

Tradeoffs among reasons for action*

Jonathan Baron[†]

Abstract

Lexical rules are rules for decision making that give one type of consideration absolute priority over another, e.g., life over money. At the normative level — the level of reasoning designed for people with perfect self-control and unlimited time for reflection — there is no good justification for lexical rules. Whenever we are tempted to say that (for example) life is always more important than money, we can imagine a situation in which a tiny probability of loss of life is pitted against a great gain in money for a great many people. Arguments for lexical rules at the prescriptive level — the level appropriate for real people to teach to their children — may be made, but these arguments are often weaker than they first appear to be. Hence, ‘simple moral systems,’ moral codes based primarily on lexical prohibitions, are fallacious. These arguments may expose other fallacies in everyday moral codes as well, such as reliance on distinctions between omission and commission. A prescriptive theory is proposed that is closer to the normative one.

A funny thing happened to me at a workshop given by Professor Lawrence Kohlberg about moral thinking. Kohlberg presented his classic dilemma (e.g., Kohlberg, 1970) about whether a husband should steal a drug in order to save his wife’s life. When Kohlberg asked how many thought that the husband should steal the drug, almost all hands went up, including mine. The next question was, roughly, ‘How many think that a person would be wrong to think that the husband should not steal the drug?’ To my surprise, my hand was one of the few not to go up. I wondered why not. Was I really a closet relativist (or subjectivist)? Was *I* the one who was wrong? This paper is the result of that wondering.

My main claim concerns what I have called the normative and prescriptive theory of action (Baron, 1985a,b). Briefly, I take prescriptive theory to set out the rules of how to make decisions about action. Normative theory represents the goal that prescriptive theory tries to achieve. These kinds of theory are of

*Pre-publicaiton version of Baron, J. (1986). Tradeoffs among reasons for action. *Journal for the Theory of Social Behavior*, 16,173–195.

[†]I thank Jane Beattie, Alan Fiske, John C. Hershey, Mark Monheit, John Sabini, Barry Schwartz, Stan Schwartz, and several anonymous students and reviewers for helpful discussions and comments on earlier drafts.

particular importance for choices that involve tradeoffs of different considerations, one weighing on the side of one action, another on the side of another. These considerations might be moral, prudential, or (most typically) both. We are sometimes inclined to follow *lexical rules*, rules in which one consideration is always primary to another, regardless of the strength of the second or the improbability of the first being relevant. For example, we might justify stealing the drug by saying that life is more important than property. My argument is that such lexical rules cannot be decisively justified from a *normative* point of view. Therefore, if a person considers both life and property, and decides that the damage to property outweighs the damage to life in a given case, he is not necessarily wrong from a normative point of view. Although lexical rules may be justifiable as *prescriptive* rules, such justifications are contingent on claims about human psychology - claims that may be correct for some people only (e.g., people of the present time).¹

Many of the arguments against lexical rules apply at the prescriptive level as well as the normative level. An implication of this is that the moral systems of many people are inappropriate in a way that is analogous to their decision-making systems (Kahneman & Tversky, 1984), which seem to deviate from normative models in ways that are likewise difficult to justify prescriptively. I thus advocate a move toward a prescriptive theory more like the normative theory. I suggest that such a prescriptive theory is practical if people try to weigh competing considerations honestly and consistently. We must also realize that we are bound to weigh our self-interest more than the interests of others, and we should try to do this consistently across situations. Although this view shares some features with moral relativism, it is saved from being fully relativistic by an emphasis on the method by which the conclusion is reached. Thus, what is not relativistic is an insistence on the use of certain methods of thinking in the drawing of conclusions, specifically, methods that abjure self-deception and treat alternative possibilities fairly. (These methods do not guarantee agreement, however.)

A legitimate task of moral psychology is to try to understand why moral thinking is not always what it ought to be. In seeking this kind of understanding, we need a clear view of what moral thinking ought to be. Such a view might also suggest where it goes wrong.

1 Some preliminaries

1.1 Normative and prescriptive theories

There is good reason to have separate normative and prescriptive theories of decision making (Baron, 1985a). A prescriptive theory is a set of rules designed to specify what people should do (or how they should decide what to do) in

¹My claim concerning tradeoffs is fully consistent with (but need not assume) utilitarianism. This fact, and the distinction between normative and prescriptive levels, makes my own views very close to those of Hare (1981).

certain situations. Most moral codes, and many prudential rules (e.g., ‘Honesty is the best policy’), are intended to serve this purpose. Prescriptive rules may be contrasted with *descriptive* statements (or theories) of how people actually behave. When people behave at variance with prescriptive rules that we have accepted, we can criticize their behavior (constructively) by pointing to the discrepancy.

A normative theory is what we would like, on reflection, to have our behavior conform to if we could. For example, the normative theory of decisions might specify that we want to maximize some sort of overall measure of preference satisfaction (utility) within the individual decision maker, and the normative theory of morality might specify that we want to maximize some sort of total good (possibly utility, possibly something else) across individuals. A normative theory is a set of standards, not ordinarily a guide to action.

One role of normative theories is to justify prescriptive ones. For example, to argue that ‘honesty is the best policy,’ we might try to show that our following this rule will maximize utility (in some class of cases). A useful way to think about normative models is that they are actually prescriptive models that would be accepted on reflection by a creature who had perfect self-control and infinite time to gather information and reflect both about the rules to be used and the application of those rules.² The normative model then becomes the best justification we have for a given prescriptive rule. In some cases, when time for decision making and calculation is not at issue given the importance of the decision, the normative model may become the prescriptive model, as when utility theory is used in decision analyses (Keeney & Raiffa, 1976). By this account, one task of normative theorists is to reflect about the rules themselves on behalf of everyone.

The distinction between normative and prescriptive theories is a way of dividing up the task of reflecting about the kind of theory we want. We first reflect on what we would want if we weren’t constrained by certain limitations, and then we use the theory developed in this way, plus the facts about actual constraints, to design the final, prescriptive, theory. The prescriptive theory is our basis for advice and criticism. The constraints themselves may excuse failure to conform to the normative theory.

This distinction is exactly analogous to Hare’s (1981) distinction between the Archangel and the Prole in moral deliberation. In Hare’s utilitarian theory, the Prole (a hypothetical creature whose methods we all may use) makes moral decisions by following simple rules, such as, ‘Life is more important than property,’ or, ‘Thou shalt not murder.’ The Archangel (another hypothetical creature whose methods we all may use) makes moral decisions by putting itself simultaneously in the position of all affected parties and balancing the relevant preferences of one against those of another — exactly as one would balance conflicting preferences against each other within oneself. These decisions are ordinarily about specific actions, which are described as specifically as possible.

