

# Chapter 11

## Behavioral Decision Theory and Business Ethics: Skewed Trade-offs Between Self and Other

George Loewenstein

Behavioral decision theory studies the trade-offs people make when they decide between options or courses of action. For example, someone choosing between jobs might trade off a high salary in one job against the high prestige of another job; for someone considering a risky prospect, such as starting a business, there would be a trade-off between certainty and expected value; and in intertemporal choices the typical choice is between earlier smaller rewards and later larger rewards. Behavioral decision theory also helps us to understand ethical decisions that involve trade-offs. My focus is on managerial decisions involving trade-offs between a manager's personal well-being and that of others.

Business decision making often calls for trade-offs between the well-being of the decision maker and that of others. The decision to blow the whistle on unsafe working conditions, for example, may result in the loss of one's job, but failure to do so may harm those exposed to the unsafe conditions. The desire to keep one's job is clearly consistent with self-interest; the desire to benefit another person is often termed *altruism*.<sup>1</sup> Altruism can be expressed formally by assuming that individuals' utilities depend not only on their own consumption ( $S$ ) but also on that of others ( $O$ ):  $U = U(S, O)$ . Altruism is said to exist when the partial derivative of  $U$  with respect to  $O$  is positive—that is, when utility increases with increasing payoffs to "other" or decreases with increasing costs borne by "other." Altruism is distinct from situations in which people care about others' outcomes only because they influence their own—that is,  $U[S(O)]$ , with  $S'(O) > 0$ .

A second, somewhat controversial, feature of behavioral decision theory is its attention to *errors* in judgment and choice. Unlike economics, which

---

The author wishes to thank Jon Baron, Eloise Coupey, and Karen Jenni for helpful comments.

assumes that people behave optimally, subject to constraints, behavioral decision theory describes human decision making with all its blemishes and warts. Proponents of the focus on error argue that errors are more interesting than "normal" behavior and that the best way to understand the norm is often to study deviations from it. This is especially true when it comes to studying ethical aspects of decision making. How many readers would make it to the second chapter of a book on ethics that trumpeted the ethical behavior of managers, and how much light could such a treatise shed on the determinants of unethical behavior? Ivan Boesky has more to tell us about business ethics than Ben and Jerry.

Fortunately, at least for ethics commentators, managerial behavior provides a virtually inexhaustible source of seemingly skewed trade-offs between personal well-being and that of others. Many examples are so notorious that they have become virtual code words for callousness: the Pinto, Johns Manville, DES, and the Dalkon Shield, to name a few. In each of these cases, a cohort of managers knowingly imposed egregious harm on an unknowing group of customers or workers for the purpose of making a profit. The question such cases raise is how such large groups of people could have deliberately traded relatively minor enhancements in their own material well-being for others' lives and/or livelihoods. I believe that behavioral decision theory, together with the allied disciplines from which it borrows, offers at least part of the answer to this question.

Behavioral decision research, I will argue, paints an extremely bleak picture of the possibilities for altruism in general, and managerial altruism in particular. The research I discuss in this chapter suggests several different reasons why managers' trade-offs between their own and others' well-being are likely to be skewed to the point where they put very little weight on the effect of their decisions on other parties, except insofar as those effects have repercussions for their own well-being.

First, considerable research points to the fact that altruism is generally a weak force in human behavior except, perhaps, in a relatively small subgroup of the population. Moreover, altruistic sentiments tend to be highly transient and ephemeral, at times surging to absurd proportions but at other times displaying an appalling feebleness. Second, it is especially easy to discount negative consequences to others when those who experience the consequences are statistics rather than known persons—a condition that is frequently satisfied in managerial decision making. Third, people are generally much more sensitive to incentives that are immediate than to those that are delayed. The fact that the impact of managerial decisions on managers themselves is generally much more immediate than their impact on others produces a further reduction in concern for others. Fourth, people typically evaluate trade-offs by comparing them with other trade-offs they have recently made or accepted. Thus, skewed trade-offs between self and other that might be unacceptable if introduced in their entirety, may well be accepted if the

decision is broken down into a series of small steps. This pattern is exacerbated when decision making takes place in insulated groups whose members use one another as points of reference. Fifth, people possess a remarkable ability to mislead themselves about the nature of the trade-offs they face—to rationalize that what benefits them also benefits (or does not hurt) others—and also to minimize their own responsibility for any adverse outcomes that occur. Finally, people underestimate—or are entirely oblivious to—the impact of many of these factors on their own decisions and, as a result, do not develop defenses against them. After reviewing research supporting each of these assertions, I discuss the possibilities for motivating ethical behavior and describe by way of contrast an example of how to ensure unethical behavior—the case of the “independent” auditor.

