Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*, 569-579.

Outcome bias in decision evaluation

Jonathan Baron and John C. Hershey
University of Pennsylvania[1]

Abstract

Subjects were given descriptions of decisions made by others under conditions of uncertainty, together with outcomes of those decisions. Some decisions were medical decisions made by a physician or a patient, and others were decisions about monetary gambles. Subjects rated the quality of thinking that went into the decisions, the competence of the decision maker, or their willingness to let the decision maker decide on their behalf. Subjects understood that they had all relevant information available to the decision maker. Subjects rated the thinking as better (or rated the decision maker as more competent, or indicated greater willingness to yield the decision) when the outcome was favorable than when it was unfavorable. In monetary gambles, subjects rated the thinking as better when the outcome of the option *not* chosen turned out poor than when it turned out well. Subjects who were asked felt that they should not take outcomes into account in making these evaluations, yet, they did so. In part, this effect of outcome knowledge on evaluation may be explained in terms of its effect on the salience of arguments for each side of the choice. Implications of these findings for the theory of rationality and for practical situations are discussed.

> `A fault condemned but seldom avoided is the evaluation of the intention of an act in terms of the act's outcome. An agent who acted as wisely as the foreseeable circumstances permitted is censured for the ill-effects which come to pass through chance or through malicious opposition or through unforeseeable circumstances. Men desire to be fortunate as much as they desire to be wise, but yet they fail to discriminate between fortune and wisdom or between misfortune and guilt ... We are ingenious in 'discovering' the defect of character we believe would account for a person's misfortune.' (Arnauld, 1662/1964, p. 285)

> `Since good decisions can lead to bad outcomes (and vice versa) decision makers cannot infallibly be graded by their results.' (Brown, Kahr, & Peterson, 1974, p. 4)

> `A good decision cannot guarantee a good outcome. All real decisions are made under uncertainty. A decision is therefore a bet, and evaluating it as good or not must depend on the stakes and the odds, not on the outcome.' (Edwards, 1984, p. 7)

Evaluations of decisions are made in our personal lives, in organizations, in judging the performance of elected officials, and in certain legal disputes, such as malpractice suits, liability cases, and regulatory decisions. Because evaluations are made after the fact, there is often information available to the judge that was not available to the decision maker, including information about the outcome of the decision. It has often been suggested that such information is used unfairly, that reasonable decisions are criticized by Monday-morning quarterbacks who think they might have decided otherwise, and that decision makers end up being punished for their bad luck (e.g., Arnauld, 1662/1964; Berlin, 1984; Nichols, 1985).

The distinction between a good decision and a good outcome is a basic one to all decision analysts. The above quotation from Edwards (1984) is labeled by the author as `a very familiar elementary point.' In this paper, we explore how well the distinction between decisions and outcomes is recognized outside the decision-analysis profession.

Information that is available only after a decision is made is irrelevant to the quality of the decision. Such information plays no direct role in the advice we might give decision makers *ex ante* or in the lessons they might learn (Baron, 1985, ch. 1). The outcome of a decision, by itself, cannot be used to improve a decision unless the decision maker is clairvoyant.

Information about possible outcomes and their probabilities falls into three relevant classes: *Actor Information*, known *only* to the decision maker at the time the decision is made; *Judge Information*, known *only* to the judge at the time the decision is evaluated; and *Joint Information*, known both to the decision-maker at the time of decision and to the judge at the time of evaluation. (In some cases, the decision maker and the judge will be the same person, at different times.) In the cases we consider, the judge has the outcome information, and the actor does not.

Although outcome information plays no direct role in the evaluation of decisions, it may play a very appropriate indirect role. In particular, it may affect a judge's *beliefs* about Actor Information. A judge who does not know the decision-maker's probabilities may assume that the probability was higher for an outcome that occurred than for the same outcome, had it not occurred. (Note, however, that outcome information tells us nothing about the *utilities* of a decision maker, even if we have no other information about them.) In the extreme, if we have no information except outcome, it is a reasonable prima-facie hypothesis that bad outcomes (e.g., space-shuttle accidents) result from badly made decisions. We do not usually set up commissions of inquiry to delve into policy decisions that turn out well.

Another appropriate *indirect* role of outcome information is that it allows a decision maker to modify beliefs about probabilities in similar situations. If I know nothing about the proportion of red cards in a deck, I can learn something about that proportion by drawing cards from the deck. (However, if I know that the deck is an ordinary one, sampled with replacement, I learn nothing by drawing cards.) This effect of outcome information can operate only within a sequence of similar decisions, not in a single decision.

At issue here is whether there is an *outcome bias*, in which people take outcomes into account in a way that is irrelevant to the true quality of the decision. This sort of bias is not established by showing that people take outcomes into account. As just argued, outcomes are relevant when they might inform us about Actor Information. One way to show an outcome bias is to give the judge all relevant information about outcome probabilities known to the decision maker, plus the outcome. That is, there is only Joint Information and Judge Information (the outcome), no Actor Information.

Information (relevant or irrelevant) may have two effects on evaluations: 1., an effect on the judged probability of outcomes, which, in turn, affects evaluation, and 2., a direct effect on the judged quality of the decision, as shown here:

```
Outcome information  ----------------------->|      evaluation
   |                                         |----> of
   -----> judged probability of outcome ---->|      decision
```

For example, we may think a decision is bad if we believe that bad outcomes were highly probable, but outcome information may also affect our evaluation even if the probability of an outcome is known.