²Different, but related, normative models might result from different specifications of exactly what these conditions mean.

In such cases, the normative model and the prescriptive model are the same. In other cases, the Archangel might reflect not about specific actions but about general rules of thumb, or heuristics, to be used by the Prole. Some of these rules would tell us when Archangel reasoning itself was appropriate (e.g., not when time is short or when self-interest is likely to lead to strong biases). This use of Archangel thinking is analogous to (or an instance of) the use of normative models to justify prescriptive ones.

Hare uses the distinction between the Archangel and the Prole to defend utilitarianism against a number of objections of a sort that may be made to any theory involving tradeoffs. One objection is that there are cases in which utilitarianism seems to justify some act that conflicts with very basic rules of moral behavior, such as prohibitions against murder. In many such cases, Hare argues, a person would do best (even from the Archangel's point of view) to follow the rules of the Prole and not to attempt Archangel reasoning. For example, conclusions of political terrorists that their acts are justified by the greater good do not, in general, seem to have panned out, and therefore potential terrorists would do well to think twice, or thrice, before acting. In other cases the Archangel might be right, despite our intuitions to the contrary. Our intuitions may result from having been taught prescriptive rules, which might not always apply.

Another objection is that rule-utilitarianism and act-utilitarianism often prescribe different actions (as may any theories concerned with individual acts versus classes of acts). By Hare's account, the former is for the Prole, the latter, for the Archangel. As noted, other rules must be invoked to determine when the categorical (Prole) rules may be overridden.

1.2 Decision theory, moral theory, and moral-decision theory

A second relevant distinction, cutting across the normative-prescriptive distinction, is that between moral theory and decision theory. Decision theory concerns the achievement of rationally chosen goals, including *both* moral and prudential goals.

An example of a normative decision theory is Multi-Attribute Utility Theory (or MAUT; see Baron, 1985a,b; Keeney & Raiffa, 1976). According to this theory, the decision maker assesses the utility of each possible outcome on each of several dimensions (e.g., profit, energy production, number of diseases, number of deaths). By multiplying the utility of an outcome by its probability given a particular choice, we may compute an expected utility for that choice on that dimension. By determining a weight for tradeoffs between one dimension and another, we can combine the expected utility on the different dimensions into an overall expected utility form each choice.³

³In the versions of MAUT most often discussed, several assumptions are made, which provide an axiomatic foundation for the system, a set of criteria for our judging its appropriateness in a given case, and, by inference, a method for applying the theory (Krantz, Luce, Suppes, & Tversky, 1971, chapter 6; Keeney and Raiffa, 1976). There are various sets of

Decision theory is silent about the proper relative weight of moral considerations versus self-interest. It accepts whatever weights are provided by the decision maker. Part of forming a life plan (in the sense of Baron, 1985, and Rawls, 1971), even for an ideal creature, is to decide how much weight to give to moral considerations. The decision to adopt morality as part of a life plan need not be (and usually isn't) all or none.

Moral theory concerns moral judgments. Here, morality is primary to all other considerations. Self-interest comes in only as the interest of one person among many. It is the perspective of moral theory that we use when we judge that someone acted immorally, even when, from the perspective of decision theory, putting ourselves in the position of the person we condemn, we can sympathize with the decision made.

An example of a moral theory that is closely analogous to MAUT is act utilitarianism, we weigh preferences of different people against each other. The only difference is that preferences are measured and traded off across individuals. As Hare (1981) points out, thinking about such tradeoffs need not be any more difficult than thinking about tradeoffs across very different dimensions in utility theory (e.g., love versus success). Like MAUT, utilitarian theory may take probabilities into account (although this fact is rarely discussed in the literature).

The kind of theory we seek here is neither a decision theory nor a moral theory. We do not want a decision theory because we want to allow stronger statements about morality than decision theory makes. We might, for example, want a theory that disallows total inattention to morality. We do not want a moral theory because we might (as I shall argue later) want to allow people to weigh self interest more than a moral theory would allow. I shall call what we seek a moral-decision theory, a theory (both normative and prescriptive) that concerns decisions, not judgments, and that does take morality into account. Although we might still sympathize with the violator of such a standard, we would have to judge him as morally wrong in a stronger way than if he had violated the purely moral theory.

2 The argument against lexical rules

I shall now argue that lexical rules are never appropriate at the normative level for any decision or moral theory. In other words, there is no consideration that ought always have infinite weight relative to any other consideration. By 'consideration,' I mean any property of a decision that weighs on one side or the other. I assume that an attribute or consideration can be imagined, in a normative, mathematical sense, as an ordered continuum. Each point on the continuum is (abstractly stated) an equivalence class of outcomes, and we may suppose that it is possible to assign each outcome (probabilistically, at least)

assumptions that will serve these purposes, but they usually include weak ordering and transitivity of preferences, independence of dimensions, and some assumptions designed to insure that scales can be constructed.

to its class in each attribute, that is, to assign it its position on the scale. I also assume that it is possible to imagine outcomes at any point along each continuum. If two outcomes differ only in a single consideration, we can imagine an infinite series of possible outcomes between these that differ along the same dimension. There are no gaps. Some of these outcomes might be gambles involving other outcomes, for example, a .1 probability of the first outcome .9 probability of the second. This gamble would be closer to the first than to the second.⁴

As an example for my argument, take the claim that the preservation of human life is always more important than money. Many people do feel that no expense should be spared to preserve a particular human life, so long as the person in question wishes to live.⁵ (I put aside those cases in which the person's desire to live is itself ambiguous.) If a close relative needs a life-saving operation, one ought to use all available funds to pay for it. If the funds are simply not available, this is a moral tragedy. People who feel this way often feel that life is always more important than money. Of course, 'money' here is a stand in for 'the *other* things that money can buy.' Thus, we do not want to spend so much money that other lives are lost.

As a practical (prescriptive) matter, adoption of such a principle takes away from physicians, families, and insurance companies the burden of deciding how much a particular life is worth. Insurance companies in particular have self-serving reasons to err on the side of saving money, and it may be just as well to try to prevent them from doing this. Moreover, any infringement on the idea that life is more important than money may lead us down a slippery slope toward a society in which life is valued very little, and we should not take this chance.