### THE WEAKNESS OF ALTRUISM

Recent empirical investigations of altruism have found it to be a surprisingly weak force in human decision making. There is no reason to believe that managers are exceptions to this rule; quite the opposite, as I will argue.

The weakness of altruism is evident in research on *social utility functions*, which specify an individual's well-being as a function of payoffs to that person and to other parties. In one study (Loewenstein, Thompson, and Bazerman 1989), we presented subjects with scenarios in which they and another party received money from the sale of jointly owned property or had to divide costs stemming from liability arising from such property. We found very little evidence of altruism; subjects' satisfaction with the outcome of the transaction rarely increased with the other party's payoff, and in many cases it declined substantially. This finding held true even when we described the relationship between the parties in very favorable terms. Rather than altruism, what we observed was a powerful loathing for coming out below the other side. This research suggests that altruism is a relatively minor force in human affairs when compared, for example, to the distaste for being “one-down.”

Similar conclusions have been reached in research on experimental games. For example, in the “ultimatum game” (Güth, Schmittberger, and Schwarze 1982), two subjects are paired off; one is assigned the role of “divider” and the other is assigned to be “chooser.” The divider is asked to split a fixed amount of money (for example, \$10) between the two players. The chooser then decides whether to accept or reject the proposed split. Neither person gets any money if the chooser rejects the offer. Although traditional game theory predicts that dividers should offer minimal amounts and choosers should accept them, dividers in fact typically offer the other subject more than a trivial amount, favoring most often an equal split, such as \$5/\$5.

Early accounts of the ultimatum game attributed the behavior of dividers to altruism; it postulated that dividers give up part of the “pot” because they

care about the well-being of choosers. Results from a closely related game, however, cast doubt on this explanation. The so-called “dictator” game (Hoffman and others, in press; Bolton, Katok, and Zwick, in press), is just like the ultimatum game except that the chooser must accept the split and therefore has no power to punish the divider. In the dictator game contributions are far lower than in the ultimatum game. Thus, consistent with the findings from research on social utility, it appears that dividers in the ultimatum game offer nontrivial amounts because they are worried that choosers will reject small amounts as unfair. This worry is well founded, since choosers often do reject inequitable offers, sacrificing personal gain to punish unfair dividers.

There are some exceptions to the general rule of weak altruism, not only across persons<sup>2</sup> but within them. Most people experience periodic surges of pity and concern for others, and at these times might even be willing to sacrifice on their behalf.<sup>3</sup> Such fluctuations, however, are probably relatively uncorrelated with the objective desirability of altruistic behavior. For example, a few weeks before writing this paper, when the massacre in Rwanda was taking place, I went to see a movie in which a boy who is somewhat older than my son is hit by a car and later dies. Later that night it struck me that my emotional response to the fictitious boy in the movie was stronger than my reaction to the slaughter of hundreds of thousands of Rwandans.

### STATISTICAL VICTIMS

In the research on the ultimatum game discussed earlier, subjects have a sense that they are playing the game with another specific individual, even though that person's identity is not revealed. In contrast, the people who will be adversely affected by many business decisions cannot specifically be identified at the time the decision is made; they are so-called statistical victims. Many commentators have lamented the public's tendency to show more concern for identifiable victims than for statistical victims.

Karen Jenni and I (Jenni and Loewenstein 1994) recently conducted research to test the validity of the *identifiable-victim effect* and to attempt to understand its underlying cause. Besides confirming that such an effect indeed exists, our research points to three major differences between statistical and identifiable victims that contribute to their differential treatment.

The first difference is the greater amount of information people have about identifiable victims. Given modern media coverage, when an identifiable person is at risk of death, a great deal is known about them almost immediately. For example, people see the school picture of a small girl who is trapped in a well, hear interviews with her tearful parents, and watch desperate attempts to rescue her. This information, necessarily unavailable for statistical victims, increases empathy for the identified victim.

