existence apart from cases. For this reason, some leading legal ethicists argue for contextualism, the view that ethical judgment must be sensitive to situational differences that cannot be captured in abstract principles.² Contextualists will not necessarily endorse rigid moralism—and the high road to integrity seems nothing if not rigidly moralistic.

The low road is so much simpler—that, of course, is what makes it the low road. If your conduct conflicts with your principles, modify your principles. This is the path of least resistance; so much so that apparently we follow it unconsciously all the time. That, at any rate, is what fifty years of research in social psychology teaches us. In situation after situation, literally hundreds of experiments reveal that when our conduct clashes with our prior beliefs—when, in the jargon of social science, we act “counterattitudinally”—our beliefs swing into conformity with our conduct, without our ever noticing.³

2. *E.g.*, William H. Simon, *The Practice of Justice: A Theory of Lawyers' Ethics* 9-10, 69-74, 138-39 (1998) (defining and defending a “Contextual View” of legal ethics, which rejects categorical norms in favor of context-sensitive norms); David B. Wilkins, *Making Context Count: Regulating Lawyers After Kaye*, *Scholar*, 66 *S. Cal. L. Rev.* 1147 (1993) (arguing that different contexts of legal practice demand different norms of professional conduct); Geoffrey C. Hazard, Jr., *Context in Ethics*, 2 *APA Newsletters* No. 2, Spring 2003, at 163 (arguing that philosophers' overreliance on categorical norms wrongly neglects the context-sensitive aspects of moral judgment).

3. Both here and in the remainder of this article, my discussion of the impact of social-psychological forces on lawyer behavior draws instruction and inspiration from work of my colleague Donald Langevoort. *See* Donald C. Langevoort, *Behavioral Theories of Judgment and Decision Making in Legal Scholarship: A Literature Review*, 51 *Vand. L. Rev.* 1499 (1998); Donald C. Langevoort, *Ego, Human Behavior, and Law*, 51 *Va. L. Rev.* 853 (1995); Donald C. Langevoort, *The Epistemology of Corporate-Securities Lawyering: Beliefs, Biases and Organizational Behavior*, 63 *Brook. L. Rev.* 629 (1997); Donald C. Langevoort, *Where Were the Lawyers? A Behavioral Inquiry into Lawyers' Responsibility for Clients' Fraud*, 46 *Vand. L. Rev.* 75 (1993) [hereinafter Langevoort, *Where Were the Lawyers?*]. I am in overall agreement with Langevoort; and the principal addition to his work in the present article lies in two observations at the heart of my argument: first, that as a matter of psychological theory, cognitive dissonance theory has a kind of primacy, because it organizes a great deal of the subsequent research into disparate topics; and secondly, that dissonance reduction bears an uncomfortable resemblance to the quest for

I. INTEGRITY AS DISSONANCE REDUCTION

A. Counterattitudinal Advocacy

In one classic dissonance experiment, subjects were asked to perform a boring, repetitive task—rotating screws in holes of a pegboard—and afterwards were paid to tell the next student waiting to perform the same task that it was really very interesting. This is “counterattitudinal advocacy,” known more colloquially as “lawyering.” The reigning behaviorist paradigm in social psychology of the day hypothesized that higher rewards would reinforce the task more strongly, and therefore predicted that the higher the pay, the more likely the subjects were to evince belief in what they were advocating. But dissonance theory yields the opposite prediction. Deceiving one’s fellows for little or no benefit to oneself creates dissonance, and so the “pro bono” advocates should be more likely to internalize the belief they were advocating. And so the experiments proved, to the dismay of behaviorists and the delight of dissonance theorists. Apparently, when my behavior makes me, as Saint Augustine says, “a great riddle to myself,”⁴ I solve the riddle in the simplest way: *if I said it, I must believe it, at least a bit; if I did it, I must think it’s right.*⁵ All this, I emphasize, goes on unconsciously. But the net effect is a happy harmony between what I do and what I believe—the textbook definition of integrity. It is, however, a kind of integrity in which my beliefs always rationalize my actions after the fact, and in which I therefore automatically inhabit the best of all possible moral worlds: the world of my own inevitable righteousness.

Subsequent research has refined the dissonance idea. Experiments reveal that we don’t always resolve dissonance between cognitions by changing our beliefs. Rather, we do so when the dissonant cognitions threaten to undermine our own self-concept—paradigmatically, when it occurs to us that we may have done something wrong.⁶ Apparently, we are all highly resistant to the thought of our own wrongdoing, and the result is that we will bend our moral beliefs and even our perceptions to fight off the harsh judgment of our own behavior. As some psychologists say, we are *intuitive lawyers*.⁷ As usual, Nietzsche

integrity.

4. St. Augustine, Confessions, Bk. 4, available at <http://www.sacred-texts.com/chr/augconf/aug04.htm> (last visited Oct. 18, 2003).

5. See Leon Festinger & James M. Carlsmith, *Cognitive Consequences of Forced Compliance*, 58 J. Abnormal & Soc. Psych. 203 (1959). See also the discussion in Lee Ross & Richard E. Nisbett, *The Person and the Situation: Perspectives of Social Psychology* 66 (1991), and additional sources cited therein.

6. See Elliot Aronson, *The Social Animal* 230-33 (7th ed. 1995), on this refinement of the dissonance idea.

7. See Roderick M. Kramer & David M. Messick, *Ethical Cognition and the Framing of Organizational Dilemmas: Decision Makers as Intuitive Lawyers*, in *Codes of Conduct: Behavioral Research Into Business Ethics* 59 (David M. Messick

put it best. "‘I have done that,’ says my memory. ‘I cannot have done that,’ says my pride, and remains inexorable. Eventually—memory yields."⁸

Of course, a little niggling voice might call this self-deception. Some might find it less than entirely admirable. If so, lawyers for pay can perhaps take comfort from the counterattitudinal advocacy experiment: it suggests that the higher they bill, the less likely they are to deceive themselves into believing what they say on behalf of clients. But they shouldn't take too much comfort, because the difference between them and their pro bono peers is only a matter of degree. Other experiments have shown that counterattitudinal advocacy, whether it be cheap or dear, typically nudges beliefs in the direction of the advocacy. Furthermore, bystanders aren't immune to the same effect. Observers hearing someone give a speech supporting or opposing Fidel Castro will believe that the speaker's own attitude is either pro- or anti-Castro *even if they know that someone else has picked which side the speaker should argue*.⁹ Indeed, observers will draw conclusions about the speaker's belief even knowing that the speaker is merely reading out loud an essay written by someone else.¹⁰ More bizarrely, subjects told that a person being described to them had been assigned to write a pro- or anti-abortion essay will believe that the person favors the conclusion supported in the essay, *without reading the essay*.¹¹ Our tendency to infer attitudes from advocacy-behavior apparently leads us to discount almost entirely information about external constraints on what others are advocating.¹² Add to this finding the fact—about which more later¹³—that what those around us believe influences what we believe, and it should scarcely be surprising that the very process of advocacy tends to swing our beliefs into line with the positions we advocate.

More importantly, well-paid lawyers have people around them, called clients, who may believe strongly in what they tell their lawyer to say—and our beliefs have a tendency to fall in line with those of the people around us. In the famous experiments of Solomon Asch, many people identified a short line-segment as longer than a long line-segment, after having heard other people (who in reality were

& Ann E. Tenbrunsel eds. 1996); see generally Aronson, *supra* note 6, at 175-245.

8. Friederich Nietzsche, *Beyond Good and Evil* § 68 (Walter Kaufmann trans., 1966).

9. Edward E. Jones & Victor A. Harris, *The Attribution of Attitudes*, 3 J. Experimental Soc. Psych. 1 (1967).

10. Arthur G. Miller, *Constraint and Target Effects in the Attribution of Attitudes*, 12 J. Experimental Soc. Psych. 325, 330-33 (1976).

11. Icek Ajzen et al., *Consistency and Bias in the Attribution of Attitudes*, 37 J. Personality and Soc. Psych. 1871, 1874 (1979).

12. See generally Edward E. Jones, *Interpersonal Perception* 138-66 (1990); Edward E. Jones, *The Rocky Road from Acts to Dispositions*, 34 Am. Psychologist 107 (1979).

13. See *infra* Part I.B.

confederates of the experimenter) say that the shorter line was longer. Asch's results have been duplicated in numerous settings, both in and out of the laboratory, and the startling conclusion that bare-bones sense perception can be influenced by the company we keep is robustly supported by the evidence.¹⁴

Lon Fuller noticed the phenomenon in the work-lives of lawyers. He puts the observation in the form of a hypothetical story about a young lawyer, five or six years out of law school,

working in an office with at least, let us say, six other lawyers—perhaps with as many as a hundred or more When he was in law school he used to worry that he might be called upon by his office to advocate causes in which he did not personally believe. He finds that this is not a real problem In those instances where he had some doubts about the client's case at the beginning, these doubts evaporate after he has worked on the case for a few days; his client's cause then comes to seem at once logical and just. He worries a little that he might have experienced the same conversion had he been working on the other side, but this slight concern does not detract from his zeal or his desire to advance his client's interests.¹⁵

In Fuller's shrewdly perceptive fable, we see dissonance-reduction in the direction of advocacy and the socially-induced perception of the justness of the client's cause as two sides of a single phenomenon.