Fischhoff (1975) demonstrated the existence of a *hindsight bias*, an effect of outcome information on the judged probability of an outcome. Subjects were given scenarios and asked to provide probabilities for different outcomes. When subjects were told the outcome and asked what probability other subjects who did not know the outcome (or they themselves if they did not know it) would give, they gave higher probabilities than those given by actual other subjects not told the outcome (or told that some other outcome had occurred). Note that these demonstrations filled our condition of eliminating Actor Information (where the `actors' are the other subjects). Subjects were asked to judge the probability for someone who had exactly the same information they had (except for outcome), no more.

Although it seems likely that the hindsight bias would lead to biased evaluations of decision quality, this has not been shown. Nor is it what we seek to show here. Rather, *we seek a direct effect of outcome on evaluation of decisions*, an effect that does not operate through an effect of outcome knowledge on a judge's assessed probabilities of outcomes. To this end, we hold probability information constant by telling subjects that probabilities are known, or by otherwise limiting probability information. Of course, in real life, the outcome bias we seek could work together with the hindsight bias (as shown in the above diagram) to distort evaluations of decisions even more than either bias alone.

Zakay (1984) showed that managers counted good outcomes as one of the criteria for evaluating decisions made by other managers. However, as we have argued, it is perfectly reasonable to do this when there are facts known only to the decision maker (Actor Information). At issue in the present paper is not whether people *use* outcome information, but whether there are conditions under which they *over*use it. Thus, we look for an effect of outcome information when the subject is told everything that is relevant. In this case, outcome should play no role at all in our evaluations of decisions, although we hypothesize that it will.

The outcome bias we seek may be related to Walster's (1966) finding that subjects judged a driver as more `responsible' for an accident when the damage was more severe. However, questions about responsibility might be understood as concerning appropriate degree of punishment or blame rather than rationality or quality of decision-making. As a general rule, it makes sense to punish actors more severely for more severe consequences; it is usually difficult to know what the actor knew and severity of consequences is a clue to the degree of negligence. Even when we know what the actor knew, use of this general rule might set clearer precedents for others (as in the utilitarian rationale for `punishing the innocent'). Walster apparently intended the question about responsibility to tap subjects' beliefs about the extent to which the driver could have prevented the accident by acting differently. Walster suggested that her results were due to subjects' desire to believe that events were controllable: if bad outcomes are caused by poor decisions or bad people, we can prevent them by correcting the decision-making or by punishing the people. If subjects interpreted the question this way, they would be making an error, but not the same error we seek in the present study.

Similarly, studies of the effect of outcomes on children's moral judgments (e.g., Berg-Cross, 1975; Leon, 1982; Stokes and Leary, 1984; Surber, 1977) have used judgments of responsibility, deservingness of punishment, or `badness,' which could be appropriately affected by outcome. Also, in most cases, no effort is made to provide the judge with all relevant information available to the actor.

Mitchell and Kalb (1981) also showed effects of outcome knowledge on judgments of both responsibility for outcomes and outcome probability. Subjects (nurses) read descriptions of poor performance by nurses (e.g., leaving a bed railing down) which either resulted in poor outcomes (e.g., the patient fell out of bed) or benign outcomes. In fact, outcome knowledge affected both probability judgments and responsibility judgments. Although the former effect may be a hindsight bias, it may also be an appropriate inference about Actor Information: outcome information might have provided information about factors that affect outcome probability from the decision maker's point of view (e.g., whether the patient was alert, and, if not, whether she slept fitfully). Mitchell and Kalb argued that the effect of outcome on probability did not explain the effect on responsibility judgment: the correlation between judged probability and judged responsibility, with outcome held constant, was nonsignificant across subjects. Of course, the problem still remains that the term `responsibility' need not refer only to quality of the decision.

In the present experiments, instead of looking at the correlation between outcome judgments and probability judgments, we fix the outcome probabilities by telling the subject what they are from the decision maker's point of view. We also explicitly ask about `quality of thinking.' All decisions are expressed in the form of gambles. For example, an operation may lead to a cure or to death, with given probabilities. We give the subjects probabilities of all possible outcomes and brief descriptions of each outcome. It is reasonable to assume that the quality of the decision depends on the probabilities of the outcomes - which summarize all the information we have about uncertain states of the world that could affect the outcome - and the desirabilities or utilities of the outcomes. Although we do not provide all necessary information about desirabilities, the outcome provides no additional information on this score. We shall say that an outcome bias exists if the evaluation of the decisions depends on their outcomes.

Why might we expect to find an outcome bias? The main reason is that the generally useful heuristic of evaluating decisions according to their outcomes might be overgeneralized to situations where it is inappropriate. It might be learned as a rigid rule, perhaps from seeing punishment meted out for bad outcomes resulting from reasonable decisions.

Of course, it can often be appropriate to use outcome information to evaluate decision quality. This has been termed the `frequentistic' way of thinking about decision making under uncertainty (Vlek, 1984, pp. 22-23). But this will be most useful when Actor Information is quite substantial relative to Judge Information or Joint Information, and it is necessary to judge decisions by their outcomes (as fallible as this may be) simply because there is little other information to go on. This is especially true when decision makers are motivated to deceive their evaluators about the nature of their own information.