Second, in situations with identifiable victims, all or almost all of the people at risk can usually be saved, whereas actions to save statistical lives generally save only a small portion of those who are at risk. People care much more if half of a group of twenty will die than if .01 percent of a group of one hundred thousand will die (Ritov and Baron 1990). Four (out of an identifiable group of four) whales trapped in the ice elicit international concern and costly rescue efforts, but hundreds of otherwise unidentified whales that will be caught in fishermen's nets are barely worthy of mention. Our empirical research suggests that this *reference group effect* is the single most important cause of the differential treatment of statistical and identifiable lives.

Third, identifiable victims are certain to be injured or to die if action is not taken, but statistical risks are probabilistic, so there is some possibility, however small, that no one will die. This factor seems to be a minor contributor to the discrepant treatment of statistical and identifiable victims.

Clearly, in most cases of business decision making, and especially those involving top management, the people who are adversely affected are not only anonymous to the decision makers (as in the ultimatum game) but are statistical in character. Smokers who eventually contracted lung cancer, workers and their families who died of asbestos poisoning, drivers and passengers who were incinerated in Pintos, and women who suffered high rates of infertility and miscarriage as a result of DES were all statistical victims from the perspective of the responsible managers at the time that the relevant decisions were being made. Specific victims couldn't be identified (and thus identified with) at the time the relevant decisions were made; it was thought that only a small fraction of groups exposed to the products would be affected; and there was some possibility, however small, that no one would be adversely affected.

### IMMEDIACY OF EFFECTS

In most business decisions, some of the consequences are felt immediately by the decision maker, whereas payoffs or consequences for others are typically delayed. Time delay is yet another factor that might cause decision makers to place undue weight on consequences to themselves relative to others. Considerable research has shown that people (and animals) place greater weight on outcomes that are immediate than on those that are delayed—a phenomenon commonly referred to as *time discounting*. People are disproportionately influenced by rewards and costs that are immediate or imminent. Because consequences to self are typically more immediate than those experienced by others, time discounting exacerbates the discrepancy in concern for consequences to self and others. For example, the whistle-blower is in danger of immediately losing his or her job, but the benefits resulting from blowing the whistle (or the costs of not doing so) are delayed.

The combination of delay and uncertainty seems to be especially pernicious, not only when it comes to managers imposing costs on others but also in people's behavior toward themselves in the future. In many of the standard examples of suboptimal individual decision making—for example, overeating, smoking, or failure to wear seat belts—the costs of changing one's behavior are immediate, but the benefits are both delayed and probabilistic. Given that people expose themselves to such risks, is it surprising that managers expose other people to similar types of risks? The Safeway supermarket chain may have adulterated its meat, mixing new meat with old and washing spoiled meat in order to sell it as new, but I ate a full rack of ribs for lunch the other day at the local house of grease, then barbecued a steak for dinner (both actions with delayed and uncertain consequences). The Ford Motor Company cynically sold Pintos with exploding gas tanks to increase profits, but for years I saved money by driving a rusty death trap (which I then sold to my neighbor's son). Executives of cigarette companies concealed evidence about the dangers of smoking but, while possessing the knowledge, they themselves smoked at a very high rate. The dictum to "do unto others as you would do unto yourself," even if followed religiously, might have little impact on the behavior of the executives of Safeway, Ford, and the cigarette companies.

Time delay and uncertainty are typically treated separately in the decision literature. Roger Brown (1986), however, argues that they can usefully be viewed as two dimensions of a more general construct that could be termed *immediacy*, which would also encompass attributes such as physical proximity and various forms of sensory contact. In general, people care more about immediate, certain, identified victims who can be seen, smelled, touched. Brown illustrated the effect of immediacy—broadly construed—by reanalyzing the results from the famous Milgram shock studies (Milgram 1974). The original Milgram shock studies demonstrated that a large fraction of relatively average people could be induced to administer what they believed were extremely painful and potentially hazardous shocks to another person (the victim) in response to verbal prodding by an experimenter.

Milgram ran twenty-one variations of his experiment, and Brown showed that the rank ordering of these conditions according to the percentage of subjects who administered the maximum level of shock could be predicted perfectly on the basis of the immediacy of the experimenter (who exhorted the subject to administer the shocks) and the victim. Shock rates were highest in the baseline condition in which subjects were in the physical presence of the experimenter but the victim was in another room. Rates declined when the victim was moved into the same room, declined further when the subject had to hold down the victim's arm to deliver the shock, and declined to virtually nil when the experimenter was absent.