But we are speaking normatively now, and at that level we may assume that dispassionate judgments of tradeoffs (between life and money) are possible. Nor is the slippery-slope argument relevant at the normative level. The ideal creature for whom the normative theory is intended must be conceived as able to overcome self-interest (or, more precisely, to weigh it appropriately) and to hold a position on even a slippery slope. Self-serving bias and susceptibility to slipperiness are actually contingent facts (which might sometimes be false) about human nature, relevant at the prescriptive level, but irrelevant at the normative level, where they may be postulated out of existence by idealization.

⁴In particular, I assume weak ordering, transitivity, and restricted solvability. Weak ordering insures that of any two outcomes A and B, either $A \succ B$, $B \succ A$, or $A \sim B$, where \succ indicates preference and \sim indicates indifference. Transitivity says that if $A \succ B$ and $B \succ C$, then $A \succ C$. Restricted solvability deals with the situation in which $A1 \succ B \succ A2$, where A1 and A2 differ only in a single dimension. In this case, there must be an outcome A3 such that $A3 \sim B$. In addition, I assume that there is a gamble in which the outcome is A1 in state a1 and A2 in state a2, such that the decision maker is indifferent between this gamble and B.

⁵There are many other examples that would do. We could take Rawls's (1971) argument that primary goods should be distributed so that the welfare of the least advantaged class is prior to other considerations, or Nozick's (1974) argument that the right to give one's money to someone else is prior to considerations of redistribution of income. As we shall see, any concept of inviolable rights usually leads to some sort of lexical ordering.

The Archangel need not be a cynic.

Lexical rules are often supported by direct moral intuition. Let us put aside whatever doubts we have about the legitimacy of justifying general rules on the basis of moral intuitions (Hare, 1981) and put ourselves in the position of someone whose intuition would indeed balk at even a small curtailment. Suppose that an insurance company has a choice of saving the life of Mr. Smith with a heroic operation costing millions of dollars or else curing a million people's arthritis. (I pick arthritis because it is painful but not usually life threatening.) Intuition may still balk. (Mine does.) But now let us suppose that the success of the operation is not a certainty, but rather a probability P . Suppose that P were .001. Would our intuition change? Suppose it were .000001. Would it change then? It seems to me that there has to be a value of P sufficiently small so that our intuition would switch. If not, I would suggest that we are actually molding our intuition to conform to the principle (which would then need some other justification) rather than justifying the principle in terms of our intuition. If such a value of P exists, then we are, in fact, willing to trade off the two attributes at issue, life and money. That is, we are willing to say that some finite interval on the economic scale is equivalent to a nonzero interval on the life scale. The nonzero interval on the life attribute is just the interval between no change at all and the gamble of the smallest change at probability P . This interval is equivalent (or less than) the interval between the status quo and the arthritis cure, on the economic scale.

Consider a more difficult case. No matter how much money I am paid to kill you, it would not be morally right, it seems. But now let us ask whether it would be morally right for me to expose you to an additional .0000000001 probability of immediate death in order for me to gain a million dollars. Perhaps our intuition still balks at this one. However, I would suggest that the reason does not concern the tradeoff itself, but rather a very strong prescriptive rule, specifically, the rule that it is morally risky to decide to sacrifice someone else's interest for one's own. Once we allow this at all, we may again be dragged down a slippery slope by self-serving bias, as in the insurance-company example described earlier. If we imagine an ideal person not subject to these failings, our intuition becomes less clearly opposed to taking the chance in question.

To remove the possibility of self-serving bias (which is normatively irrelevant) let us ask whether it is morally right for me to expose person A to this probability of death in order than person B (not me) can gain the million. Concretely, suppose B is sick in bed while I hear that his last chance to claim an inheritance is about to pass, and I (whose driver's license has been revoked) instruct A (who I know will obey) to drive to the courthouse to pick up the papers for B to sign, thereby placing A's life in jeopardy (of death from an accident). If we are truly unwilling to trade off life for money, this instruction would also be an immoral one. However, when I take into account the preferences of both parties (supposing, for example, that I must make a decision out of self-interest before I know which party I shall be), it seems clear that the risk would be worth it, provided the probability were sufficiently small. Even if we balk at this example, let us consider the fact that B is only a single person. There is only so much

utility that he can get out of money, no matter how great the amount of money. Imagine that I make you risk your life (e.g., drive two blocks) for the sake of eliminating poverty from the world forever. Thus, by combining the preferences of many people, we can extend the range of variation of any given attribute, and thus see more clearly how that attribute might trade off with others.⁶

Perhaps there is something special about risking one's life by driving, because this is a risk we take all the time without thinking about it. So let us change these examples so that the risk involves a (fantastic) game of Russian Roulette with the identical (very low) probability of death as driving two blocks. How might this change be morally relevant? For one thing, you are aware of the danger of death from Russian Roulette, but you may have forgotten that driving is also dangerous. Well, suppose, in the driving version of the story, that I remind you of the danger of driving before I make you do it; would that then make it immoral to make you, given that it was moral without the reminding? Hardly. Another possible distinction between driving and Russian Roulette is that driving is normal and Russian Roulette is not. I shall argue later that such distinctions are also morally irrelevant; they amount to framing effects in the sense of Kahneman and Tversky (1984).

One argument against my position is that the kinds of examples needed to refute lexical ordering are, for some pairs of attributes, totally hypothetical and could never occur. One answer is that sometimes we are surprised by the problems that do occur (as in many recent examples concerning transplants, genetic engineering, and reproduction). If we adopt fundamental moral principles that do not admit such surprises, we shall be at a loss when the cases arise; we shall have to start from scratch. Beyond this, the distinction between the normative and prescriptive levels is intended to accommodate just such assumptions about what could and could not happen at the prescriptive level only. The normative level thus provides justifications that are independent of such assumptions.

Another objection (following an argument of Rawls [1971] concerning the 'Difference principle') is that we cannot use hypothetical situations involving probabilities, because we may not know the probabilities in question, and we need principles that will serve us when such ambiguity is present. In such cases it is usually argued that we should err on the side of caution, minimizing the maximum possible loss. Thus, we would choose life over money when we do not know P because we would make this choice if P were sufficiently high. Thus, the argument may go on, we should always choose life over money.

There are three answers to this objection. First, the last conclusion does not follow; sometimes we do have probability information. Normatively, there is no reason we should blind ourselves to information.