Ordinarily, it will be relatively harmless to overgeneralize the heuristic of evaluating decisions according to their outcomes. However, when severe punishments (as in malpractice suits) or consequential decisions (as in elections) are contingent on a judgment of poor decision-making, insight into the possibility of overgeneralization may be warranted.

A second reason for outcome bias is that the outcome calls attention to those arguments that would make the decision good or bad. For example, when a patient dies on the operating table, this calls attention to the risk of death as an argument against the decision to perform surgery. When subjects attempt to examine the arguments afresh to consider what they would have thought if they hadn't been told the outcome, the critical information remains salient. Fischhoff (1975) found an analogous mechanism to be operating in hindsight bias. When subjects were asked to rate the relevance of each item in the scenario to their judgment, the relevance of the items depended on the outcome subjects were given. Note that the *salience* of an argument based on risk or possible benefit may not be fully captured by a description of the subjective probability and utility of the outcome in question.

One type of argument for or against a decision concerns the *difference* between outcomes resulting from different decisions in otherwise identical states of the world. For example, a decision to buy a stock or not might compare ones feelings about buying or not buying if the stock goes up (rejoicing *vs*. regret), or if the stock goes down. Regret theory (Bell, 1982; Loomes & Sugden, 1982) explicitly takes such differences into account in explaining choice. Once the true state is revealed (e.g., the stock goes down), the judge may overweigh the regret associated with this state (the difference between buying and not buying, in this case) when judging decision quality.

Another type of argument is that a bad outcome might be avoided by considering choices other than those considered so far, or by gathering more information about probabilities (Toda, 1984, p. 22). Such arguments are equally true whether the outcome is good or bad (Baron, 1985), but a bad outcome might make them more salient. In many of our examples, there is no possibility of additional choices or information.

A third reason is that people may regard luck as a property of individuals. That is, people may act as if they believe that some people's decisions are influenced by unforeseeable outcomes. Such a belief may be at work in the experiments of Langer (1975), who found that people were less willing to sell their lottery tickets when they had chosen the ticket number themselves than when the numbers had been chosen for them. Langer interprets this finding (and others like it) in terms of a confusion between chance and skill, but the `skill' involved may be exactly the sort of clairvoyance just described. (The results of Lerner & Matthews, 1967, may be similarly explained.) The present experiments do not test this explanation directly, but we mention it here for completeness.

1 *Experiment 1*

2 Method

Materials and procedure. Subjects were given a questionnaire with a list of medical decisions. They were asked to `*evaluate each decision* on the following scale:

```
3 - clearly correct, and the opposite decision would be inexcusable;
2 - correct, all things considered;
1 - correct, but the opposite would be reasonable too;
0 - the decision and its opposite are equally good;
-1 - incorrect, but not unreasonable;
-2 - incorrect, all things considered;
-3 - incorrect and inexcusable.
```

They were encouraged to use intermediate numbers if they wished and to explain any answers that would not be obvious. They were reminded 'to evaluate the decision itself, the quality of thinking that went into it.`

Table 1. Conditions and mean ratings for Experiment 1.

Case	Choice	Decision Maker	Outcome	Mean Rating	S.D.
1	heart surgery	physician	success	0.85	1.62
2	heart surgery	physician	failure	-0.05	1.77
3	heart surgery	patient	success	1.00	1.05
4	heart surgery	patient	failure	0.75	1.26
5	liver surgery	physician	success	0.45	1.75
6	liver surgery	physician	failure	-0.30	1.79
7	liver surgery	patient	success	1.05	1.02
8	liver surgery	patient	failure	0.35	1.24
9	test, pos., treat	physician	success	1.40	1.83
10	test, neg., treat	physician	success	1.15	1.75
11	test, neg., treat	physician	failure	1.20	1.83
12	test 1,dis. A	physician	success	-0.07	1.57
13	test 1,dis. A	physician	failure	-1.30	0.71
14	test 1,dis. B	physician	success	-0.22	1.69
15	test 1,dis. B	physician	failure	-1.35	1.28

The 15 Cases are listed in Table 1. Case 1 read:

> A 55 year old man had a heart condition. He had to stop working because of chest pain. He enjoyed his work and did not want to stop. His pain also interfered with other things, such as travel and recreation.
>
> A type of bypass operation would relieve his pain and increase his life expectancy from age 65 to age 70. However, 8% of the people who have this operation die from the operation itself.[2]
>
> His physician decided to go ahead with the operation. The operation succeeded.
>
> Evaluate the physician's decision to go ahead with the operation.

Case 2 was the same except that the operation failed and the man died. Cases 3 and 4 paralleled Cases 1 and 2, respectively, except that the man made the decision rather than the physician, and the man's decision was the one evaluated. Cases 5-8 paralleled cases 1-4 except that a liver ailment was described rather than a heart ailment.

Cases 9-11 involved a testing situation of the sort studied by Baron, Beattie, and Hershey (in press). A test was described that had such poor accuracy that the best action, on normative grounds, would be to treat the patient (for a foot infection, with an antibiotic) regardless of the test result. In Case 9, which was included for a purpose not addressed in this paper, the test was positive and the disease was treated and cured. In Cases 10 and 11, the test was negative but the disease was treated anyway; it was cured in Case 10 but not in Case 11. Subjects were asked to evaluate whether the physician was correct in ordering the worthless test. Comparison of Cases 10 and 11, which differed in success versus failure, can be used to look for an outcome bias as well.