B. *Diffusion of Responsibility and Social Cognition*

The socially-influenced character of perception also helps to explain so-called "diffusion of responsibility," the well-known fact that groups of people are often less likely to respond helpfully in emergency situations than are individuals. (The relevance of diffusion of responsibility to the behavior of organizational lawyers who do nothing about client wrongdoing should be clear.¹⁶) When a college student has an apparent epileptic seizure in the company of five bystanders, he receives help only a third of the time.¹⁷ Does this show heartless indifference? Not necessarily: when the student has a seizure in front of a single bystander, he receives help 85 percent of the time.¹⁸ In a classic set of experiments by Darley and Latané, subjects either heard a crash of glass and a woman screaming from the next room, or witnessed smoke coming through the vent into their

14. See, e.g., Solomon E. Asch, *Effects of Group Pressures upon the Modification and Distortion of Judgments*, in *Groups, Leadership, and Men* 177 (Harold Guetzkow ed., 1951). See Ross & Nisbett, *supra* note 5, at 30-35.

15. Lon L. Fuller, *Philosophy for the Practicing Lawyer*, in *Principles of Social Order* 287-88 (Kenneth I. Winston ed., 1981).

16. See Langevoort, *Where Were the Lawyers?*, *supra* note 3.

17. John M. Darley & Bibb Latané, *Bystander Intervention in Emergencies: Diffusion of Responsibility*, 8 *J. Personality & Soc. Psych.* 377 (1968).

18. *Id.*

own room. When the subjects were alone, most responded to the apparent emergency; but when another person sat next to them and failed to respond, most subjects mimicked the other person and did not respond themselves.¹⁹ Evidently, we respond to situations by checking to see how other people respond, and their response in large measure determines how we perceive the situation and therefore how we ourselves will respond. And of course the phenomenon is reciprocal: as we watch the other, the other watches us. We reinforce each other, in wrong beliefs as well as accurate ones (a phenomenon psychologists call *pluralistic ignorance*). The shaping and reciprocal reinforcement of perception by seeing how others perceive the same thing constitutes the basic phenomenon of socially influenced cognition, or, for short, *social cognition*. Pedestrians stepping around the body of a homeless man collapsed in the street may simply be taking their cues from each other; the evidence suggests that they would stop to help if they were alone. Our moral compass may point true north when we are by ourselves, but place us next to a few dozen other compasses pointing East, and our needle will fall into alignment with theirs—and, in doing so, influence the needles of others' compasses.

It may appear that social cognition theory and cognitive dissonance theory represent two distinct ideas: one, that we conform our beliefs to the perceived beliefs of the people around us, and the other, that we conform our beliefs to our own prior actions. In fact, however, psychological theorists regard them as two aspects of the same theory. Here, the significant idea came from a 1967 paper by Daryl Bem.²⁰ Recall that dissonance theory and behaviorism originally seemed to be rival theories. Bem provided a behaviorist reinterpretation of dissonance. The Achilles' heel of classical dissonance theory consisted of its need to postulate a felt inner tension—the irritating psychological experience of dissonance—that we need to resolve. The problem was that subjects did not report an inner experience of tension, and the theory therefore had to stipulate that the felt tension was unconscious. The unconscious experience of tension thus turned out to be a postulate founded not on observed data, but on the needs of the theory, always an embarrassment in a supposedly experimental science.²¹

19. See generally Bibb Latané & John M. Darley, *The Unresponsive Bystander: Why Doesn't He Help?* (1970); for a literature review, see Bibb Latané & Steve Nida, *Ten Years of Research on Group Size and Helping*, 89 *Psych. Bull.* 308 (1981).

20. Daryl J. Bem, *Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena*, 74 *Psych. Rev.* 183 (1967).

21. *But cf.* Aronson, *supra* note 6, at 235-37 (arguing that dissonance does cause actual psychological arousal). However, the experiments Aronson cites infer the existence of psychological arousal rather than measuring it directly. I suspect that this issue can only be settled by experiments in which subjects are tested for physiological indicators of arousal.

Bem suggested a simpler mechanism than unconscious experiences (unexperienced experiences) to explain the alteration of belief after counterattitudinal action. According to Bem, we have no direct introspective access to our own beliefs. Man, remember, is a great riddle to himself. Instead, we infer our own beliefs in exactly the same way we infer other people's beliefs: by observing our behavior and its context, and reasoning from outward manifestation to inner belief.

In effect, Bem invites us to regard our own self as a sequence of different selves, each one a time-slice of the four-dimensional space-time worm called "me." Inferring my own beliefs from the observed behavior of the society of my own past selves seems, on this interpretation, no different from inferring my own beliefs about the length of a line-segment or the urgency of a scream in the next room from the observed behavior of those around me. Dissonance reduction turns out to be a special case of social cognition.

C. *Recursively Reinforcing Commitment and the Road to Perdition*

One consequence of dissonance theory is that once I act, my beliefs will rationalize the action and therefore impel me to further action of the same sort—which, in turn, calls for renewed rationalization, and further action. Action, we might say, breeds *commitment*, and commitment breeds further action in an ever-steeper slippery slope. The pattern

action → rationalization → commitment → further action

has a recursive character. Psychologists call this the foot-in-the-door effect. In 1966, Freedman and Frazer dramatically demonstrated that persuading someone to place a one-inch-square pro-driving-safety sticker in their window quadruples their willingness to consent two weeks later to having a ferociously ugly, large pro-driving-safety sign in their yard.²²

To moralists, the step-by-step road to perdition forms a familiar trope. "Let me tell you how you will start acting unethically," Patrick Schiltz writes in a well-known article on life in large law firms.²³ It starts with the time sheets, Schiltz tells us—with the moment when you first pad a time sheet just a little bit, intending to pay back the "loan" from the client with a little unbilled time next month. Soon the loans become more frequent, and after a while you lose the desire to pay them back. Then come the lies—first, the white lie about why you missed a deadline, then the darker lie that you carefully proofread the prospectus that you didn't, then the misleading answer to a deposition

22. Jonathan L. Freedman & Scott C. Fraser, *Compliance Without Pressure: The Foot-in-the-Door Technique*, 4 J. Personality & Soc. Psych. 195 (1966).

23. Patrick J. Schiltz, *On Being a Happy, Healthy, and Ethical Member of an Unhappy, Unhealthy, and Unethical Profession*, 52 Vand. L. Rev. 871, 917 (1999).

that you prep your client to give, then the smoking gun document that you don't turn over in discovery. In every case, you will have a rationalization. Speaking, he tells us, from personal experience, Schiltz ruefully observes that "[a]fter a couple years of this, you won't even notice that you are lying and cheating and stealing every day that you practice law."²⁴

C. S. Lewis concurs. According to Lewis, "To nine out of ten of you the choice which could lead to scoundrelism will come, when it does come, in no very dramatic colours. Obviously bad men, obviously threatening or bribing, will almost certainly not appear."²⁵ According to Lewis, the fatal first step on the road to perdition comes simply because you want oh so much to belong to the Inner Ring, or, as we Americans say, the inner circle or in-crowd. A member of the in-crowd offers you a hint of friendship and a glimpse of life on the inside, and tempts you to do something that "we always do"—"and at the word 'we' you try not to blush for mere pleasure"²⁶—that isn't quite kosher. Once you have been tempted,

next week it will be something a little further from the rules, and next year something further still, but all in the jolliest, friendliest spirit. It may end in a crash, a scandal, and penal servitude: it may end in millions, a peerage and giving the prizes at your old school. But you will be a scoundrel.²⁷

The trope, I have said, is familiar—but what strikes me as less familiar is that this pattern, the very picture of lost integrity, exemplifies to social psychologists something close to the opposite—the self's incessant *pursuit* of integrity, of harmony between belief and action. That first small departure from the straight and narrow is like the one-inch-square sticker in Freedman and Fraser's experiments.²⁸ Agreeing to place it in our window leads us to reformulate our self-concept in a way that rationalizes the action, and the new self-concept impels us toward further action of the same sort as its own vindication. What Lewis implies, and Schiltz says outright, is that all the while that you're giving the Devil his due, a little bit more each day, you're also persuading yourself that the Devil is a misunderstood fellow whose hidden virtues are only now becoming transparent to you.

Perhaps the most famous example of the slippery slope to fatal commitment is the Milgram experiments, in which subjects are ordered to administer escalating electrical shocks to another subject in an experiment on the effects of pain on learning. The victim, of course, is a confederate of the experimenter, and the shocks are fake;

24. *Id.* at 918.

25. C.S. Lewis, *The Inner Ring, in They Asked for a Paper: Papers and Addresses* 139, 146 (1962).

26. *Id.* at 146-47.

27. *Id.* at 147.

28. Freedman & Fraser, *supra* note 22.

the real goal of the experiment is to study subjects' responses to destructive orders. Milgram's astonishing discovery, replicated many times in several countries, is that almost two-thirds of the subjects prove willing to go all the way to the end of the sequence of shocks, despite the fact that the victim spends much of the time screaming for the experiment to stop, and eventually falls ominously silent, while the label on the shock-generator reads "Danger: Severe Shock."²⁹ The literature on these famous and alarming experiments is large, and many explanations have been offered for the depressingly high rate of compliance. The one I find most satisfactory focuses on the gradually escalating character of the shocks. The shock-generator has thirty switches, beginning at 15 volts and going up in 15-volt increments to 450. The result is that Milgram's subjects never confronted the pure question "Should I administer this 330-volt shock?" Instead, the question was "Should I administer this 330-volt shock given that a minute ago I administered a 315-volt shock—and I did that after administering twenty previous shocks?" To think "That would be wrong!" virtually requires a subject to conclude that the previous shock, only insignificantly less severe, was also wrong—and cognitive dissonance makes that a very difficult conclusion to accept. It's not that Milgram's subjects became committed sadists. Rather, each shock they administered committed them to the belief that the next shock is neither sadistic nor even wrong (because the prior shocks were not sadistic or wrong). The dissonance-induced commitment leads the subjects unconsciously to gerrymander the boundary between right and wrong.³⁰ Commitment breeds commitment, and leads to overshooting the bounds of reasonableness as outside observers perceive those bounds, but as the subject clearly does not. The subject is keeping faith with his own commitments, and the shock-victim must suffer at the hands of the subject's pursuit of integrity.