⁶My claim amounts to an assertion that goes beyond the assumptions stated in footnote . If the claim is accepted, we can add additional assumptions: If $A1$ differs from $A0$ on one dimension and $A2$ differs from $A0$ on another, and if $A2 \succ A1$, then there exists $A3$ such that $A3$ differs from $A0$ on the same dimension as $A2$ and such that $A3 \sim A1$. Also, there is a gamble, $A2$ in state $a2$ and $A0$ in state $a0$, such that the decision maker is indifferent between the gamble and $A1$. (Alternatively, these statements could be derived from some sort of Archimedean assumption; see Krantz et al., 1971.)

Second, there is no good reason that ambiguity should lead to risk aversion of this sort. Arguably, when one has no information about probability, this amounts to saying that there is no reason to think that either possible state is more likely than any other. In such cases, one's best bet would be to assume that the probabilities of the two states are equal. If we do have information about the probabilities, then that information ought to allow us to make an informed guess about just how likely, and it is better to act according to this guess than to assume the worst. If we do not think our best guess about the probability is our best guide to action, then it is not our best guess (see Savage, 1954).

Third, when we adopt principles in advance for a class of situations, ignoring the information that will distinguish members of the class, we are adopting prescriptive rules, not normative ones. In normative theory, we are free to make decisions about individual cases in which probabilities are actually known. The only reason for stating a rule for cases with totally unknown probabilities is that we want to anticipate many such cases before knowing their details. When we do this, we must bind ourselves to follow the rule, despite any knowledge we might come to have about individual cases. If there is reason to bind ourselves at all in such a way (and if we can actually do it), there must be a strong justification for throwing away the information we have about specific cases. Such a justification would be exactly of the sort used more generally to make prescriptive rules and to justify them at the normative level (e.g., slippery slopes and self-serving bias). Lexical rules may well, as I have admitted, have a place at the prescriptive level, but this is not an argument for their having a place at the normative level.

In summary, given any claim that one attribute is normatively prior to another, we could always counter that claim by imagining cases in which the former attribute is only probabilistically relevant or in which the second is multiplied by being relevant over many people. Thus, there will never be any case in which we feel, on reflection, that one attribute is infinitely more important than another. Attributes can always be traded off. If this is true, then it becomes at least more difficult for any normative theory to specify the relative weight of one attribute against another. If the relative weight is infinite, there is no need to measure and compare differences. If not, then there can always be disagreement among people about the value of P at which a gamble becomes worth taking, or how many people must benefit before a loss of some sort is justified. Given that such disagreement is possible within the bounds of normative theory, it is difficult to say that any particular person is wrong in any particular case about the relative weight to be given to two considerations, e.g., life and money. Later, I shall suggest some ways in which this difficulty may be partially overcome.

3 Implications

3.1 Simple moral systems

One common type of moral system, to which most of us subscribe to some degree, consists of a manageable list of moral rules, usually of the form, ‘Thou shalt not ...’ The rules have absolute priority, but as long as we follow the rules, we are free to ignore moral considerations completely, if we so choose. No tradeoff is allowed between violation of the rules (or even, in the extreme, risk of violation of the rules) and moral considerations outside the rules (e.g., effects of our actions on other people’s preferences). The term ‘simple’ indicates only that morality is defined by lexical rules rather than tradeoffs (which require judgment).

There are various variants of this basic theme. Many rights-based moral theories (Dworkin, 1977; Frey, 1984; Nozick, 1974) would fall into this category: one type of lexical rule is that a certain type of right should never be violated. (These theories, of course, are not simple at all in other senses of the term.)⁷ Outside of philosophy, such theories are alive and well in the American Civil Liberties Union, the Libertarian Party, the followers of Ayn Rand, and both sides of the abortion debate. Such organizations justify their stands in terms of various rights such as the rights of the biological parents of adopted children (to raise the children if they wish), the right to keep what one earns, the right to control one’s body, or the right to life. Each such right is assumed to be lexically prior to some other right. Usually, salient counterexamples are handled by the addition of stipulations and conditions, e.g., ‘Free speech is an inalienable right, except that one may not yell ‘fire’ in a crowded theater, sell pornography to minors, etc.’ What is explicitly rejected is the idea that rights, or duties, once stated, may be traded off with *any* sort of consideration under appropriate conditions; this is, of course, the idea I advocate.

3.2 Omission and commission

Simple moral systems are often maintained by a distinction between acts of omission and acts of commission. For example, a rule within some systems of this sort says, ‘Thou shall not kill a person (except in self-defense or war).’ It is immoral to kill someone to prevent him from stealing \$50 from you (supposing there is no other danger). The same moral system might applaud a person who contributes money to prevent starvation in the Sahel, but it is not immoral to *fail* to contribute \$50 so that a life there may be saved. One is enjoined to avoid killing at (nearly) all costs, yet, the taking of positive action to save life is a moral nicety, not a requirement.

⁷In Ross’s (1930) scheme of ‘*prima-facie* duties’ (keeping promises, helping others when we can, etc.) it is possible for any duty to be overridden by another. However, the claim of an overridden duty still holds after the overriding (p. 28; Williams, 1981, makes the same sort of argument). One way to interpret this is to say that Ross’s system is lexical as a moral theory but not as a decision theory.

The distinction between omission and commission may be seen as the result of a cognitive illusion, specifically, a framing effect (Kahneman and Tversky, 1984). When we think of a certain outcome as normal, expected, or part of the status quo (e.g., death from starvation), then we see the prevention of that outcome as an act of commission. If an outcome is not expected, then we see causing it as an act of commission. We tend to feel that the evil done by tearing up one's check to Oxfam on the way to the post office is greater than the good done by (someone else) writing the check for the same amount, and we institutionalize feelings of this sort into our prescriptive moral codes. Similarly, we are happy with a discount for paying cash (compared to a credit card) but unhappy with a surcharge for using a credit card (compared to cash; see Kahneman & Tversky, 1984); we favor a tax deduction for charity but oppose an equivalent government subsidy to charities; and we oppose 'heroic measures' to save a doomed newborn but favor equally expensive and futile 'standard procedures.' These distinctions are normatively irrelevant because they amount to *redescriptions of the same situation*. We can avoid these framing effects by focusing on the *difference* between the consequences of two actions we might take. In the case of the newborn, for example, what matters is whether the benefit to be gained from some treatment — standard or not — is worth the cost (including the cost in resources taken away from others in need) and risk. (See Singer, 1979, chapters 7 and 8, for further defense of this position.)