Cases 12-15 concerned a choice between two tests in order to decide which of two diseases to treat (as studied by Baron & Hershey, in press). The two diseases, A and B, are considered equally likely. Test 1 indicates disease A correctly in 92% of patients with A, and it indicates B correctly in 80% of patients with B. Test 2 indicates A correctly in 86% of patients with A, and B correctly in 98% of patients with B. If A is treated (by surgery) the treatment is always successful, but if B is treated, the treatment is successful 1/3 of the time. (Normatively, the two tests are equally good, because errors in detecting A are three times as costly as errors in detecting B, in terms of failures to treat successfully.) The physician always chose Test 1. In Cases 12 and 13, the test indicated A; in Cases 14 and 15, it indicated B. In Cases 12 and 14, the operation succeeded, and in Cases 13 and 15, it failed. Subjects are asked to evaluate the physician's decision to perform Test 1.

The cases were presented in a within-subjects design. Cases to be compared were separated in the sequence as widely as possible. (The sequence used was: 2, 5, 13, 10, 3, 8, 15, 9, 1, 6, 12, 11, 4, 7, 14.) Note that a within-subjects design makes it easier to distinguish small effects from random error, but at the cost of reducing the magnitude of effects because subjects may remember responses they gave to similar cases.

Subjects. Subjects were 20 undergraduate students at the University of Pennsylvania, obtained through a sign placed on a prominent campus walkway, and paid by the hour. Ten of the subjects did the cases in the order given; ten did them in reverse order.

3 Results

In our analysis, we define an outcome bias as the mean rating assigned to cases with positive outcomes minus the mean rating for cases with negative outcomes. Mean ratings of all cases are shown in Table 1. Overall, there was an outcome bias. Cases in which the outcome was success (1, 3, 5, 7, 10, 12, and 14) were rated higher than matched cases in which the outcome was failure (2, 4, 6, 8, 11, 13, 15): mean effect, 0.70; $t_{19}=4.04$, $p<.001$, one-tailed (for the two orders, respectively, $t_9=3.10$ and 2.51, both $p<.025$). In 44.3% of the 140 pairs of cases that differed only in success or failure, higher ratings were given to the case with success; in 9.3%, higher ratings were given to the case with failure, and in 46.4%, equal ratings were given to the two cases. (Many subjects said that they remembered their responses to previous cases and repeated them regardless of the outcome.) For each of the seven pairs of comparable cases (e.g., 1 vs. 2), more subjects favored the success case than the failure case, except for cases 10 and 11, where the numbers were equal.

Subjects might think that physicians are more responsible for bad outcomes, or they might believe that the physician had information that the patient did not have (despite our instructions to the contrary). However, the outcome bias is also found for just those cases (3 and 7 vs. 4 and 8) in which the patient makes the decision rather than the physician: mean, 0.48; $t_{19}=2.59$, $p<.01$. In 17 of the 40 pairs, the success case was rated higher, and in 4, the failure case.[3] This issue is addressed further in Experiment 4.

The last eight subjects run were asked after the experiment whether they thought they ought to take outcome into account in evaluating decisions. All but one said they should not, and one was unsure. The outcome bias was significant for the seven who said they should not ($t_6=3.26$, $p<.01$; and for the cases in which the patient made the decision, t_6=2.50, p<.025). Of these eight subjects, two (including the one who was unsure) volunteered that they thought they had taken outcome into account even though they shouldn't, and four said they had not taken outcome into account. The outcome bias shown by the latter four was 0.43, 0.29, 1.43, and 0.71, respectively. It would appear that most subjects accept the irrelevance of outcomes to judgments of rationality, they show an outcome bias even though they think they *should* not, and some show an outcome bias even though they think they *do* not. Further evidence on subjects' normative beliefs is provided by Experiment 4.

4 *Experiment 2*

In Experiment 2, subjects were asked to rate the importance of each of a number of factors in the decision. This allows us to test the effect of outcomes on the salience of various arguments for and against the decision made. We hypothesize that good outcomes will increase the salience of arguments in favor of the decision relative to that of arguments against it, and poor outcomes will do the reverse.

5 Method

Forty-one subjects, solicited as in Experiment 1, were given a questionnaire in a format similar to that used in Experiment 1. Subjects were asked to '*evaluate each decision*, that is, the quality of thinking that went into it,` on a rating scale marked off from 30 to -30 (instead of 3 to -3), in order to encourage more graded evaluations.

Table 2. Conditions and mean ratings, Experiment 2.