Unfortunately, managers' trade-offs between their own and others' well-being typically occur in circumstances that approximate the baseline condi-

tion in the Milgram experiment (the experimenter is physically present and the victim is not), in which shock rates were at a maximum; the impact of managers' decisions on themselves is generally far more immediate in all senses of the term than the impact on other parties.

### ADAPTIVE TRADE-OFFS

Trade-offs, or the awareness of trade-offs, between self and other often change over time. For example, during the early stages of development of products such as cigarettes, asbestos, or DES, the executives responsible for producing and selling the products probably were not aware or only dimly aware of their hazardous qualities. Thus, they thought that the profit they were earning was at the expense of only minimal risks to their customers; awareness of risks emerged gradually.

Such a gradual unfolding of hazards to the public would not be problematic if managers evaluated risks and returns in terms of absolute levels. At a certain point, executives would simply decide that the product was too dangerous to sell to consumers, and they would take the product off the shelves. But human judgment is generally more sensitive to change than to absolute levels. A clever experiment conducted by Amos Tversky (1969) illustrates the hazards of making decisions on the basis of changes rather than levels.

Tversky presented subjects with a series of choices between simple gambles offering a chance of winning a small amount of money. For each choice, one gamble always paid \$25 more but had a 1/24 smaller chance of winning. For example, the first choice was between \$4.00 with probability 11/24 and \$4.25 with probability 10/24; the next choice in the series was between \$4.25 with probability 10/24 and \$4.50 with probability 9/24. The probabilities of winning those gambles were not presented numerically, but as spinners (like a wheel of fortune) with a certain fraction shaded black. If the spinner landed in the black area, the subject won the amount specified by the gamble. Probably because they could not visually detect the difference between the probabilities of adjacent gambles, most subjects opted for the higher payoff—smaller probability gamble from each of these choices. Thus, presented with a series of stepped choices, subjects aggregated responses pointing to a preference for the last gamble in the series (\$5.00, 7/24) over the first (\$4.00, 11/24). When Tversky presented subjects with a direct choice between these two gambles, however, a majority of subjects stated a preference for the \$4.00, 11/24 choice. This experiment shows that people can arrive at very different outcomes depending upon whether they are presented with a single large choice or a series of smaller incremental choices.

Many instances of unethical behavior seem interpretable in these terms. People often violate their own moral precepts in a series of small steps; they are "led down the garden path." Thus, in the famous Milgram experiments,

subjects were not asked at the outset to administer a potentially lethal shock but were given a series of requests to increase the voltage marginally. Having given someone a 100-volt shock, one finds it difficult to justify stopping at precisely that point rather than acquiescing to the experimenter's request to increase the voltage again by a small amount. Similarly, R. J. Lifton (1990) argued that it was the incremental character of ethical decay that made it possible for German doctors to become active killers, even though they had taken the Hippocratic oath to do no harm. He described a process whereby doctors were first present when euthanasia took place, were later asked to add their signature to a document, still later were asked to supervise a mercy killing, and so on, to the point where many actually administered lethal injections to eugenically "undesirable" persons.

This failure to notice step-by-step degradation of ethical standards is likely to be exacerbated when managers are imbedded in insulated groups, as is common. It is well established that people in groups tend to compare attitudes and opinions among themselves and that such social comparisons can have an important formative impact (see, for example, Asch 1951; Myers and Lamm 1976; Janis 1972). Just as a rude comment from a shopkeeper would be more noteworthy in Atlanta than in Manhattan, behavior that is extreme with respect to the standards of the general population will appear much less so in a group context if all members of the reference group change together. Choices that appear unethical to outsiders may seem perfectly justified to a group of managers.

### THE HUMAN CAPACITY FOR RATIONALIZATION

People are not objective information processors. One of the most important nonobjective influences on information processing is self-interest, which poses yet another impediment to incorporating concerns for or about others in business decisions. It is by now well established that people tend to conflate what is personally beneficial with what is fair or moral. For example, David Messick and Keith Sentis (1979) asked subjects to specify the "fair" rate of pay for two people (self and other) who had worked at the same task, one for ten hours and one for seven. The person who worked seven hours was always paid \$25; subjects were asked how much the person who worked ten hours should be paid. When told that it was the subject who had worked seven hours and the other person had worked ten, a large number of subjects advocated strict equity or piecework wages—\$25 for each worker—as fair. When told that the other person had worked seven hours and the subject had worked ten, however, subjects tended to advocate a fixed hourly wage, resulting in a higher payoff for themselves, as fair.