D. *Advocacy to Excess*

These reflections on the self-reinforcing character of commitment may help us explain puzzling cases in which lawyers whom one would have expected to dislike a particular cause or client, and to do the bare minimum that competent advocacy requires, instead go the extra mile on behalf of a cause they presumably detest. A celebrated example is Francis Bacon, who at one point in his career found

29. See generally Stanley Milgram, *Obedience to Authority: An Experimental View* (1974); Arthur A. Miller, *The Obedience Experiments: A Case Study of Controversy in Social Science* (1986).

30. That, at any rate, is the analysis of the Milgram experiments I develop (following Milgram's own suggestion) in David J. Luban, *The Ethics of Wrongful Obedience*, in *Ethics in Practice: Lawyers' Roles, Responsibilities, and Regulation* 94 (Deborah L. Rhode ed., 2000).

himself prosecuting his own friend and patron, the Earl of Essex. As Macaulay reports, Bacon

did not confine himself to what would have been amply sufficient to procure a verdict. He employed all his wit, his rhetoric, and his learning, not to ensure a conviction,—for the circumstances were such that a conviction was inevitable,—but to deprive the unhappy prisoner of all those excuses which, though legally of no value, yet tended to diminish the moral guilt of the crime, and which, therefore, though they could not justify the peers in pronouncing an acquittal, might incline the Queen to grant a pardon.³¹

Then, after the execution of Essex, Bacon published a pamphlet to traduce his memory.³² It was this miserable last straw that led Macaulay to his often-quoted rhetorical question about lawyers, “whether it be right that a man should, with a wig on his head, and a band round his neck, do for a guinea what, without those appendages, he would think it wicked and infamous to do for an empire.”³³ But those who quote this celebrated epigram sometimes forget that Macaulay did *not* criticize Bacon for following the role-morality of lawyers—the separation of the personal and the professional. Macaulay had no wish to question the prevailing professional rules. “If . . . Bacon did no more than these rules required of him, we shall readily admit that he was blameless, or, at least, excusable.”³⁴ The problem was that Bacon went so much further than the rules required, as Macaulay goes to great and entertaining rhetorical lengths to emphasize.³⁵ Bacon did everything in his power to ensure that no mercy would or could be shown to Essex. Why? Macaulay thinks he knows the reason. “The real explanation of all this is perfectly obvious The moral qualities of Bacon were not of a high order.”³⁶ According to Macaulay’s diagnosis, Bacon’s “desires were set on things below.”³⁷ Bacon was simply too enamored of “[w]ealth, precedence, titles, patronage, the mace, the seals, the coronet, large houses, fair gardens, rich manors, massy services of plate, gay hangings, curious cabinets.”³⁸

Perhaps so; Bacon would hardly be the first or last lawyer to succumb to such temptations. But before accepting this diagnosis, let us consider a more contemporary example. In 1993, a black lawyer named Anthony Griffin made headlines by representing the grand

31. Thomas Babington Macaulay, *Francis Bacon*, in 2 *Critical and Historical Essays* 290, 314 (1926).

32. “The faithless friend who had assisted in taking the Earl’s life was now employed to murder the Earl’s fame.” *Id.* at 315.

33. *Id.* at 317.

34. *Id.*

35. *See id.* at 318-19.

36. *Id.* at 319-20.

37. *Id.* at 320.

38. *Id.*

dragon of the Texas Knights of the Ku Klux Klan as an ACLU volunteer. The KKK had engaged in a campaign to terrorize black residents who wished to move into an all-white housing project, and the State of Texas attempted to obtain the Klan's membership lists in an effort to prosecute them. The ACLU agreed to represent the Klan's grand dragon, and steered him to Griffin—who, as it happens, not only worked with the ACLU, but was also General Counsel for the Port Arthur Branch of the NAACP, which soon fired Griffin because he was representing the Klan.³⁹

No great mystery exists about why the ACLU—or even a black ACLU lawyer—might wish to defend the Klan's right to keep its membership lists secret. During the civil rights movement, the state of Alabama harassed the NAACP by demanding membership lists, and the Supreme Court's decision in *NAACP v. Alabama* that the lists need not be given to the state stands as a landmark First Amendment protection of political association.⁴⁰ The puzzle is that Griffin went much further than defending *NAACP v. Alabama* by representing the Klan. Apparently, Griffin couldn't stop. He went on to represent the Klan in its attempt to “adopt-a-highway” next to the housing project, a transparent effort to evade a restraining order to keep away from the project.⁴¹ The integrity of the First Amendment hardly required this further representation, any more than Bacon's persecution and assassination of the Earl of Essex had anything to do with the requirements of the prosecutor's role. Yet no one would suggest that Griffin had fallen under the sway of curious cabinets or massy services of plate.

David Wilkins, who has analyzed the ethics of Griffin's representation of the Klan in great detail, gets closer to the truth than Macaulay in explaining the puzzle of counterattitudinal advocacy that goes above and beyond the call of duty in its zeal:

[I]t is virtually impossible for someone in an adversarial role to keep their clients, and more importantly their client's view of the world, at arms length. It is a familiar truth in social science that those who are called upon to support positions that they initially find morally abhorrent will search for ways to reduce the distance between their beliefs and their practices.⁴²

In short, cognitive dissonance strikes again. If Wilkins is right, Griffin—and, I am speculating, Bacon—went the extra mile in their advocacy to keep faith with their initial decision to undertake the advocacy in the first place. Apparently, their integrity called for nothing less.

39. David B. Wilkins, *Race, Ethics, and the First Amendment: Should a Black Lawyer Represent the Ku Klux Klan?*, 63 *Geo. Wash. L. Rev.* 1030, 1030 (1995).

40. *NAACP v. Alabama*, 357 U.S. 449 (1958).

41. Wilkins, *supra* note 39, at 1051-53.

42. *Id.* at 1055-56.

E. *Group Polarization in Adversary Systems*

I have suggested, in the spirit of Bem's version of dissonance theory, a structural similarity between social cognition (the way our beliefs adapt to the beliefs of others) and individual belief-modification (the way our beliefs adapt to our own prior actions).⁴³ The similarity runs in both directions. Not only do we become committed to our own courses of action—out of solidarity, one might say, with the company of our prior selves—but we become equally committed to other members of our own “team” in social competitions. Here, too, the relevance of the research to lawyer behavior should be obvious: the adversary system sets up a social competition in litigation, and the free enterprise system sets up a social competition in transactional practice. Henry Tajfel told subjects that a test of their aesthetic tastes showed that they prefer Klee to Kandinsky—and that bit of misinformation proved sufficient for them to discriminate in favor of others who supposedly prefer Klee, and against those who supposedly prefer Kandinsky.⁴⁴ Similarly, thirty-two young boys were told, after an experiment in visual perception, that they belong to a group that systematically overestimates (or underestimates) the number of dots flashed on a screen. They were then given the task of dividing money among the group. They systematically discriminated in favor of those supposedly in the same perceptual group and against those in the other group—the phenomenon of in-group favoritism resulting from *group polarization*, intimately familiar to participants in adversarial proceedings.⁴⁵ Remember that advocacy makes others think that we believe what we are advocating, even when they ought to know better—and, through a combination of commitment and taking cues from those others, we are likely to believe it ourselves. Group polarization and belief-change in the direction of one's group reinforce one another.

F. *Blaming the Victim*

In a 1973 experiment, subjects were assigned tasks in pairs, in which a “worker” would carry out the task for pay while the “supervisor” would give instructions. They then watched an event in which a supervisor bungled the task, ruining a highly successful effort by the worker. Knowing the results of the other group polarization experiments, it should not surprise us that those who expected to be

43. See *supra* notes 21-22 and accompanying text.

44. See Henri Tajfel et al., *Social Categorization and Intergroup Behaviour*, 1 Eur. J. Soc. Psych. 149 (1971); see also *Social Identity and Intergroup Relations* (Henri Tajfel ed., 1982).