A related example concerns the changing of bureaucratic procedures. Dawes (1976, 1979) has argued that the use of simple mathematical formulas for decisions such as admission of students to graduate school can reduce mistakes and save effort, but his arguments have been ignored. One possible cause (among others) of this neglect is that we imagine the harm done to those (few) students who would be rightly accepted under the old system but wrongly rejected by the formula, and we ignore the harm done to those (relatively many) who would be rightly accepted by the formula but wrongly rejected by the old system. The harm done from changing the rules is seen as an act of commission, hence a moral no-no, but the improvement that results from changing the rules is seen as a moral nicety. Because the avoidance of moral harm is lexically ordered (in our prescriptive system) ahead of the benefit of positive moral acts, we avoid change, even though net benefit would increase.

What arguments could justify a *prescriptive* system based on a distinction between omission and commission? Slippery-slope and self-serving-bias arguments do not seem to apply here. One possibility is that prescriptive rules are designed to be easily recalled, given certain triggering conditions. The rule, 'Do not take money,' is triggered by the intention to take money, but the rule, 'Give money,' requires an additional condition (when? every Sunday?) It is harder to find out whether one has given to those in need than whether one has been a thief. Also, many prohibitions leave no ambiguity about the behavior that satisfies them (e.g., not taking any amount of money) but positive injunctions are usually open ended (how much should I give?) These differences seem insufficient justification. The function they serve is to permit us to ignore the good we might have done (but didn't do) and thereby to convince ourselves that we

are being 100% moral.

3.3 Morality and self-interest

Simple moral systems put morality ahead of self-interest. Morality, it is claimed, should be lexically prior to any considerations of self-interest. Prescriptively, the priority of morality over self-interest has the advantage of simplifying moral decision making. People are not burdened with constantly having to make decisions between morality and self-interest. If I follow the simple rule, ‘keep promises regardless of self-interest,’ I do not have to weigh my self-interest against the moral value of each promise. Further, when self-interest and morality conflict, people have a self serving reason to ignore the moral, and our judgments might be distorted. Slippery-slope arguments may also be made; once I allow self interest to override moral principles, I am on my way to pure selfishness. However, the disadvantage of the priority of morals (normatively and prescriptively) is that it limits moral considerations to those that are easily stated as simple lexical rules, usually prohibitions.

Once we drop the distinction between omission and commission, the consequences of a theory in which we must weigh all interests equally are profound (Singer, 1979, ch. 8). For example, suppose I have been persuaded that my failure to donate \$50 to save the life of a starving child is morally equivalent (roughly) to my killing someone for \$50. Of course, the \$50 means less to me than to a person who would die without it, so I would increase net benefit (and surely net good as well) by giving it away. This type of reasoning could be used over and over, a sufficient number of times so that my own resources would be substantially depleted if I gave all the money that I morally should. If I weigh my self-interest more than the interests of others, however, I could make a decision that would be normatively correct as a decision (although not morally correct) if I gave far less than this, because of the weight I given to my (and my family’s) self interest. The only normative requirement is that I weigh such self-interest consistently, by the same factor in every case. Otherwise, I subvert my own ends (Baron, 1985, ch. 2); for example, if I donate \$1000 to save a single life through heroic medical treatment, I may lose the chance to save 20 lives through immunization, and I therefore save fewer lives overall.⁸

At the prescriptive level of a moral-decision theory, equal weighing of self and others is self-defeating, for it is apparently impossible in practice. Later, I shall suggest a specific prescriptive rule for relative weighing of self and others. (At the normative level, the appropriate relative weight depends on the nature of the hypothetical creature the theory is for [Baron, 1985, ch. 2]. For an educator [who represents society] a student’s personal [non-moral] plans are relevant at least because the moral domain is enriched by our chance to facilitate each other’s personal plans. From the student’s perspective, moral concerns may enrich personal plans. These considerations may not yield a unique relative weight, but they may establish an appropriate range.)

⁸See Hare, 1981, 11.6-11.7, and Singer, 1979, ch. 8, for other proposals on resolving this sort of problem.

3.4 Self-referential altruism

In his theory of ‘self-referential altruism,’ Broad (1942) argues for a set of ‘concentric circles’ of obligations, e.g., first to family, then to friends and colleagues, then to nation, and last to humanity as a whole. Theories of this sort seem to be held by many people, and for some who hold them, there is an apparent lexical ordering. Some nationalists (such as my congressman) regard the national interests of their country as lexically prior to the interests of people elsewhere. Foreign aid must be justified entirely by national interest.

Broad, whose distinction between derived and ultimate obligations (pp. 55-56) anticipates the prescriptive-normative distinction, justifies such principles as derived (prescriptive) obligations by arguing that our limited knowledge and power enables us to serve best the interests of those closest to ourselves. ‘... in consequence of this, the maximum balance of good over evil among conscious beings as a whole is most likely to be secured if people do not aim directly at it. It is most likely to be secured if each aims primarily at the maximum balance of good over evil in the members of a limited group consisting of himself and those who stand in more or less intimate relations to him.’ In addition, says Broad, nations whose citizens concern themselves with humanity as whole may not survive.

It follows from the arguments I have made that the lexical form of such a system is normatively unjustifiable. Prescriptively, the arguments in favor of such a system do not justify lexical rules either. Broad’s arguments do not imply that the ‘bias’ toward the inner circles must be lexical. Aside from these arguments, these biases may be justified in terms of the goodness of the institutions that create the obligations in question (and the prescriptive rules they imply), e.g., the family, the nation, and the institution of making commitments itself, which in turn creates additional obligations to spouses, children, etc. The point here is that these arguments for prescriptive rules in favor of biases toward family over foreign strangers are not arguments for *lexical rules*. The biases may perfectly well be quantitative, just as the bias toward self-interest discussed in the last section is quantitative.

3.5 Cooperation, the prisoner’s dilemma, and the commons dilemma

One type of moral argument is of the form, ‘What if everybody did that?’ (see Regan, 1980). If the answer is, ‘Everyone would be better off,’ then, it is argued, everyone should take the action in question, and vice versa. This argument can be applied to such actions as paying one’s taxes honestly, not building arms in the context of an arms race, not accepting bribes, not polluting the environment, and not having too many children. These situations are often referred to as ‘commons dilemmas’ because they are analogous to grazing one’s cow on a common pasture (Hardin, 1968; Schelling, 1978). Many of the prescriptions of simple moral systems (e.g., ‘Do not lie’) are also justified by this cooperative principle. Another use is in the justification of such morally ‘pure’ actions as

refusing to buy the products of a poorly behaving company or government (even when there is no chance of affecting the offending policy).