In studies by Roth and Murnighan (1982), pairs of players bargained over one hundred chips that determined their chances of winning a monetary

prize (for example, thirty-seven chips gave a 37 percent chance of winning). One player's prize would be \$20 and the other's prize would be \$5. Notice that there are two ways to split the chips "equally"; one can give fifty chips to each, giving them equal chances of winning, or one can give twenty chips to the \$20-prize player and eighty chips to the \$5-prize player, equalizing the expected dollar winnings). When neither player knew the prize amounts, they agreed to divide the chips about equally. When the players knew the prize amounts, however, they tended to hold out for the distribution of chips that favored themselves, producing a higher rate of disagreement.

In my own work with Linda Babcock, Sam Issacharoff, and Colin Camerer (Babcock and others 1995; Loewenstein and others 1993), we have presented subjects with diverse materials from a lawsuit resulting from a collision between an automobile and motorcycle. Subjects are assigned the roles of plaintiff and defendant and try to negotiate a settlement. If they are unable to settle, the amount paid by the plaintiff to the defendant is determined by an impartial judge who has read exactly the same case materials. Before they negotiate, we ask subjects to predict the judge's ruling, and we offer a monetary reward for accuracy. Nevertheless, plaintiffs' predictions of the judge's award amount are typically substantially higher than defendants'.

Recent research by Chris Hsee (1994a) points even more directly to the prevalence of rationalization. Hsee exposed subjects to decisions involving a trade-off between instrumental attributes (those that are easy to justify) and affective attributes—for example, the choice between two job candidates, one of whom is more competent, the other more physically attractive. Hsee found that when both attributes were defined precisely, people generally opted for the candidate stronger on the instrumental attribute—in this case, the more competent candidate. If provided with an excuse for hiring the more attractive candidate, however, such as a remote possibility that the more attractive candidate is in fact more competent, people used the excuse to justify hiring the more attractive candidate.

All this research shows that when there are competing norms of fairness, subjects will tend to select as relevant those that materially favor themselves. In addition, when there is ambiguity about the consequences of alternatives, people will be able to rationalize taking the option they personally prefer as opposed to the one that is normatively or ethically more justifiable. Since the consequences of business decisions are typically highly ambiguous, these types of rationalizations are likely to be very common in business decision making.

A second form of ambiguity that is pervasive in business settings probably contributes to self-serving judgments and to the undermining of ethical behavior by managers. This is the ambiguity of who is *responsible* for a particular outcome. People seem to be quite adept at relieving themselves of responsibility for harming another person, as illustrated by the very high rates of maximum shock delivery (90 percent) in a variant of the Milgram

experiment in which the subject did not actually pull the switch that delivered the shock but carried out an operation that was ostensibly necessary for the shock to be delivered (see Sabini and Silver 1982). As in this experiment, many people in businesses perform functions that are necessary for a particular project to be completed, but few perform functions that are sufficient for its completion, so that any consequences can be traced purely to their own actions. In business and other institutional settings, as Sabini and Silver (1982) note, "The relation between an individual's action and the rules and commands of an organization obscures personal responsibility" (p. 65).

## FAILURES OF SELF-PREDICTION

The final nail in the coffin of managerial altruism is the tendency to underestimate the influence of the factors just discussed on one's own behavior. In principle, people may want to behave in a fashion that reflects a substantial weighting of others' welfare, but in practice their failure to recognize the force of these factors may undermine their ability to do so.

For example, people underestimate the influence of immediacy on their decisions. The failure to predict the impact of immediacy is well illustrated by an oft-cited study by Christensen Szalanski (1984). He interviewed expectant mothers about their desire for anesthesia during childbirth. A majority of women expressed a preference for natural childbirth until after labor began, at which point a majority shifted in favor of anesthesia. Apparently, the women were unable to anticipate the severity of or the motivating quality of the pain they would experience, as if they did not somehow fully empathize with their future (in-pain) self. Interestingly, this was also true of women who had given birth previously, suggesting that the lack of empathy also extended to their past (in-pain) self. The tendency to underestimate the effect of immediacy can also explain why the results of the Milgram shock experiment are so surprising to people. When subjects were presented with a description of the Milgram experiment but not told about the final result, most predicted that only a very small fraction of subjects would administer high levels of shock and that they themselves would not (Milgram 1974). People seem to underestimate the impact of the experimenter's immediate presence.