45. Harvey A. Hornstein, *Out of the Wilderness?*, 29 Contemp. Psych. 11, 11 (1984).

workers themselves blamed the supervisor, while those who expected to be supervisors blamed the equipment or circumstances, but did *not* blame the supervisor. More surprisingly, however, when it became clear that the mishap would cost the worker the money he had earned, the other “supervisors” went one step further, and severely disparaged the personal qualities of the worker.⁴⁶

This last phenomenon—blaming the victim—occurs even in settings where us-and-them polarization is not the issue. In an ingenious series of experiments, Melvin Lerner confirmed repeatedly that the worse someone is treated, the more likely observers are to regard the victim as an unattractive, flawed person.⁴⁷ Lerner explains this phenomenon as an unconscious attempt to ward off the scary thought that if unfair treatment can happen to the victim, it can happen to me. We disparage the victim in order to find a distinction, some distinction, between her and us in order to reassure ourselves that we won’t be victimized next.⁴⁸

Lerner’s explanation of blaming the victim has the ring of truth, but I think that it may work in tandem with another phenomenon, more akin to the dissonance-based effects discussed above. Strikingly, out of over a thousand subjects in Lerner’s experiments, not a single one tried to help the victim or walked out of the experiment in protest.⁴⁹ Why not? One explanation, in the spirit of Darley and Latané’s diffusion-of-responsibility research, is that the subjects witnessed the scene of victimization in groups—and groups don’t act as readily as individuals do.⁵⁰ Whether or not that explanation is right, the fact that no subject protested the injustice done to the victim suggests, according to dissonance theory, that these subjects will justify their own inaction to themselves by minimizing the injustice—and the simplest way to minimize the injustice is to denigrate the victim. This, too, is familiar common sense about lawyers: even if lawyers dislike the side they are representing, they often wind up disliking the other side even more, and they relieve their own discomfort at what representing their client requires them to do by the consoling thought

46. Alan L. Chaikin & John M. Darley, *Victim or Perpetrator?: Defensive Attribution of Responsibility and the Need for Order and Justice*, 25 J. Personality & Soc. Psych. 268 (1973).

47. Melvin J. Lerner, *The Belief in a Just World: A Fundamental Delusion* 89-103 (1980).

48. *Id.* at 21.

49. *Id.* at 51-52.

50. The diffusion-of-responsibility phenomenon should not be confused with the well-known economic result, based in the free-rider problem, that groups typically underprovide collective goods. See generally Mancur Olson, Jr., *The Logic of Collective Action* (1965). What Darley and Latané discovered was not the logic of collective action, but the illogic of collective action—that is, they discovered not the rational decision to free ride on others, leading to mutual inactivity, but an irrational overreliance on our perception of what others think, leading to mutual inactivity because nobody wants to make the first move.

that at least the party they are harming is a bad person who has it coming.

G. *The Scripted Self: Playing Roles*

Another important feature of the worker-supervisor experiment is that the subjects conformed their own pro- and con-attitudes to the role they themselves anticipated playing. When placed in the role of worker before witnessing a scene in which a supervisor's mistake costs the worker his pay, one faults the supervisor. When placed in the role of supervisor, one blames the equipment or circumstances and denigrates the worker. Along with the other dissonance-based phenomena discussed above—belief modification,⁵¹ social cognition,⁵² diffusion of responsibility,⁵³ commitment-escalation,⁵⁴ the foot-in-the-door effect,⁵⁵ group polarization,⁵⁶ and blaming the victim⁵⁷—conformity to social role plays a prominent part in our psychic makeup.

Undoubtedly the most famous of all experiments in the power of roles to shape cognitions is the Stanford Prison Experiment ("SPE"), conducted by Philip Zimbardo, Craig Haney, and their associates. Volunteer undergraduate subjects were divided randomly into "guards" and "inmates" in a mock prison. In less than a day, guards began bullying and brutalizing the inmates, while the inmates began to exhibit the pathologies of real-life prisoners, so much so that five had to be released very soon because of "extreme emotional depression, crying, rage and acute anxiety."⁵⁸ By the second day, the prisoners revolted and the guards put down the rebellion by blasting them with fire extinguishers.⁵⁹

The transformation of the subjects almost defies belief. One guard wrote in his diary before the experiment, "[a]s I am a pacifist and nonaggressive individual, I cannot see a time when I might maltreat other living things."⁶⁰ By day five of the experiment, this same student wrote the following in his diary:

51. See *supra* text accompanying notes 44-46.

52. See *supra* text accompanying notes 16-22.

53. See *supra* text accompanying notes 16-22.

54. See *supra* text accompanying notes 23-31.

55. See *supra* text accompanying notes 23-31.

56. See *supra* text accompanying notes 44-46.

57. See *supra* text accompanying notes 47-51.

58. Craig Haney et al., *Interpersonal Dynamics of a Simulated Prison*, 1 Int'l. J. Criminology & Penology 69, 81 (1973).

59. Philip Zimbardo et al., *The Mind Is a Formidable Jailer: A Pirandellian Prison*, N.Y. Times, Apr. 8 1973, §6 (Magazine), at 41.

60. Craig Haney & Philip Zimbardo, *The Socialization into Criminality: On Becoming a Prisoner and a Guard*, in *Law, Justice, and the Individual in Society: Psychological and Legal Issues* 198, 207 (Tapp & Levine eds., 1977).

This new prisoner, 416, refuses to eat. That is a violation of Rule Two: "Prisoners must eat at mealtimes," and we are not going to have any of that kind of shit Obviously we have a troublemaker on our hands. If that's the way he wants it, that's the way he gets it. We throw him into the Hole ordering him to hold greasy sausages in each hand. After an hour, he still refuses I decide to force feed him, but he won't eat. I let the food slide down his face. I don't believe it is me doing it. I just hate him more for not eating.⁶¹

Part way through the experiment, some of the inmates' parents came to visit them, and were horrified by the degraded state their sons had been reduced to. But after a little tough talk from Warden Zimbardo, the parents simply backed down.⁶² Later, the inmates were visited by a prison priest and a lawyer. Like the parents, neither of these professionals had been instructed to act in a role, or had agreed to do so—but both of them did. For example, it simply never occurred to the lawyer (or parents, or priests) to tell the students that they were not inmates, but rather volunteers in an experiment that they could leave at any time. Astoundingly, the prison script seemed to induce everyone to act in role.

II. TAKING STOCK OF SITUATIONISM

The Stanford Prison Experiment seems to portray a world in which the very idea of personal integrity seems absent—a Goffmanesque world where there are no selves, only selves-in-roles, selves who slide frictionlessly from role to role, in each case conforming to the expectations of the role and whatever principles of right behavior come attached to its script. Behind the mask, another mask; behind all the masks, a vacuum; beneath the vacuum, a mask once again. Indeed, some theorists have drawn conclusions very close to the claim that when it comes to character, there's no there there. John Doris, in his recent book *Lack of Character*, concludes from a careful examination of the experimental literature that human character, defined as a fixed set of dispositions toward certain behavior, is

61. *Id.* at 209.

62. Philip G. Zimbardo, *Stanford Prison Experiment: A Simulation Study of the Psychology of Imprisonment Conducted at Stanford University* (1999), at <http://www.prisonexp.org/slide-24.htm>.

Some of the parents got upset when they saw how fatigued and distressed their son was. But their reaction was to work within the system to appeal privately to the Superintendent to make conditions better for their boy. When one mother told me she had never seen her son looking so bad, I responded by shifting the blame from the situation to her son. "What's the matter with your boy? Doesn't he sleep well?" Then I asked the father, "Don't you think your boy can handle this?" He bristled, "Of course he can—he's a real tough kid, a leader." Turning to the mother, he said, "Come on Honey, we've wasted enough time already." And to me, "See you again at the next visiting time."

Id.

largely a myth.⁶³ Gilbert Harman, drawing on the same experimental literature, agrees.⁶⁴ Both draw on an interpretation of the experiments I have been reviewing called *situationism*, defended most cogently by Lee Ross and Richard Nisbett.⁶⁵ At this point, I want to detour briefly to discuss the merits of situationism, in order to distinguish the argument I am offering from a version of situationism that I think is wrong.

The thesis of situationism is, quite simply, that differences in situations account for much more of the observed variation in human behavior than do differences in personality. Our tendency to believe otherwise, that is, to ascribe people's behavior to their personality or character—or, for that matter, their voluntary choices—rather than the situation they are in, is what situationists criticize as the “fundamental attribution error.”⁶⁶ Situationists point, for example, to some of Milgram's findings.⁶⁷ In one variant of his shock experiment, Milgram placed the subject on a team with another subject—actually, of course, a confederate of Milgram. When the confederate uncomplainingly obeyed orders to continue the shocks, 90% of the subjects went along; but when the confederate refused to administer high level shocks and walked away from the experiment, compliance by subjects plummeted to 10%. Clearly, it would be implausible to assume that the subject population in one variant differs radically in propensity to comply from the subject population in the other. Instead, the conclusion must be that situational differences generate this dramatic swing from near-universal compliance to near-universal rebellion. Similarly, Doris points to experiments by Isen and Levin that showed that people who find a dime in the coin return of a pay telephone were vastly more likely to help a stranger pick up papers she has dropped than people who find no dime.⁶⁸ Again, the power of the situation appears to dominate or even dwarf the power of personality and character in determining behavior. The lesson of situationism, drawn by Harman and Doris, seems to be that we have *no* character that disposes us to behave consistently across situations.⁶⁹

63. See generally John M. Doris, *Lack of Character: Personality and Moral Behavior* (2002).

64. Gilbert Harman, *Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error*, 99 *Proc. Aristotelian Soc'y* 315 (1999).

65. See generally Ross & Nisbett, *supra* note 5.

66. *Id.* at 29-89.

67. Doris, *supra* note 63, at 46; Milgram, *supra* note 29 at 116-22.

68. Doris, *supra* note 63, at 30-31 (discussing A. M. Isen & P. F. Levin, *Effect of Feeling Good on Helping: Cookies and Kindness*, 21 *J. Personality & Soc. Psychol.* 384 (1972)). Fourteen of sixteen subjects who found the dime helped pick up the stranger's fallen papers, while only one out of twenty-five subjects who found no dime stopped to help. *Id.* at 387.