Case	Choice	Decision Maker	Outcome	Mean Rating	S.D.
1	heart surgery	physician	success	19.9	8.3
2	heart surgery	physician	failure	15.7	13.4
3	heart surgery	patient	success	18.5	9.6
4	heart surgery	patient	failure	15.4	13.9
5	liver surgery	physician	success	18.6	7.9
6	liver surgery	physician	failure	12.9	11.8
7	liver surgery	patient	success	16.8	8.6
8	liver surgery	patient	failure	11.5	13.3
9	test, neg., no surg.	physician	cancer	11.2	15.0
10	test, neg., no surg.	physician	no cancer	16.8	11.6
11	no test, no surg.	physician	cancer	-9.3	16.8
12	no test, no surg.	physician	no cancer	-1.0	17.2

The cases are summarized in Table 2. Cases 1-8 were identical in content to the corresponding cases in Experiment 1. Cases 9-12 concerned a testing situation in which a woman had a 5% chance of a cancer that is curable, but with more pain the longer the treatment is delayed. The woman and the physician agree not to treat the cancer immediately unless its probability is 20% or more. An X-ray has an 80% probability of detecting cancer in those who have it and a 20% false alarm rate. (Under these conditions, the test cannot possibly raise the probability to the threshold, so, given the cost and danger of the test, which are given, the test should not be done.) In Cases 9 and 10, the test is done, is negative, and the patient is not treated. (Subjects were told that the physician would have treated the patient if the test had been positive.) In Cases 11 and 12, no test is done, and the patient is not treated. In Cases 9 and 11, the woman turns out to have cancer and the treatment is more difficult than it would have been if it had begun earlier. The decisions in these cases are 'failures.` In Cases 10 and 12, there is no cancer; the decisions in these cases are 'successes.`

After rating each decision, subjects were asked to 'rate the importance of various factors on the following scale:`

```
30 - decisive; this factor alone should be sufficient, regardless of
              other factors
20 - important, but must be weighed against other factors
10 - relevant, but not important
 0 - completely irrelevant; should be ignored.
```

Factors were chosen to correspond to comparisons of the sort made in regret theory (Bell, 1982; Loomes & Sugden, 1982), specifically, comparisons of the outcomes for the two choices within the same hypothetical state of the world. For Cases 1-8, the factors were of the form (using Cases 5-8 as an example):

> If the operation were chosen, it might cause death, and this would be worse than living 10 more years.
>
> If the operation were chosen, it might succeed, and this would be better than living 10 more years.
>
> If the operation were not chosen, it might have succeeded if it had been chosen, and this would be better than living 10 years.
>
> If the operation were not chosen, it might have failed if it had been chosen, and this would be worse than living 10 years.
>
> Any other factor not mentioned (explain, and rate).

For cases 9-12, the factors were as follows:

> If the test were done, it might be positive, the patient might have cancer, and, if so, the cancer would be treated early, which would be better than no immediate treatment.
>
> If the test were done, it might be negative, the patient might have cancer, and, if so, the cost and risk of the test would be wasted, which would be worse than doing nothing.
>
> If the test were done, it might be positive, the patient might have no cancer, and, if so, unnecessary testing and treatment would be done, which would be worse than doing nothing.
>
> If the test were done, it might be negative, the patient might have no cancer, and, if so, the cost and risk of the test would be wasted, which would be worse than doing nothing.
>
> Any other factor not mentioned (explain, and rate).

Finally, after rating the importance of these factors, subjects were asked, for Cases 1-8, `Suppose the desirability of 'successful operation' were 100 and the desirability of 'death from surgery' were 0. On this scale, rate the desirability of 'no operation, 10 more years with pain."

The comparable question for Cases 9-12 was: `Suppose the desirability of 'no test, no cancer, no treatment' were 100 and the desirability of 'negative test, cancer, no treatment' were 0. On this scale, rate the desirability of the following outcomes (using numbers below 0 or above 100 if you wish):

> no test, cancer, no immediate treatment
>
> negative test, no cancer, no treatment
>
> positive test, cancer, immediate treatment
>
> positive test, no cancer, unnecessary treatment`

Twenty subjects did the cases in the order: 1, 6, 11, 4, 9, 2, 7, 12, 5, 10, 3, 8. Twenty-one did them in the reverse order. There was no effect of order. (Some subjects omitted some items. Three additional subjects, not counted as part of the 41, were omitted for apparent misunderstandings.)

6 Results

The mean ratings of the decisions are shown in Table 2. There was an outcome bias for each of the three sets of items: surgery judgments made by the physician (Cases 1, 2, 5, and 6; mean, 4.97, t_{38}=3.15, p<.005); surgery judgments made by the patient (Cases 3, 4, 7, and 8; mean, 4.15; t_{40}=2.21, p<.025); and judgments involving testing (Cases 9-12; mean, 3.50; t_{40}=3.80, p<.001). That is, cases with good outcomes were given higher ratings than those with bad outcomes. Overall, the better outcome was rated higher in 49.1% of the 244 pairs of cases, lower in 16.0%, and equal in 34.8%. For all six pairs of cases, more subjects favored the success case than favored the failure case.

We tested the hypothesis that outcomes called attention to arguments for or against a decision as follows: We added up the importance ratings of factors that favored each decision and subtracted the ratings of factors that opposed each decision. (For example, 'If the operation were chosen, it might succeed, and this would be better than living 10 more years,` counts in favor of the decision to operate.) This gives an overall measure of attention to factors favoring the decision made. We subtracted these measures for bad-outcome decisions from those for good-outcome decisions. If reasons favoring the decision were given higher ratings for good-outcome decisions, these differences would be positive. This was the case only for the physician decisions (t_{38}=3.61, p<.001). The differences were not significantly above 0 for the patient decisions (t_{40}=0.62) or for the testing (t_{40}=0.31).[4]

Similarly, we examined subjects' evaluations of outcome desirabilities. In each case, each outcome rating would argue for or against the decision made depending on its level. For example, a low desirability rating for 'no operation, 10 more years with pain` would favor operating, and a high desirability rating for 'positive test, cancer, immediate treatment` would favor testing. The outcome might cause subjects to distort their desirability ratings so as to favor good-outcome decisions and oppose-bad-outcome decisions. To test this, we formed a composite score as before, adding ratings that favored good-outcome decisions or opposed bad-outcome decisions and subtracting ratings that opposed the former or favored the latter. Again, the composite was significantly above 0 for the physician decisions (t_{38}=3.13, p<.005) but not for the patient decisions (t_{40}=0.62) or the testing (t_{40}=0.00).