People also underestimate the effect of adaptation on their own behavior. For example, a robust finding in behavioral decision research is that when people are endowed with an object, they are typically very resistant to giving it up—far more so than one would expect on the basis of their desire for the object in the first place. This *endowment effect* is typically explained on the basis of adaptation (to possession) and *loss aversion* (the aversion to losing what one has). My own research on the endowment effect (Loewenstein and Adler 1995) shows that people are unable to predict that they will become

attached to objects once they possess them. Thus, they seem to underestimate the effect of adaptation on their own behavior. In the context of business decisions, this research suggests not only that managers will adapt to the status quo, perhaps influenced by their peer group and by a series of small changes, but also that they will be unaware of the adaptation they have undergone.

Finally, people seem to underestimate their own powers of rationalization. In the studies of the legal dispute discussed earlier (Loewenstein and others 1993; Babcock and others 1995), we paid plaintiffs and defendants for the accuracy of their prediction of the judge's ruling; the fact that this motivation had no effect on the predictions is consistent with the conclusion that the self-serving bias is unconscious and nondeliberate. Moreover, in one set of studies we informed subjects about the bias before they gave their predictions of the judge's award; we also asked them to predict their opponent's prediction of the judge's award. Telling them about the bias had no impact on their own predictions of what the judge would do, but it did change their prediction of their opponent's prediction; subjects informed of the self-serving bias believed the result and thought that their opponents would exhibit the bias but that they themselves were somehow immune to it.

In each of these cases, failure to appreciate the effect of a particular factor is likely to leave managers vulnerable to it. Unaware of or doubting the effect of immediacy, adaptation, and rationalization, managers exposed to situations in which these factors are operative, although confident in the belief that they will not be affected, are likely to succumb to these forces.

## POSSIBLE SOLUTIONS

Early social thinkers such as Adam Smith and Thomas Hobbes were impressed by the power of human empathy and altruism but nevertheless concluded that these tendencies were not sufficiently powerful to ensure socially constructive behavior. Smith argued that the desire for personal gain was a much more reliable force with which to motivate socially beneficial behavior, while Hobbes believed that the coercive force of strong government was needed to rein in human behavior. The research I have reviewed supports their skepticism concerning the possible role of altruism in social life.

Probably the best way to encourage behavior that benefits other people is to provide a personal incentive for doing so. To a great extent such incentives are built into daily life. For example, there is a strong *reciprocity norm*; acts that benefit others are often paid back in some form, and moreover, as the ultimatum game demonstrates, people are often willing to sacrifice their own well-being to punish people they feel have harmed them or treated them unfairly. Thus, people have built-in behavioral responses that motivate others to treat them fairly (see Frank 1988).

The legal system is also designed, if not to encourage generosity toward others, at least to discourage the imposition of egregious harm on others. To the extent that legal behavior is also ethical behavior, fear of sanctions is undoubtedly a major contributor to the latter. Finally, social norms and the sanctions that result from violating them are probably an important source of behavior that takes others into consideration. For many people, managers included, loss of reputation or social standing is as much to be feared as any material penalty. Thus, people's personal constitutions and the institutions and norms that prevail in society are all arranged in ways that reduce the necessity for personal altruism.

Whether or not it is necessary or even socially desirable to promote individual altruism is an unresolved (and probably unresolvable) question. To the extent that one wants to do so, however, the foregoing discussion suggests several ways to encourage business managers to consider the concerns of others and the impacts that their decisions may have. First, managers should be in close contact with the people who are affected by their decisions, making these people "identified" and "immediate." Second, measures should be taken to prevent managers from forming an insulated subgroup, in order to minimize the chance that group norms will deviate substantially from societal norms. Ideally there should be a periodic introduction of new personnel into the ranks of any group of managers. New people will not only see any incremental changes that have occurred in the aggregate but will take time to adapt to the evolved norms of the management subgroup in which they are placed. Third, steps should be taken to keep people out of situations in which they are faced with temptation and in which the various factors that undermine altruism are operative. With respect to this third point, it may be informative to examine a case in which precisely the opposite occurs—the case of so-called independent auditors.