69. Doris puts the idea nicely in a chapter entitled “The Fragmentation of Character”:

[I]t's not crazy to think that someone could be courageous in physical but

The disconcerting picture seems to be a near-determinism of situations, in which minuscule differences in the situation—a dime or no dime, the presence or absence of other people in the room—turn into major differences in behavior, and individual idiosyncrasy explains very little of the differences. (It is a *near*-determinism, not a strict determinism, because individual idiosyncrasy still plays some explanatory role.⁷⁰)

I believe that caution is in order about what conclusions to draw from these observations, however. There's no denying the situationists' point that minute changes in situation can dramatically affect the proportion of people exhibiting a given behavior. But the situationists have a hard time explaining why different people behave differently in the *same* situation. In Milgram's basic experiment, two-thirds of the subjects complied, but one-third did not. The point is

not moral extremity, or be moderate with food but not sex, or be honest with spouses but not with taxes. If we take such thoughts seriously, we'll qualify our attributions: "physical courage" or "moral courage," instead of "courage," and so on. Would things were so simple. With a bit of effort, we can imagine someone showing physical courage on the battlefield, but cowering in the face of storms, heights, or wild animals. Here we go again: "battlefield physical courage," "storms physical courage," "heights physical courage," and "wild animals physical courage." Things can get still trickier: Someone might exhibit battlefield courage in the face of rifle fire but not in the face of artillery fire. If we didn't grow sick of it, we could play this little game all day.

Doris, *supra* note 63, at 62 (internal citation omitted). Doris's point is that the quest to explain experiments by reference to personality traits should simply be abandoned.

The salience to legal ethics of this critique of character should be clear, but nowhere clearer than in the bar admissions process, with its requirement of screening applicants on the basis of character. In the leading article on the subject, Deborah Rhode points to the lack of empirical support for belief in character traits that are consistent across situations, and the lack of predictive power of behavior in one situation for behavior in another. She argues, in a fashion similar to Doris, that the experimental findings undermine the entire enterprise of character-assessment, which she believes that the bar should abandon. Deborah L. Rhode, *Moral Character as a Professional Credential*, 94 *Yale L.J.* 491, 555-62 (1985).

70. One must be careful about this question, however. Psychologists sometimes write as though what I am calling "individual idiosyncrasy" is itself a set of deterministic factors—personality traits, perhaps biological in origin, that cause whatever component of behavior the situations do not cause. *See, e.g.,* Lerner, *supra* note 47, at 120 ("[A]s any reasonable psychologist will tell you, all behavior is 'caused' by a combination of antecedent events and the genetic endowment of the individual."). We leave the framework of determinism only when we insist that "individual idiosyncrasy" includes a component of individual choice. Thinking in this way of course lands one squarely in the midst of the problem of free will; but when were we ever out of that problem? Here I agree with John Doris, who argues that situationism pretty much leaves the problem of free will, with its unsettled and unsettling debates between hard and soft determinism, compatibilism and incompatibilism, untouched. Doris, *supra* note 63, at 132-33. *But cf. id.* at 133-46 (Doris's ensuing arguments about ways in which situationism complicates some approaches to the free-will problem). In an earlier paper, I tried to lay out the assumptions necessary for a compatibilist analysis of the Milgram experiments. Luban, *supra* note 30, at 112-16.

more general. Throughout the experimental literature of social psychology we find striking and statistically significant correlations between experimental variables and subjects' responses—but, significant as they may be, the correlation coefficients seldom exceed 0.5, which by the standards of physics is a low correlation, signifying that the manipulated variable accounts for only one-fourth of the variance in behavior. People differ, and the question for situationism is how these differences should be explained.

The answer, Ross and Nisbett tell us, lies in the fact that people construe situations differently, that “it is the situation as construed by the subject that is the true stimulus.”⁷¹ Thus, the differences in response arise not from differences in human character, but rather from differences in perception and construal.

Perhaps. Yet I find this explanation, according to which the one-third noncompliance rate in the Milgram experiment is explained by arguing that the noncompliant third perceived the situation differently from the two-thirds of compliers, to be both too convenient and too ad hoc, given that we don't actually know anything about how Milgram's subjects construed the situation. In Ross and Nisbett's view, what I have called a near-determinism of situations becomes more like a true determinism. Ross and Nisbett localize individual idiosyncrasy in the capacity for perception and construal, while accepting a version of stimulus-response determinism according to which the stimulus (the situation-as-construed) leads subjects to uniform responses. But why? Why parse the individual this way, into a perception/construal capacity that exhibits idiosyncrasy and a responsive capacity that exhibits little or none? To do so seems arbitrary, and borders on downright inconsistency. After all, construing a situation is itself a kind of action, and one would suppose that a consistent situationist should posit that situations account for most of the variation in construals as well as responses. In that case, however, the situationist is left with *no* explanation for variation among individuals placed in the same situation. Moreover, even if the situationist is right about individual variability in construing situations, one can reply that personality lies in large part in our habits of perception and construal, so at least some form of personality theory survives the situationist objection.

I prefer to think of situations—the independent variables that experimenters manipulate—as sources of pressure or temptation. Quite simply, the experiments demonstrate how difficult—but not impossible—acting against the situational tide is. It's so difficult that in the basic Milgram experiment only a third of the subjects are able to bring it off. Adding a compliant teammate makes it more difficult still, so that only one subject out of ten was able to resist; while adding

71. Ross & Nisbett, *supra* note 5, at 11; see generally *id.* at Chapter 3.

a noncompliant team-mate makes resistance easy enough that nine out of ten subjects were able to succeed at resisting the orders to continue administering shocks. In the terms of our initial metaphor, situational changes alter the relative gradient of both the high road and the low road. What the experiments *do* show, quite graphically, is that seemingly-minor manipulations of the environment can cause astonishingly large changes in the ease or difficulty of action, the angle of incidence between the two roads.

Putting the situationists' point in these terms—that is, that situations transform the ease or difficulty of certain courses of action—avoids the implication of determinism. The situation sets conditions under which we choose, but the numbers strongly imply that these conditions do not render choice impossible. Notice that if every subject complied with Milgram's experiment, it would provide evidence that the experiment had uncovered a mechanism akin to a physical reflex, over which we have no choice or control. And if only a few subjects out of the thousand complied, we might regard them as pathological cases, and excuse them from blame on the grounds that they have a screw loose somewhere. The actual two-thirds/one-third split precludes us from drawing either of these deterministic conclusions about the compliers.

Because I resist situational determinism, I resist as well the radical suggestion that our deep-seated propensity to fall into predetermined roles (as in the SPE) means that we have no core self whose integrity matters, only a collection of selves-in-roles, each seeking its own harmony between behavior and principle, but without any larger unity of self. The experiments do show that we lack robust consistency across situations. This should not surprise us, however. After all, if some of my roles impose inconsistent moral demands—if, for example, with a wig on my head and a band round my neck I will be asked to do for a guinea what I would otherwise think it wicked and infamous to do for an empire⁷²—and my daily life leads me to occupy all these roles; and if, further, the actions I take in each role lead me to adopt beliefs that vindicate those actions, then dissonance theory predicts that I will preserve my conception of myself as a morally upright individual in the only way left: by abandoning the belief that my other beliefs should be consistent.⁷³ We purchase integrity, what Gerald Postema calls the “unity of practical consciousness,” at the price of logic, the unity of theoretical consciousness.⁷⁴ The experimental demonstration that we lack robust consistency across

72. See Macaulay, *supra* note 31, at 217.

73. The philosopher Sidney Morgenbesser, asked if he believes in the law of non-contradiction, replied, “Yes and no.”

74. Gerald J. Postema, *Self-Image, Integrity, and Professional Responsibility*, in *The Good Lawyer: Lawyers' Roles and Lawyers' Ethics* 286, 296 (David Luban ed., 1983).

situations shows that integrity consists of a complex unity, stitched together with a great deal of self-deception that allows us to deny inconsistencies and the dissonance they induce. Integrity remains something that we seek. The problem, then, remains the one we began with: that the quest for integrity, manifested in all the psychological phenomena we have been reviewing, can drive us to behavior as disconcerting and morally repellent as that shown in the Stanford Prison Experiment or in Milgram's demonstration.

III. THE DIFFERENCE BETWEEN INTEGRITY AND DISSONANCE REDUCTION

I do not suppose that I am reporting anything novel or recondite. All of these experiments are widely known, and the Milgram and Stanford experiments are famous, almost legendary. My argument, which is perhaps less familiar, is that these are not simply an array of discrete, unrelated psychological curiosities. Rather, they are aspects of cognitive dissonance, in its social and individual guises. They emerge, therefore, from our drive toward inner harmony—our drive toward integrity. The quest for integrity kills, and in killing it leaves the survivors with their own sense of rectitude intact, like a tattered flag flapping in the wind over the fallen.

To all this there is a simple reply: *You are not talking about genuine integrity.* Integrity does not consist of molding and adapting one's principles to whatever behaviors we and those around us find convenient. Integrity consists of taking the high road, the road of conforming our behavior to our principles. I mentioned earlier that the word 'integrity', like 'integer', 'integral', and 'integrate', comes from the Latin *integrare*, to make whole. That word, in turn, derives from *in-*, "not", plus *tangere*, "touch." An entity is whole if it is untouched, unsullied; the Latin *integer vitae* meant innocent, pure, blameless in life.⁷⁵ And thus the person of integrity is not merely the person whose principles and behavior harmonize, regardless of how that harmony gets achieved, but rather the person who has kept her principles intact ("intact" is another word whose Latin root means "untouched"). We think of the person of integrity in the terms C. S. Lewis uses in *Perelandra* to describe Ransom, his protagonist: "even if the whole universe were crazy and hostile, Ransom was sane and wholesome and honest."⁷⁶ His moral compass never turns from true north, no matter how many other compasses point elsewhere.