Laboratory analogs of commons-dilemma situations have been developed, and these are usually called ‘prisoner’s dilemmas’ (based on a two-person analog, even though the situations of interest involve many people; see Dawes, 1980). In general, these situations face people with a choice of two actions, C (cooperate) and D (defect), with the cooperative principle standing squarely on the side of C. At issue are the normative and prescriptive models for such situations. It is implicit in the literature that C is always the correct response; this cooperative principle seems to be the normative model. Indeed, Dawes (1980) defines the relevant situations as those in which D increases self benefit (holding constant the actions of others) but the benefit of *each* is maximized if *all* choose C.

One problem with the cooperative principle is that it often conflicts with the maximization of net benefit or goodness (Regan, 1980). If everyone has already trampled the grass to death, why should I be the only one to obey the ‘Keep off the grass’ sign? I would only decrease net benefit by hurting myself. (Let us put aside such arguments as the fact that by visibly walking around the grass I might encourage others to do likewise.) In prisoner’s-dilemma situations, C need not be the morally correct response, if it does not increase overall benefit or good given the actions that others can be expected to take. It would make sense to replace Dawes’s definition with one (based on Pruitt, 1967) in which the relevant situations are those in which C maximizes net benefit over all players but D maximizes benefit to the self.

Following the arguments of the last section, if we take self-interest into account, D may be a normatively appropriate response even if C would maximize overall benefit or goodness. The temptation to win for oneself by defecting may be legitimately too great. For example, if a subject in a prisoner’s dilemma experiment had a general policy of weighing his own self-interest 10 times as much as the interests of anyone else *consistently* (a policy that might make him one of the more generous contributors to Oxfam, if he thought that the marginal utility of \$1000 to him was less than a tenth of its utility to the 20 people whose lives it could save), he would have perfectly good reason to take advantage of an opponent who looked like a good sucker.

In sum, the appropriate normative principle would seem to require, first, cooperation only when cooperation can accomplish some net good, and, second, *consistent* tradeoffs (across situations) between self-interest and the expected benefits of cooperation. When the benefit from D is overwhelming, I can be excused for giving in to the temptation (although I can still be *morally* censured for doing so). Investigators of social dilemmas should not assume that C is always the normatively correct response.⁹

⁹These two principles together, the requirement that cooperation do some good and the legitimacy of self interest, can lead to a kind of ‘When in Rome, do as the Romans’ relativism. The argument, ‘Why shouldn’t I accept favors if everyone else in my position does it?’ does have some normative force, especially if there is little chance of affecting the practice by setting an example. Of course, this is not true relativism, because the basic principles are unchanged across cultures.

3.6 Guilt

There are times when we are put in a situation in which any action we take seems to violate a prescriptive moral rule. For example, I was recently in a situation in which I had to disobey the rules of my university (which required that I abide by the judgment of an appeals court) or else leave a case of blatant cheating go essentially unpunished. Such dilemmas are especially common for holders of simple moral systems, unless these systems are carefully designed to insure that rules never conflict (e.g., Rawls, 1971). Some (e.g., Williams, 1981) have suggested that in such situations it is impossible to do what is right. This problem disappears when we take the normative point of view in which tradeoffs are always possible. One must simply choose the course that does the least harm, all things considered. If our moral emotions are tied to the normative level, we will not feel guilt after we do the best we can.

Guilt, then, may often be traced to the use of a simple moral system. (It may also result from our tendency to evaluate decisions and actions as if bad outcomes are always caused by bad decisions or actions; e.g., Walster, 1966). Guilt may always be with us, for we may always violate our own prescriptive rules, but some of our guilt may be avoidable.

4 Toward a prescriptive theory

The question of the best prescriptive theory is not entirely a philosophical one. If we are concerned about how best to approximate the results of the normative theory in practice, we must take into account the facts of human psychology, of the culture in question, and of the world situation (see Hare, 1981, chapter 10.2). But even these facts do not dictate the best theory. The construction of a prescriptive theory is a problem of *design*, the design of child rearing, of education, of institutions, and, most generally, of ways of living. Such questions of design are the ones that arouse the fiercest and most passionate conflicts. Philosophy, by clarifying and answering questions about normative theory, may help to allay some of these conflicts, but not all of them.

With this warning, let me tread into deeper waters and make some proposals about a prescriptive theory ‘for our time.’ I propose that our actual thinking about moral questions should become closer to the normative model of decision-making and rational life-plans that I have sketched, and, as a result, farther from any sort of simple moral system. We should think about consequences of our decisions for preferences of relevant people, including our own preferences, and about other relevant attributes of our actions (if any), and we should weigh the relevant factors without being influenced by framing effects or lexical rules. We should give as high a weight as we can manage to moral considerations, and then act according to the more general (not specifically moral) normative theory (as closely as we can) without guilt, knowing that we are doing the best we can.

By this view, we can criticize a person morally in two ways: for having a life plan that gives too little weight to moral considerations and for failing to act in

accord with his own plan. We cannot criticize him for failing to do all he could morally.

The sacrifice of a reasonable amount of money on behalf of famine victims — an amount commensurate with a higher weight for self-interest than for the victims' interest but not commensurate with pure utilitarian neutrality — is an example of the type of reasoning I advocate.

Since we have given up the protection of lexical rules, we must also be aware of the distortions that can affect moral reflection, in particular, self-interested rationalization (e.g., 'My donation to Oxfam won't help anyway'), and intellectual bias (e.g., 'Assassinating X will do more to save lives than would millions in charity'). We may try to avoid these distortions by following rigid prescriptive rules (e.g., 'Don't assassinate anyone'). But, in a world in which it is difficult — for good reason — for any moral authority to force his views on everyone else, these prescriptive rules must be seen as voluntary commitments to bind ourselves, made with full understanding of the reasons they are required and of the fact that they are not normatively absolute. They are analogous to the ex-smoker's rule: 'Do not have one cigarette, even though you know just one will do no harm.'

We must also respect the principles of good thinking itself (Baron, 1985a). These include a willingness to think in proportion to the importance of the topic. It has often been argued that evil is more often a result of thoughtlessness than of sustained thinking that reaches incorrect conclusions. People need to be taught to think about moral issues as part of being taught to think about everything else that is worthy of thought. Good thinking also requires a fairness to possibilities (e.g., possible actions) other than the first one to enter one's mind. This requires some effort or willingness to consider alternative possibilities, to search for evidence against one's favored possibility as well as evidence in favor of it, and to take account of evidence without regard to whether it supports what we favor or not. For moral thinking, relevant evidence usually involves the preferences of others, so we must make an effort to consider many points of view and to assess preferences accurately from each point of view. These injunctions about fairness are especially important when we are motivated — e.g., by self-interest or by pride in our ability to guess — to believe that our initial answer is correct. They apply to the thinking we do about our life plans as well as the thinking we do about individual decisions. In this case, they make for a certain kind of honesty, a lack of self-deception.