## HOW TO GUARANTEE UNETHICAL BEHAVIOR: THE CASE OF AUDITING<sup>4</sup>

In theory, auditors are supposed to represent the interests of external users of financial statements, such as stockholders, potential stockholders, financial advisers, underwriters, and potential creditors. Auditors are paid by the company they audit, however, who can hire or fire them at will. Moreover, auditors often socialize with the management of the company they audit. The American Institute of Certified Public Accountants (AICPA) acknowledges the pressures on auditors but argues that personal integrity is sufficient for objectivity. Rule 102 of the AICPA (1988) code of professional ethics states,

In the performance of any professional service, a member shall maintain objectivity and integrity, shall be free of conflicts of interest, and shall not knowingly misrepresent facts or subordinate his or her judgment to others.



The AICPA (1988) seems to feel that such integrity can be maintained by exhortation alone:

Members should accept the obligation to act in a way that will serve the public interest, honor the public trust, and demonstrate commitment to professionalism

In light of the points I have raised, this standard can be seen as entirely unattainable for most people. First, the people who will be hurt by any misrepresentation of information are statistical. Many of them might lose a small amount of money; it isn't clear who will do so; and there is some chance that no one will be adversely affected by a minor misrepresentation. In contrast, the auditor is likely to be intimately acquainted with those who would be hurt by a negative ("qualified") opinion on an audit. Second, the negative consequences of a qualified opinion are likely to be immediate—loss of the client's friendship, likely loss of the contract, and possible unemployment—whereas the effects of a false negative (an unqualified report where qualification is merited) are likely to be delayed in time. Third, auditors form an ongoing working relationship with the organizations they audit, and any deterioration in the audited company is likely to unfold gradually. Auditors may unknowingly adapt to small changes year after year in the company's financial practices. Fourth, financial records are inherently ambiguous, so it is very easy for an auditor to rationalize arriving at a judgment that is consistent with self-interest rather than with the actual financial figures. In sum, if one wanted to create a business setting that would virtually *guarantee* unethical behavior, it would be difficult to improve on the existing case of independent auditing.

## CONCLUSIONS

Every generation would like to believe it is more civilized than the last, but inevitably there comes a moment of truth. As recent events in Bosnia and Rwanda show, our generation enjoys no exemption. In the aftermath of each period of mass cruelty and murder, a soul-searching follows in which people ask themselves how human beings could have done these things to one another. Their answers are as diverse as intellectual thought itself: mob psychology, toilet training, authoritarian family structure, a "death instinct," capitalist greed, obedience. The purpose of this chapter has been to consider what insights behavioral decision theory and allied subdisciplines have to offer on the age-old question of "man's inhumanity to man" and, more specifically, on the behavior of managers.

In light of the behavioral decision theory research, I believe, the prevalence of unethical business behavior poses no puzzle whatsoever. The para-

dox in search of explanation, if any, is the relative infrequency of such behavior. The key to business ethics does not lie in altruism.

## ENDNOTES

1. The exact definition of altruism has been debated endlessly. For example, some writers would exclude from altruism situations in which people help others because they feel good as a result of doing so. Distinctions are also commonly made between "genetic" altruism and "psychological" altruism. I define altruism simply as the weight placed on others' costs and benefits relative to one's own in decision making. This definition does not count as altruism situations in which people help others because they expect to be compensated in some material fashion.
2. There probably are some "true" altruists in the population. Indeed the research on social utility functions and dictator games, as well as studies of behavior in the prisoner's dilemma and in social dilemmas, suggests that a subgroup of the population—perhaps 20 percent—may be noncontingent altruists. Although such people deserve commendation, their aggregate influence is probably relatively small.
3. Much of the social psychology literature on altruism and helping behavior has focused on the issue of how to elicit or inhibit such pity and concern and on the exact mechanism by which concern leads to actual helping behavior (see, for example, Batson and others 1991; Cialdini and others 1987).
4. This section borrows from an unpublished term paper by Kimberly Morgan, "Auditors' Perceptual Biases: A Threat to Independence" (University of Pittsburgh, Katz Graduate School of Business).

CODES OF CONDUCT  
Behavioral Research into  
Business Ethics

DAVID M. MESSICK AND ANN E. TENBRUNSEL  
EDITORS

Russell Sage Foundation / New York

(1996)