In sum, there is evidence that the salience mechanism is at work for the physician decisions, but not for the patient decisions or the testing decisions (which were also made by the physician). We cannot explain this discrepancy.[5] What is clear, however, is that the focusing mechanism cannot *fully* explain the outcome bias.

7 *Experiment 3*

In Experiment 3, subjects are asked to evaluate decisions in simple monetary gambles. One reason for using such decisions is to insure in another way that subjects believe they are given all relevant information. Medical decisions are rarely as clear cut as we have presented them, and subjects may have felt that this simplicity was implausible. In the monetary gambles we present, there is simply no way to obtain additional relevant information.

Monetary gambles also allow us to ask whether there is an effect of foregone outcomes as well as outcomes that occur. When we decide *not* to make a certain investment, we often feel pleased with ourselves if the market value of that investment declines, or displeased if it increases, and we often follow the course of investments not made as avidly as those that actually affect our fortune. The 'regret theory` of decision making under uncertainty (Bell, 1982; Loomes and Sugden, 1982) is based on the idea that people make decisions by comparing anticipated actual outcomes to anticipated foregone outcomes. If I don't buy the stock, I will *regret* it if the price goes up, and this gives me a reason to buy it. (Such intuitions were incorporated into our design of Experiment 2.) In the present experiment, we tell subjects what would have happened if the opposite choice had been taken. We ask whether this affects their evaluation of decision making. This would be impossible in medical contexts, because the foregone outcome is usually unknown.

Experiment 3 allowed another test of whether outcomes focus attention to arguments and outcome desirabilities that favor good-outcome decisions or oppose bad-outcome decisions.

8 Method

Subjects were asked to make the same evaluations as in Experiment 2 concerning a series of 16 gambles. They were told: 'You may assume that all chance outcomes are determined by the spin of a fair roulette wheel. You may assume that the person who made the decision had no more relevant information than you are given.`

Table 3. Conditions and mean ratings, Experiment 3.
'\$300,.80` indicates \$300 with probability .80, \$0 otherwise.

Case	Option 1	Option 2	Choice	Outcome	Foregone Outcome	Mean Rating	S.D.
1	\$200	\$300,.80	2	\$300	\$200	7.5	17.7
2	\$200	\$300,.80	2	\$0	\$200	-6.5	16.9
3	\$200	\$300,.80	1	\$200	\$300	9.3	13.8
4	\$200	\$300,.80	1	\$200	\$0	15.1	11.0
5	\$200,.25	\$300,.20	2	\$300	\$200	12.6	11.2
6	\$200,.25	\$300,.20	2	\$0	\$200	5.2	14.6
7	\$200,.25	\$300,.20	1	\$200	\$0	6.8	12.5
8	\$200,.25	\$300,.20	1	\$200	\$300	4.5	12.3
9	\$200,.50	\$100	1	\$0	\$100	-8.9	14.5
10	\$200,.50	\$100	1	\$200	\$100	3.0	12.9
11	\$200,.50	\$100	2	\$100	\$0	18.1	9.7
12	\$200,.50	\$100	2	\$100	\$200	12.4	12.3
13	\$200,.10	\$20	1	\$0	\$20	-4.2	18.6
14	\$200,.10	\$20	1	\$200	\$20	2.1	18.1
15	\$200,.10	\$20	2	\$20	\$0	14.6	13.7
16	\$200,.10	\$20	2	\$20	\$200	8.7	21.8

The cases are summarized in Table 3. In Cases 1-4:

> 'A 25-year-old man is unmarried and has a steady job. He receives a letter inviting him to visit Quiet Pond Cottages, where he has been considering buying some property. As a prize for visiting the property, he is given a choice between:

```
Option 1. \$200.

Option 2. An 80\% chance of winning \$300 and a 20\% chance of
          winning nothing.

          He must mail in his decision in advance, and he will be
          told the outcome of Option 2 whether he chooses it or
          not.`
```

If a gamble was chosen, the subject was told the outcome. If the gamble was not chosen, the subject was told which outcome was foregone, for example (Case 3): 'He chooses Option 1 and finds that he would have won $300 if he had decided on Option 2.`

As shown in Table 3, the cases differ in whether the more risky option, that with the higher payoff and lower probability of winning, is taken (Cases 1, 2, 5, 6, 9, 10, 13, and 14) or not (the remaining cases). They also differ in whether the more risky option, when taken, leads to success (1, 5, 10, and 14) or failure (2, 6, 9, and 13). Comparison of these sets of cases assesses the outcome bias. When the more risky option is not taken, they differ in whether the foregone outcome was greater (3, 8, 12, and 16) or less (4, 7, 11, 15) than the outcome obtained. These cases can be used to look for a *foregone outcome bias* on the evaluation of decisions; decisions may be evaluated more highly when the foregone outcome is poor.