I am not asking that we think through every little decision. We may still use rules such as 'Honesty is the best policy,' or, 'Always keep your promises,' in everyday interaction. But when we turn our attention to questions of public policy, important decisions involving life and death, or decisions about the overall course of our individual lives, these rules are not enough. A fact of our culture is that we seem to be faced with more and more of these decisions.

These arguments imply that moral education should place less emphasis on the kind of rules that simple moral systems contain and more emphasis on a kind of reflective thinking in which moral considerations are *routinely* brought into play, among others. We should try to convince the young that they will

be happier (more worthy of the respect of others and themselves, more part of the community of humanity, more purposeful in their lives) if they incorporate moral goals into their life plans. An example of the need for this is the way many people, for example, college students, choose their life work. People who regard themselves as completely moral will absolutely fail to take moral considerations into account in this most important decision. Decisions are often made largely on the basis of various personal preferences for ambience and style. Very few students have anything resembling a theory of how they might fit into the world so as to help their fellows and make a reasonable living at the same time. Even those who pursue obvious careers of service, such as medicine, rarely give their desire to serve as a reason for choosing the career. The same could be said for careers in research, except that the ideology of self-indulgence (of one's curiosity and the like) has become such a part of the institution of research that it is hard to call this a career of service. The ultimate moral effect of large numbers of young people simply admitting moral reasons as one determinant (among many) of their career plans might have a larger effect for the good than an equivalently effortful increase in their conformity to simple moral systems. A bi-product might even be that people would take the more elementary rules (e.g., do not cheat) as more a matter of course, once we had incorporated moral goals into their life plans.

Lexical prescriptive rules (such as prohibitions against killing except in self-defense or war), in contrast to prescriptive rules of thumb (e.g., 'Don't lie'), cannot be maintained in a society in which people understand the normative theory that justifies them, for the same normative theory dictates that the rule does not always apply. If the normative theory were understood by all, there would be only rules of thumb, no absolutes except the theory itself. As educators (which most of us are at some time or another), we would be faced with a choice of withholding knowledge of the normative model from our students or risking destruction of their faith in absolute prescriptive rules (assuming, of course, that we accept the normative model ourselves). To withhold knowledge of the normative theory throughout a person's education (as opposed to withholding it until the student is ready to understand it) amounts to a kind of paternalistic deception that is hard to justify on a large scale. It would presuppose some sort of moral elite who understand the rules that everyone else is taught to accept without question. At the very least, those who question spontaneously should be initiated.

The same objections made to the normative model may also be made to a prescriptive model based on it, and in many cases they have more force. One is that situations in which certain tradeoffs must be made simply will not occur, or will be so infrequent that it is best not to try to think normatively. For example, situations in which the loss of life is traded off for property rights will simply not occur, and even when they do it is on the average better to save life. However, the conflict between the beneficiaries of utility companies that pollute the air and the people who lose their lives from diseases caused by air pollution suggest that there may be a real tradeoff here. Would we (if it came to that) be prepared to double our electric bills to pay for anti-pollution equipment? For

any number of lives saved? I suspect not. (See Cohen, 1983, and Fischhoff et al., 1980, however, on how we might save more lives by being more consistent.)

The slippery-slope argument may also be offered as an argument for lexical rules at the prescriptive level. However, this argument simply will not work once a person understands the normative theory. If I took it seriously, I would have to convince myself that I must not oppose pornography lest I be cajoled into opposing political speech as well, that I must not condone euthanasia lest I be cajoled into condoning murder, etc. And if I could convince myself of this, I would have to go on to convince myself that euthanasia is just as bad as murder, in order to include it in a lexical prescriptive rule. Or I might hypocritically denounce euthanasia, thinking that others, but not I, could go down the slippery slope. Slippery-slope arguments gain their force mainly (perhaps only) when we distrust *others* to reason as well as we ourselves can reason. They are associated with the same kind of paternalism I discussed earlier, in which justifications are withheld from the masses by a moral elite.

A better argument for lexical prescriptive rules is that they protect us against the effect of self-serving biases, in which we justify the pursuit of our self-interest by referring to an imagined utilitarian argument. Rules against breaking promises, cheating on one's taxes, embezzling, etc., might fall in this category, because it is easy to justify such actions. (I'll give less money to the government, which only wastes it, and more to charities. I'll take a little off the top to make up for the raise my employer didn't give me.)¹⁰ One problem here is that such rules are not very effective. A great many people, it seems, do break promises (e.g., to return my books) and cheat a little on their taxes. Most people seem to think of rules of this sort as rules of thumb rather than absolutes in any case. On the other hand, Ainslie (1975) and Thaler and Shefrin (1981) have argued that such rules of thumb are essential elements of self-control. But they will serve this function if they are seen as devices for self-control, not moral absolutes.

Another argument against the claim that lexical rules are required to prevent self-serving bias is that, in the prescriptive system I advocate, the incentive for self-deception would be reduced. In simple moral systems, it is sometimes very difficult to follow a lexical rule, because self-interest weighs so heavily on the other side. If one can deceive oneself into thinking that no rule is violated, the difficulty is removed. In the prescriptive system I advocate, self-interest is a recognized and legitimate consideration in all decisions. One need not ignore it; rather, one should try to be consistent in weighing it against the interest of others, and to give those interests as much weight as one comfortably can. One might strive for a level of weight that would maximize net benefit, including the discomfort of self-control, if everyone were to adopt roughly the same level.¹¹ With such a rule, one need not deceive oneself at all. Moral self deception may be a symptom of unrealistic standards.

¹⁰Broad (1942) argues that (prescriptive) obligations are strongest when the temptation to break them is strongest.

¹¹The same rule might be part of a normative moral-decision theory, as well as the prescriptive one described here.