Of course, as an analysis of the concept of integrity this must be right. When we are done in by situational forces that distort our moral judgment, we are hardly "untouched." Just the opposite: we are all too touched. The low road to integrity is simply not the same as

75. See Horace, Odes and Carmen Saeculare, I:22.

76. C.S. Lewis, *Perelandra* 6 (1944).

the high road, and bending your principles to rationalize your actions is not the same as bringing your actions into conformity with your principles.

The problem, however, lies in telling them apart *from the inside*. As I noted above, merely asserting categorical principles and refusing to deviate from them regardless of the situation we find ourselves in may be Ransom-like integrity, but it may also be an inability to learn from experience, a kind of fatal priggishness and narrow-minded inflexibility. Every normal life contains episodes of learning from experience, during which principles are reinterpreted, contexts are distinguished, and precommitments modified, along with episodes of sticking to your guns, drawing lines in the sand that you will not cross, and keeping faith with your ideals. This implies that a life of integrity—the high-road, genuine kind of integrity, not the low-road, ersatz kind—will normally contain episodes in which pre-existing moral judgments get discarded in the face of experience. Ransom may always have been sane and wholesome and honest, but his judgment of what particular behaviors are sane, wholesome, and honest may well have changed between the ages of fifteen and fifty. One supposes and hopes that they did. His fifteen-year-old self might view some of the fifty-year-old's beliefs as sold-out ideals, where the fifty-year-old Ransom sees a story of growing wiser (and maybe sadder). From the agent's point of view—from the inside—how do you tell which is which?

One plausible answer is that genuine integrity consists not simply of adherence to principles, but adherence to the right principles, or at any rate to reasonable principles. As Deborah Rhode puts it in her contribution to this symposium,

At a minimum, persons of integrity are individuals whose practices are consistent with their principles, even in the face of strong countervailing pressures. Yet the term also implies something more than steadfastness. Fanatics may be loyal to their values, but we do not praise them for integrity. What earns our praise is a willingness to adhere to values that reflect some reasoned deliberation, based on logical assessment of relevant evidence and competing views. Some theorists would add a requirement that the values themselves must satisfy certain minimum demands of consistency, generalizability, and respect for others.⁷⁷

That may be true in a formal sense, but from the agent's point of view the formula won't help. For one thing, many of the

77. Deborah L. Rhode, *If Integrity Is the Answer, What Is the Question?*, 72 Fordham L. Rev. 333, 335-36 (2003). In support of this point, Rhode cites Sharon Dolovich, *Ethical Lawyering and the Possibility of Integrity*, 70 Fordham L. Rev. 1629, 1654-56 (2002), who in turn refers to Mark S. Halfon, *Integrity: A Philosophical Inquiry* 32-33, 133-37 (1989) and Jody L. Graham, *Does Integrity Require Moral Goodness?*, 14 Ratio 234, 244 (2001).

psychological forces discussed here leave our principles untouched, instead affecting our judgment of whether or not a case falls under a principle. It seems likely, for example, that Milgram's compliant subjects believed before, during, and after the experiment that it is wrong to inflict undeserved suffering on the innocent. What changed was their perception of whether the 330 volt shock they were administering was an instance of inflicting undeserved suffering on the innocent. Even when cognitive dissonance reduction causes a change in values, it will not help the agent to be told that integrity consists in adherence to values that are right and reasonable, because the agent knows only what is right and reasonable to her, and what seems right and reasonable to her may have been corrupted by the psychological forces we have been examining.

But perhaps matters are not as hopeless as they appear. Many of the phenomena revealed by experimentalists are short-lived aberrations, recognized as such even by the subjects once the spell wears off. As Doris puts it, the motives induced by the experiments "are not readily enmeshed in . . . biographies; they look like psychological tics or glitches."⁷⁸ Asch's subjects did not continue to perceive shorter lines as longer after they left the experiment. Nor did most of Milgram's compliant subjects continue to believe that compliance was the right thing to do once the experiment ended and they talked it over with the experimenter and the man they had supposedly been shocking. In fact, they probably *never* believed that compliance was the right thing to do. When Milgram described his experimental set-up to audiences and asked them to guess the rate of total compliance, and whether they themselves would comply, most guessed around one percent (as compared with the sixty-five percent compliance rate in the actual experiment); and no one believed that they themselves would comply—an unmistakable sign that normal people believe compliance would be wrong. We have no reason to suppose that Milgram's compliant subjects would have responded any differently to the question. Instead, as I suggested earlier, the experiment seemed to corrupt their judgment temporarily, disabling their capacity to apply their principles correctly to the situation they found themselves in.

Perhaps the most dramatic evidence that dissonance-induced belief change is an ephemeral thing comes from Zimbardo's personal recollections of the SPE. At one point, Prisoner #819 became ill and broke down emotionally. Zimbardo found him "sobbing uncontrollably while in the background his fellow prisoners were yelling that he was a bad prisoner" because the guards had ordered them to do so.⁷⁹ Remarkably, when Zimbardo tried to lead him away

78. Doris, *supra* note 63, at 143 (internal citation omitted).

79. Zimbardo, *supra* note 62, at <http://www.prisonexp.org/slide-31.htm>.

he refused, because he had to show his fellow-inmates that he was *not* a bad prisoner. Zimbardo said to him, “Listen, you are not #819. You are [his name], and my name is Dr. Zimbardo. I am a psychologist, not a prison superintendent, and this is not a real prison. This is just an experiment, and those are students, not prisoners, just like you. Let’s go.’ He stopped crying suddenly, looked up at me like a small child awakened from a nightmare, and replied, ‘Okay, let’s go.’”⁸⁰

The suggestion, then, is that dissonance-induced belief change does not resemble genuine integrity, even from within, because outside the experimental situation it fades and vanishes, unlike our genuine long-term moral and personal commitments.

Unfortunately, there is one crucial state of affairs in which this will not be true: the state of affairs in which the agent returns again and again to the situation that caused the belief-change. Recall the earlier argument that the pattern

action → rationalization → commitment → further action

has a recursive character. Put the same subject in the belief-altering situation day after day—or, better yet, have the subject put herself in the situation day after day—and it seems overwhelmingly likely that the transitory will become permanent.

For example, suppose that the belief-altering situation is *your job*, and that each day you voluntarily go to the office and put yourself back in the situation—for, let us say, 2,400 billable hours a year, year in and year out. And each night you take yourself out of the situation. By day, with a wig on your head and a band round your neck, you occasionally have to do things for a guinea that at night you would think it wicked and infamous to do for an empire. It seems very likely that before too long you will find yourself believing that a special professional morality, distinct from the morality of your extra-professional life, justifies what you do—and this belief will be no transitory thing, but rather a fixed part of your moral personality. Nor will this dualistic view of morality bother you. You will effortlessly negotiate the transition from one form of life to the other, with no sense of tension or contradiction.

Let me give two literary examples. One comes from *Lawyerland*, Lawrence Joseph’s fictionalized account of a dozen or so Southern Manhattan lawyers.⁸¹ In a chapter entitled “Something Split,” a

80. *Id.*

81. Lawrence Joseph, *Lawyerland* (1997). Joseph states that the book is “a work of nonfiction.” See *id.* at A Note to the Reader. It is unclear, however, what exactly that means, for Joseph also indicates that many details are changed, and the book is “truthful rather than factual, but solidly based on facts.” *Id.* On the fictional nature of Joseph’s portrayals, see David Luban, *The Art of Honesty*, 101 *Colum. L. Rev.* 1763, 1763-65 (2001).

corporate dealmaker named Wylie relates a story about his partner, Jack, who is in psychoanalysis.

So what does the mind doctor say? He tells Jack that, as a lawyer, he has to be capable of deep moral compromise Well, you can't argue with that. We all know there are times when you're working on some deal that, if you were to think it through, you'd realize that it was going to ruin the lives of thousands of people and their families. We all do it—in one size, shape, form, or other.⁸²

But Wylie and Jack think the psychoanalyst is a fool—and Wylie finds Jack's reply to the soul doctor hilarious: "Well, yes, doctor, that is what I do Yes, I am a lawyer. That is how I make my living, doctor. I make my living by committing acts of violence against myself and acts of violence against others."⁸³ Jack baits the psychoanalyst until he flees his own office—and Wylie and Jack laugh about it later.

My guess is that most lawyers would respond to the psychoanalyst the same way, and I must admit that it is hard not to sympathize with Jack. But Wylie agrees with the gravamen of the shrink's accusation ("We all do it," that is, "ruin the lives of thousands"), and the point of the story seems to be that lawyers like Wylie and Jack have no difficulty living with that diagnosis—that where one would expect to find "something split," eerily enough the protagonists experience no split at all.