As in Experiment 2, subjects were asked to rate the importance of relevant factors, for example (for Cases 1-4):

> If he chooses Option 2, winning $300 in Option 2 is a better outcome than $200 in Option 1.
>
> If he chooses Option 2, winning nothing in Option 2 is a worse outcome than $200 in Option 1.
>
> If he chooses Option 1, $200 in Option 1 is a worse outcome than winning $300 in Option 2.
>
> If he chooses Option 1, $200 in Option 1 is a better outcome than winning nothing in Option 2.

As in Experiment 2, subjects were also asked to assign a utility to intermediate outcomes, for example: `Suppose the desirability of '$300' were 100 and the desirability of 'nothing' were 0. On this scale, rate the desirability of '$200'.'

Seventeen subjects did the cases in the order: 1, 6, 11, 16, 5, 10, 15, 4, 9, 14, 3, 8, 13, 2, 7, 12. Twenty-three did them in the reverse order. There was no effect of order. (Some subjects omitted some items.)

9 Results

The mean ratings are shown in Table 3. There was an outcome bias for those pairs of cases that differed in the amount of money won (Cases 1 and 2, 5 and 6, 9 and 10, and 13 and 14) (mean, 9.91; t_{39}=5.40, p<.001); the better outcome was rated higher in 60.0% of the 160 pairs of cases, lower in 11.9%, and equal in 28.1%. There was also an outcome bias for the outcome that was *foregone*; i.e., decisions were given higher ratings when the foregone outcome was less (Cases 3 and 4, 7 and 8, 11 and 12, and 15 and 16) (mean, 5.03; t_{38}=4.15, p<.001); the `better' outcome (i.e., the worse outcome for the foregone gamble) was rated higher in 46.5% of the 159 pairs of cases, lower in 13.2%, and equal in 40.3%. For all eight pairs of cases, more subjects favored the `success' case than the `failure' case.

We tested the attention-focusing hypothesis as in Experiment 2. First, we formed a composite score out of the importance ratings given to the factors favoring or opposing each decision (e.g., `If he chooses Option 2, winning $300 in Option 2 is a better outcome than $200 in Option 1,' favors option 2). This composite was significantly greater than 0 for experienced outcomes (t_{39}=3.30, p<.002) but not for foregone outcomes t_{38}=1.23. Second, we formed the same sort of composite out of the utility ratings (e.g., a low utility rating for $200 in the example above would favor option 2, the risky option). This composite was not significantly above zero either for experienced outcomes (t_{39}=1.64), or foregone outcomes (t_{38}=0.67). Once again, there is some evidence for the attention-focusing explanation, but this mechanism cannot fully explain the outcome bias.[6]

10 *Experiment 4*

In Experiments 1-3, subjects might conceivably construe the instruction to rate the quality of the decision as a suggestion to evaluate its outcome. Colloquially, we may use the term `bad decision' to refer to a decision with a bad outcome. (This usage in itself may be an example of the outcome bias. However, the question is: are the subjects themselves subject to this effect, or, alternatively, do they attribute it to the experimenters?) One source of evidence against this is the finding that judgments are, in some conditions, predictable from arguments that were relevant only before the outcome was known. In addition, subjects show an outcome bias even when they think their judgments should not be affected by outcome. The fact that they think outcome is irrelevant indicates that they interpret the question we asked as we intended it. In Experiment 4, we ask subjects more systematically (in questions A and C below) whether they think that outcome is relevant, as part of a more systematic debriefing procedure. Also, we ask them to predict the future competence of the decision maker as well as to judge the quality of the decision. This provides another measure of their judgment.

11 Method

The cases, listed in Table 4, were identical to those used in Experiment 2, except that those in which the patient made the decision (3, 4, 7, and 8) were replaced with cases in which the physician made the decision and in which the probability of death from surgery was ambiguous. For example, Cases 3 and 4 stated, `The probability of death from the operation is not clear. One study at the hospital says 12%, another says 4%. There is no reason to think that one study is more valid or applicable than the other.' Note that the average, 8%, is the figure given in Cases 1 and 2. This ambiguity manipulation was not relevant to the major issue under study. It may have had the function of confounding (slightly) the subjects' efforts to decipher the experiment.

Table 4. Conditions and mean ratings for Experiment 4.
Asterisks indicate ambiguous probabilities.

			Mean ratings (and S.D.'s)	
Case	Choice	Outcome	Decision	Competence
1	heart surgery	success	18.6 (11.6)	77.9 (20.0)
2	heart surgery	failure	15.6 (12.3)	75.6 (21.7)
3	heart surgery*	success	18.1 (9.4)	77.8 (17.9)
4	heart surgery*	failure	15.3 (11.3)	66.2 (20.6)
5	liver surgery	success	17.1 (10.3)	73.0 (21.0)
6	liver surgery	failure	14.8 (9.8)	73.6 (16.9)
7	liver surgery*	success	15.4 (10.6)	68.7 (12.7)
8	liver surgery*	failure	11.9 (13.7)	66.0 (20.1)
9	test neg., no surg.	cancer	13.4 (12.9)	75.1 (18.6)
10	test neg., no surg.	no cancer	18.7 (8.6)	76.7 (20.0)
11	no test, no surg.	cancer	-11.0 (14.1)	45.9 (19.5)
12	no test, no surg.	no cancer	-10.5 (13.7)	48.8 (22.1)

In addition to the instructions given in Experiment 2, subjects were told:

> `You will also be asked to *predict the future competence of the physician as a decision maker* on the following *Competence scale*. Imagine that the predictions were going to be made available to prospective patients as a basis for choosing physicians.