Changes in the direction these arguments imply may stress the system of corporate capitalism. This system is to some extent dependent on the tendency of young people to live by simple moral systems. It offers the young, to varying degrees, a chance to pursue their self-interest by working for a salary and then using the salary to purchase the fruits of the labor of others, in a cycle of production and consumption. The theorists of capitalism know that this system accomplishes a great deal to satisfy real human preferences, but the individuals in the system have difficulty seeing themselves as moral actors, as cooperating in a scheme that works to the benefit of others as well as themselves. They are encouraged to see themselves as workers and consumers, pursuing self-interest alone. As a result, morality and work are separated in people's life plans. From the prescriptive point of view I advocate, they ought to be intimately related. Work consumes so much of our efforts that it is our greatest opportunity to have a moral effect, and if we can have such an effect and still make a living, we surely should do so.¹²

Another area in which moral considerations are now often neglected is science. Many scientists seek to pursue truth for its own sake, with certainty being the only goal (lexically ordered above all others), regardless of social importance of the questions asked. It is often said that social importance cannot be foreseen, and therefore should be ignored. In fact, one who examines the situation honestly will find that social importance and application can be guessed with sufficient accuracy to make them relevant considerations in a theory that considers probabilistic consequences as relevant.

The prescriptive theory I have proposed requires a number of assumptions about what is possible. It requires that people learn to weigh conflicting considerations honestly, without self-deception. This may be impossible, but we cannot know this unless we make the effort. If we are to make the effort, it is clear that we will have to emphasize personal honesty (the absence of self-deception) as a primary virtue.¹³

5 Conclusion

Let me return to the husband dilemma. If the husband decides not to steal the drug, because he happens to weigh the owner's preferences above his wife's, can we say he is wrong? By the normative view I have outlined and the prescriptive view I have advocated, we cannot say so *for sure*. Depending on the circumstances, however, we might guess that the husband has deviated from the prescriptive theory. He might have weighed his self-interest more in this case than he would in others or more than people in general ought to weigh our self-interest. He might also have deceived himself — in order to protect his

¹²Radical change might not be required. Employers might simply have to compete for workers in part on the basis of the goodness — and lack of badness — of their projects. Arguably, some of this happens now, as reflected in the payment of lower salaries to people who perform obvious service and higher salaries to the hired guns of corporate law. If you can't compete on moral grounds, you have to put cash on the barrel.

¹³Elster, 1983, discusses several related concepts.

self-interest — in his estimates of the effects of his decision on wife and owner. Whether he was morally wrong thus depends not on his decision nor even on his stated arguments for it, but rather on whether the psychological process of reaching the decision was an honest one made according to the prescriptive theory, without self-deception. This is something we cannot know for sure in a given case, although we can and do have our suspicions. Although we cannot know whether a given decision was honestly made, we can teach people (I hope) how to make decisions honestly.

Am I a relativist? Yes, of a sort, if we look at the content of moral decisions and the overt arguments for them. There are certain decisions that would arouse my suspicion, but very few I could regard as utterly beyond an honest justification. No, if we look at the processes behind decision making. There is a right way to think about moral questions. It involves fair consideration of preferences of all relevant people, an effort to criticize one's first conclusions and to remain open to alternatives and counterevidence, and a degree of time and effort that is commensurate with the importance of the decision. Poor moral thinking involves insufficient search, especially for evidence (including relevant beliefs) concerning effects of decisions on others, lack of self-criticism, and self-deception. Poor moral thinking is indeed a cause for criticism.

References

- Ainslie, G. (1975) Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82, 463–496.
- Baron, J. (1985a) *Rationality and intelligence*. Cambridge University Press.
- Baron, J. (1985b) Rational plans, achievement, and education. In M. Frese & J. Sabini (Eds.), *Goal directed behavior: the concept of action in psychology*. Hillsdale, N.J.: Erlbaum.
- Broad, C.D. (1942) Certain features in Moore's ethical doctrines. In P.A. Schlipp (Ed.), *The philosophy of G.E. Moore*. New York: Tudor.
- B.L. Cohen (1983) Risk and risk aversion in our society. *Ceramic Bulletin*, 62, 1285–1288.
- Dawes, R. (1976) Shallow psychology. In J.S. Carroll & J. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum.
- Dawes, R. (1979) The robust beauty of improper linear models. *American Psychologist*, 34, 571–582.
- Dawes, R. (1980) Social dilemmas. *Annual Review of Psychology*, 31, 169–193.
- Dworkin, R. (1977) *Taking rights seriously*. Cambridge, MA: Harvard University Press.
- Elster, J. (1983) *Sour grapes: studies of the subversion of rationality*. Cambridge University Press.
- Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S.L., & Keeney, R.L. (1981) *Acceptable risk*. New York: Cambridge University Press.
- Frey, R.G. (Ed.) (1984) *Utility and rights*. University of Minnesota Press.

- Fischhoff, B., Gotein, B., & Shapira, Z. (1982) The experienced utility of expected utility approaches. In N. Feather (Ed.) *Expectations and actions: expectancy-value models in psychology*. Hillsdale, NJ: Erlbaum.
- Hardin, G.R. (1968) The tragedy of the commons. *Science*, *162*, 1243–1248.
- Hare, R.M. (1963) *Freedom and reason*. Oxford University Press.
- Hare, R.M. (1981) *Moral thinking: its levels, method and point*. Oxford University Press.
- Kahneman, D., & Tversky, A. (1984) Choices, values, and frames. *American Psychologist*, *39*, 341–35.
- Keeney, R.L., & Raiffa, H. (1976) *Decisions with multiple objectives*. New York: Wiley.
- Kohlberg, L. (1970) Stages of moral development as a basis for moral education. In C. Beck & E. Sullivan (Eds.), *Moral education*. University of Toronto Press.
- Nozick, R. (1974) *Anarchy, state, and utopia*. New York: Basic Books.
- Pruitt (1967) Reward structure and cooperation: the decomposed prisoner's dilemma game. *Journal of Personality and Social Psychology*, *7*, 21–27.
- Rawls, J. (1971) *A theory of justice*. Cambridge, Mass.: Harvard University Press.
- Regan, D. (1980) *Utilitarianism and cooperation*. Oxford: Oxford University Press.
- Ross, W.D. (1930) *The right and the good*. Oxford: Oxford University Press.
- Richards, D.A.J. (1971) *A theory of reasons for action*. Oxford University Press.
- Savage, L.J. (1954) *The foundations of statistics*. New York: Wiley, 1954.
- Schelling, T.C. (1978) *Micromotives and macrobehavior*. New York: Norton.
- Simon, H. (1957) *Models of man: social and rational*. New York: Wiley.
- Singer, P. (1979) *Practical ethics*. Cambridge University Press.
- Thaler, R.H., & Shefrin, H.M. (1981) An economic theory of self-control. *Journal of Political Economy*, *89*, 392–406.
- Walster, E. (1966) Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, *3*, 73–79.
- Williams, B. (1981) *Moral luck*. Cambridge University Press.