The second example comes from a very different quarter, Montaigne's essay "Of Husbanding Your Will."⁸⁴ The essay praises those who remain aloof and emotionally detached from causes and enterprises, and in part it is a reflection on Montaigne's own tenure as mayor of Bordeaux. At one point he writes,

The mayor and Montaigne have always been two, with a very clear separation. For all of being a lawyer or a financier, we must not ignore the knavery there is in such callings. An honest man is not accountable for the vice or stupidity of his trade, and should not therefore refuse to practice it: it is the custom of his country, and there is profit in it.⁸⁵

A few paragraphs earlier, Montaigne wrote, "I have been able to take part in public office without departing one nail's breadth from myself, and to give myself to others without taking myself from myself."⁸⁶

These boasts bear the telltale signs of dissonance-reduction at work, especially striking in a self-observer and psychologist as acute as

82. Joseph, *supra* note 81, at 41.

83. *Id.* at 43.

84. Michel de Montaigne, *Of Husbanding Your Will*, in *The Complete Works of Montaigne* 766 (Donald M. Frame trans. 1943).

85. *Id.* at 774.

86. *Id.* at 770.

Montaigne. As Gerald Postema points out, Montaigne seemed to be making two different claims: first, that “the mayor and Montaigne have always been two people, clearly separated,” suggesting that the professional self and the personal self are two distinct selves; and second, perhaps inconsistently, that because Montaigne served as mayor “without moving the length of my nail from myself,” there is only one true self, and Montaigne (but not the mayor) is the true self.⁸⁷ In Postema’s terminology, the first claim is a “schizophrenic” strategy for proving that an honest man “is not accountable for the vices or stupidity of his calling”; the second is a “restricted identification” strategy.⁸⁸ Both strategies seem like self-deception, convenient ruses for denying that a person should be held responsible for the “knavery” of his calling. On the schizophrenic strategy, Montaigne is not responsible because the other fellow—“the mayor”—should bear the blame. On the restricted identification strategy, *no one* bears the blame, because Montaigne’s true self is not invested in the mayoralty. Apparently, the mayoralty itself bears the blame—a rhetorical strategy similar to lawyers’ frequent recourse to the excuse that “the adversary system did it.”⁸⁹ Both arguments neglect the fact that Montaigne *is* the mayor, and regardless of whether Montaigne is invested in the mayoralty, it is he who performs the mayor’s duties.

Strikingly, Montaigne resorts to these psychological fictions of schizophrenia and restricted identification in order to argue that one need not abandon professions that are customary and profitable, regardless of their knavery. If I am right that schizophrenia and restricted identification *are* fictions, and that Montaigne’s arguments for non-accountability therefore fail, we are left with the situation of someone whose practice of a customary, profitable profession drives him to stable, self-justifying belief changes whose only drawback is that they happen to be lies. Montaigne claims that he has kept his integrity—he has not departed one nail’s breadth from himself. But he has kept it, it appears, by fooling himself. This is a significant point for legal ethics, because one of the first modern articles on the subject quotes these sentences from Montaigne and takes them as a model of Stoic morality to justify the ethics of advocacy.⁹⁰ As Trollope, another

87. Postema, *supra* note 74, at 292 (quoting Charles P. Curtis, *The Ethics of Advocacy*, 4 Stan. L. Rev. 3, 20 (1951-52)).

88. *Id.*

89. To cite one example, Charles Fried distinguishes between a lawyer’s personal wrongs and institutional wrongs in which the lawyer is merely the occasion, the person who pulls the legal levers enabling the system to operate. Charles Fried, *The Lawyer as Friend: The Moral Foundations of the Lawyer-Client Relation*, 85 Yale L.J. 1060, 1084-85 (1976). For criticism of the idea that the adversary system provides robust moral excuses for otherwise-objectionable acts performed by lawyers and for criticism of Fried’s views, see David Luban, *The Adversary System Excuse*, in *The Good Lawyer: Lawyers’ Roles and Lawyers’ Ethics* 107-08 (David Luban ed., 1983).

90. See Charles P. Curtis, *The Ethics of Advocacy*, 4 Stan. L. Rev. 3, 20 (1951-52).

astute literary psychologist, observed, “Men will not be talked out of the convictions of their lives. No living orator would convince a grocer that coffee should be sold without chicory; and no amount of eloquence will make an English lawyer think that loyalty to truth should come before loyalty to his client.”⁹¹

The problem, in the end, comes to this: the ethical value of integrity is experienced from the inside as a kind of harmony or equilibrium between values and actions, whereby one does what one does without departing a nail’s breadth from oneself. But the experiments show that integrity has a kind of evil twin, induced by our need to see ourselves as ethically righteous people regardless of the knavery of our calling. From the inside, the quest for integrity and the process of rationalizing our actions prove nearly impossible to distinguish. We would like our moral compass to point true north, but our only instrument for detecting true north is our moral compass. And so, even though integrity and its evil twin may differ, the *quest* for integrity can drive us to the high road or the low road, without any landmarks to alert us about which path we have taken.

IV. IS THERE A CURE FOR INTEGRITY?

A. *The Truth Cure*

I fear that no cure for integrity exists. The problem is, quite simply, that the dissonance-based phenomena we have been examining, our urgent desire as intuitive lawyers to arrange our world so that we remain upstanding citizens in it regardless of what we do, all operate unconsciously.

One comforting idea is that the truth will set us free—or, more precisely, that understanding the dynamics of self-corruption (integrity’s sturdy twin) can help us fend it off. Robert Cialdini has written an admirable textbook on the power of social-psychological forces to influence us in directions we don’t want.⁹² At the end of each chapter, Cialdini offers a section entitled “Defense,” which distills from the experimental literature recommendations about how not to be taken in by the fundamental forces of reciprocation, commitment and consistency, social proof, liking, authority, and scarcity—the “weapons of influence.”⁹³ Cialdini’s basic defensive recommendation is enhanced awareness.⁹⁴ Yet he understands that

91. Anthony Trollope, *Orley Farm* 130 (1951).

92. Robert B. Cialdini, *Influence: Science and Practice* (4th ed. 2000).

93. *Id.*

94. For example: “The only effective defense I know against the weapons of influence embodied in the combined principles of commitment and consistency is an awareness that, although consistency is generally good, even vital, there is a foolish, rigid variety to be shunned.” *Id.* at 90. “If we can become sensitive to situations where the social proof automatic pilot is working with inaccurate information, we can

enhanced awareness of unconscious forces may be impossible precisely because the forces are unconscious.⁹⁵ His recommendations may well be the best we can do, but I have doubts that the best we can do will often be good enough.⁹⁶

A personal recollection: Two years ago, I was walking across a park in Dublin with my wife and in-laws. It was a nice summer day, and I noticed a man napping on the grass. As we drew nearer, I noticed that he was lying on his stomach, not his back. Then, as we walked by, I saw that his head was not turned to one side like someone asleep. His face was pressed directly into the ground. His limbs were splayed at awkward angles, and he was completely motionless. I had what I can only describe as a moment of listless recognition that he seemed to be dead—listless, because although I recall the thought that he was dead passing through my mind, I kept walking. The listlessness was not too surprising, as we had taken a red-eye from America and spent the whole day touring: all four of us had been awake more than twenty-four hours. It just seemed so natural to keep walking.

At that time, I had been a consumer of experimental social psychology for more than five years, and had discussed the Darley-Latané experiments on bystander passivity in my classes at least three times. It wasn't until we were past the motionless man that I suddenly recognized why none of us were doing anything. It had nothing to do with the red-eye. It had everything to do with diffusion of responsibility. I said, "That guy looks dead! We should do something." We all turned back to look at him, and saw a good Samaritan with a cell-phone standing next to the fallen man, excitedly phoning for help. Just as the experiments predicted, the Good Samaritan was all by himself. Score one for Darley and Latané.

Of course, one explanation for my passivity is that Luban is a weak vessel, who talks the talk of morality and compassion but won't walk the walk—or rather, who walks the walk right past collapsed strangers in a park. Perhaps that's it. But another example might persuade you that the case is not simply one of man's fecklessness and hypocrisy. The example comes from Philip Zimbardo's recollections of his

disengage the mechanism and grasp the controls when we need to." *Id.* at 134. "A better understanding of the workings of authority influence should help us resist it." *Id.* at 196.

95. "Here's our predicament, then: Knowing the causes and workings of scarcity pressures may not be sufficient to protect us from them because knowing is a cognitive act, and cognitive processes are suppressed by our emotional reaction" *Id.* at 228.

96. In saying this, I take issue not only with Cialdini, but also with Doris, *supra* note 63, at 153 (expressing cautious optimism that knowing situationist results can help us lead better lives) —and also with myself, for in Luban, *supra* note 30, at 116, I argued that "[p]erhaps the best protection is understanding the illusions themselves, their pervasiveness, the insidious way they work on us. Understanding these illusions warns us against them, and forewarned truly is forearmed, at least to some extent."

Stanford Prison Experiment. Zimbardo recalls a critical moment several days into the experiment:

One of the guards overheard the prisoners talking about an escape that would take place immediately after visiting hours. The rumor went as follows: Prisoner #8612, whom we had released the night before, was going to round up a bunch of his friends and break in to free the prisoners.

How do you think we reacted to this rumor? Do you think we recorded the pattern of rumor transmission and prepared to observe the impending escape? That was what we should have done, of course, if we were acting like experimental social psychologists. Instead, we reacted with concern over the security of our prison. What we did was to hold a strategy session with the Warden, the Superintendent, and one of the chief lieutenants, Craig Haney, to plan how to foil the escape.⁹⁷

Haney was one of the psychologists conducting the experiment. Zimbardo continues:

After our meeting, . . . I went back to the Palo Alto Police Department and asked the sergeant if we could have our prisoners transferred to their old jail. My request was turned down . . . I left angry and disgusted at this lack of cooperation between our correctional facilities (I was now totally into my role).⁹⁸

It only got worse:

I was sitting there all alone, waiting anxiously for the intruders to break in, when who should happen along but a colleague and former Yale graduate student roommate, Gordon Bower. Gordon had heard we were doing an experiment, and he came to see what was going on. I briefly described what we were up to, and Gordon asked me a very simple question: "Say, what's the independent variable in this study?"

To my surprise, I got really angry at him. Here I had a prison break on my hands. The security of my men and the stability of my prison was at stake, and now, I had to deal with this bleeding-heart, liberal, academic, effete dingdong who was concerned about the independent variable! It wasn't until much later that I realized how far into my prison role I was at that point—that I was thinking like a prison superintendent rather than a research psychologist.⁹⁹

When I first described the Stanford Prison Experiment, I noted that the "prisoners" and "guards" were not the only ones to become captives of their roles. Their parents, the priest who visited them, and even the lawyer who came in to consult with them about their "parole

97. Zimbardo, *supra* note 62, at www.prisonexp.org/slide-25.htm.

98. *Id.* at www.prisonexp.org/slide-26.htm.

99. *Id.* at www.prisonexp.org/slide-27.htm.

hearings” all did as well. Why not? None of these people were trained to recognize psychological mechanisms of influence at work. But Haney and Zimbardo were. If anyone should have recognized the “Pirandellian prison” of the mind, it is psychologists who devote their careers to mapping its gates and cell-blocks. Apparently, it isn’t so. Nor should that surprise us: If cognitive dissonance and social cognition truly represent universal psychological forces, it is a little much to expect that scientific expertise can free us of them. Understanding how Snell’s Law explains the bent-stick effect does not make the partially-submerged stick look any less bent to the physicist.

B. *The Canary in the Mineshaft*

I have said that there is no real cure for integrity—the low-road kind of integrity, that is, the unconscious gerrymandering of principles to rationalize commitments and actions that are too inconvenient to forego. I do have a few suggestions, however. If you really fear the gradual unconscious corruption that performing in role induces, you must decide in advance what line you won’t cross—and then, when you find yourself standing at that line, or, worse, when you find yourself having just crossed it, you will know that it’s time to quit. The inspiration for this suggestion comes from David Heilbroner, a former New York City prosecutor who wrote a fascinating memoir of his time in the D.A.’s office—a story of inexperience and naiveté gradually replaced by competence and cynicism. Heilbroner underwent a deep immersion in the seamy side of life where the good guys and the bad guys all lie sometimes, and where even doing good often leaves a bad taste. Heilbroner writes:

Before joining the DA’s office I had promised myself that above all, I would never take a case to trial if I had any doubt about the defendant’s guilt. At the time it seemed an easy enough standard to abide by. But during the past few weeks I realized that the Quintana case would probably force me to put my personal ethics to the test.¹⁰⁰

Heilbroner was prosecuting Quintana for theft, and had just learned that his star witness, the clean-cut, appealing, young victim was really a drug dealer, bail jumper, and liar. Heilbroner’s supervisor was unimpressed by the revelations about the witness, and insisted that Heilbroner take the case to trial. He did so, and Quintana was acquitted. Soon after, Heilbroner quit his job. “To stay on much longer meant maintaining a blindered belief in the rectitude of our work, wanting to punish defendants, believing that our policies were all to the good: becoming the very sort of prosecutor I had always

100. David Heilbroner, *Rough Justice: Days and Nights of a Young D.A.* 261-62 (1990).

disliked and distrusted. It was time to leave.”¹⁰¹ Heilbroner admits that he was temperamentally unsuited to the prosecutor’s job,¹⁰² and that some Assistant District Attorneys “loved prosecuting in an unquestioning way that I never could.”¹⁰³ Perhaps, then, Heilbroner’s resignation was inevitable and overdetermined. Nevertheless, I like the way he set himself a mental tripwire, or, switching metaphors, a single action that would serve as his canary in the mineshaft. The moment the canary died, he knew that it was time to evacuate. Heilbroner’s canary was taking a case to trial when he wasn’t convinced that the defendant was guilty. Other lawyers, in other practices, must choose their own canaries. The formula is simple: “Whatever else I do, and however else my views change, I will never, ever” You name it. Cover up someone else’s crime. Lie about money. Falsify a document. Let a colleague suffer the consequences for my own screw-up. Do something where I couldn’t look my father in the eye if I told him about it.

My advice is to choose your canary carefully, understanding that before you enter a role your ideas about what ethical demands it entails may well be naive. But, once you’ve selected the canary, never ignore it. If necessary, write down the “I will never, ever” formula. Put it in an envelope, keep it in a drawer, and pull it out sometimes to remind yourself what it says. And, the moment the canary dies, get out of the mineshaft.

C. *Noticing When You Are Deflecting Blame to Someone Else*

A second recommendation takes its inspiration from Milgram’s research. When he debriefed subjects after the electric-shock experiment, Milgram asked them to apportion responsibility for shocking the victim among the three protagonists—the subject himself, the “scientist” giving him orders to proceed, and the victim repeatedly earning electric shocks by giving wrong answers. As one might expect, compliant subjects seldom attributed the horrible outcome of the experiment to themselves. Characteristically, they blamed it on the scientist, and often on the victim. Taking a cue from this, my recommendation is the following: whenever you find yourself doing things but denying (to yourself or to others) that you are responsible for doing them, treat it as a sign that you have succumbed to the unconscious psychological drive toward intuitive lawyering.

This recommendation may sound peculiar, given the situationists’ warning that assigning responsibility for behavior to personality, not situational pressures, amounts to a “fundamental attribution error.” Am I now suggesting that you must *not* blame the situation for what

101. *Id.* at 279.

102. *Id.* at 283.

103. *Id.* at 284.

you have done, that you must take personal responsibility—in short, that you must commit the fundamental attribution error? Well, yes, in a way. Recall my earlier critique of situational determinism, where I argued that situations do not determine behavior, but merely alter the difficulty gradient, making it easier or harder to behave in certain ways. This, I suggested, is wholly compatible with a view that emphasizes the responsibility of agents in dealing with situations. To blame others—one's boss, one's co-workers, one's situation—amounts (to borrow Sartre's term) to a kind of bad faith.¹⁰⁴ Regardless of whether or not it is bad faith, however, my suggestion at the moment is simply that whenever you find yourself blaming others for your actions, treat that as an alarm-bell, signaling that you may well be in the grips of the psychological forces of rationalization.

D. Socratic Skepticism

My third and final suggestion is less specific, but perhaps more important. Throughout this discussion I have been emphasizing the dangers of our innate tendency to falsify facts and abandon principles in order to avoid the belief that we are doing wrong. Apparently, the need to believe in our own righteousness runs deep. One possible antidote to the drive toward self-righteousness is a stance toward the world that might be labeled "Socratic skepticism."

In Plato's *Apology*, Socrates tells the Athenian jury at his trial that throughout his life he has listened to an inner voice, a *daimon*. The voice tells him when he is in danger of doing wrong. It never speaks when he is doing right—only when he is doing wrong.¹⁰⁵ To give an example, Socrates mentions a period in which Athens was ruled by the Thirty Tyrants. These rulers wished to implicate as many Athenians as possible in their crimes. At one point, they called Socrates and some others in and ordered them to arrest Leon the Salaminian so that he might be executed. The others went off to fulfill the command, but (Socrates tells his hearers), his *daimon* spoke up, and, at risk of forfeiting his own life, he simply went home. Socrates adds that his own life was spared only because the Tyrants were overthrown very soon after.¹⁰⁶

Socrates goes on to explain that throughout his life he has made it his mission to seek out those who claim to know, and test them with probing questions, hoping (in vain, he informs us) to find someone whose high opinion of his own wisdom stood the test. He insists that

104. See Jean-Paul Sartre, *Being and Nothingness* 71-79 (Hazel E. Barnes trans. 1956). I would make an exception in genuine cases of duress—for example, the situation of soldiers who are told that unless they execute prisoners they will themselves be shot. In such cases, it does not seem like bad faith to blame their commander.

105. Plato, *Apology* *31d.

106. *Id.* at *32c-d.

he himself knows only that he does *not* know.¹⁰⁷ Although Socrates does not draw the connection between his *daimon* and his skeptical stance toward his own knowledge and that of others, it seems straightforward enough: the *daimon* tells him when his action has no justification, and his skepticism leads him to test every justification that he hears.

This stance toward the world—a stance of perpetual doubt toward one’s own pretensions as well as the pretensions of others—is what I am calling Socratic skepticism. It aims to combat our basic drive to believe in our own righteousness in the most straightforward way possible: by trying to make a habit of doubting one’s own righteousness, of questioning one’s own moral beliefs, of scrutinizing one’s own behavior—“know thyself!”—with a certain ruthless irony.

This advice will no doubt seem strange and disagreeable to many. Americans admire confident, can-do leaders who never second-guess their own decisions, and who avoid skepticism and self-doubt, the telltale signs of neurotics and losers. In the world of business and government, feelings of guilt or regret are career destroyers, best cabined to ceremonial occasions like Bible breakfasts and sentencing hearings. Nevertheless, I suggest chronic skepticism and discomfort with oneself as a possible antidote for integrity—if, that is, any antidote for integrity can be found.

107. *Id.* at *23a-b.