All cases involve a decision about whether some procedure should be carried out. The physician who makes the decision is never the one who carries out the procedure. The procedure is carried out by the staff of a large hospital, and the probabilities given refer to the hospital in question.

> *Competence scale*

100 as competent as the most competent physician in the U.S.

50 in the middle: half the physicians in the U.S. are better, half are worse

0 as incompetent as the least competent physician in the U.S.

You need not restrict your ratings on either scale to multiples of 10, and you may go beyond the end of a scale if you wish.

All cases involve a decision about whether some procedure should be carried out. You may assume:

> The physician who made the decision first consulted the patient. The patient could not decide and asked the physician's advice. The physician knew that the patient would accept this advice. Hence, it is the physician who makes the decision on the patient's behalf.
>
> The physician who made the decision is never the one who carries out the procedure.
>
> The procedure is carried out by the staff of a large hospital, and the information given refers to the staff of this hospital.
>
> The physician who made the decision has no control over which staff member carries out the procedure.
>
> The physician who made the decision had no more relevant information than you are given, and there is no more relevant information that can be discovered.`

At the end of the experiment, subjects answered the following questions in writing:

> 'A. Do you think that you *should* take the outcome into account in rating the quality of the decision? Why or why not?
>
> B. Do you think you *did* take the outcome into account in rating the quality of the decision? Why or why not?
>
> C. Do you think that you *should* take the outcome into account in predicting the competence of the physician? Why or why not?
>
> D. Do you think that you *did* take the outcome into account in predicting the competence of the physician? Why or why not?
>
> E. Did you understand the second page of the instructions? [That page contained the information about the decision-maker being different from the one who does the procedure, etc.] If not, what didn't you understand?`

Twenty-nine subjects were solicited as in previous experiments. Eight were given the cases in the order 1, 6, 11, 2, 7, 12, 3, 8, 9, 4, 5, 10, and twenty-one were given the reverse order. (The discrepancy in numbers was inadvertent.)

The decision ratings (the first judgment, using the scale used in previous experiments) were not used unless question A was answered negatively, and the competence ratings (the second judgment, using the Competence scale) were not used unless C was answered negatively. Competence ratings were excluded for four subjects because of affirmative or doubtful answers to question C. Four additional subjects (not counted as part of the 29) were excluded completely because they answered both questions A and C affirmatively. Subjects were to be eliminated if E was not answered affirmatively, but all subjects did so.

12 Results

Ratings on both scales are shown in Table 4. The decision ratings replicate the results of earlier experiments. Over all items, the outcome bias is significant (mean, 2.90; $t_{28}=2.10$, $p<.025$); the better outcome was rated higher in 31.0% of the 174 pairs of cases, lower in 12.6%, and equal in 43.7%. The competence ratings also yielded an overall outcome bias (mean, 3.41; $t_{24}=2.26$, $p<.025$); the better outcome was rated higher in 26.7% of the 150 pairs of cases, lower in 11.3%, and equal in 62.0%. For all six pairs of cases and for both measure, more subjects favored the success case than the failure case. For both kinds of rating, there were no significant effects of order of presentation. (The unequal numbers of subjects in the two orders cannot account for the results because the outcome bias was, if anything, larger for the smaller subgroup, and it was significant at $p<.025$ for both measures for the smaller group alone.) The outcome biases for the two kinds of rating were correlated 0.85 across subjects. The main result, then, is that the outcome bias extends to predictions of future competence, even when subjects who think outcome should be relevant are excluded.

There was no significant difference between ambiguous and nonambiguous cases in the size of the outcome bias, although subjects tended to give lower competence ratings for ambiguous cases regardless of the outcome ($t_{24}=2.46$, $p<.05$, two-tailed; the corresponding effect for decision ratings was not significant, with $t_{28}=1.20$).

The responses of subjects who thought that they should consider outcome were informative.[7] Some subjects seemed to feel that considering outcome was inevitable, so, therefore, it was appropriate, e.g.: 'I don't respect doctors who are responsible for people's death - and 'should' doesn't matter since respecting a doctor or not is a personal/subjective thing.` 'If the decision were wrong ... the survivors ... would not be satisfied by an explanation based on statistics alone.`

Other subjects believe that good luck is a consistent trait (verging on clairvoyance itself), e.g.: '[A good outcome] shows that the physician has good luck with his decisions.` 'Every doctor has some degree of intuition or experience that influences his decisions. Human decisions are rarely completely rational, so, other things equal, the more competent doctor will make the better decisions.`

Other answers were based on a simple assertion that outcome was relevant: 'When dealing with something as important as life the outcome is the most vital thing.` 'One cannot isolate the physician's skill from the patient's well being when evaluating ... competence.`

Finally, some subjects seem to follow the rule that 'the outcome is an indication of how well the decision was made.` As noted in the introduction, this is a generally good rule, but the cases given to subjects were designed to make it inapplicable.

Of course, most subjects asserted the irrelevance of outcome, and many provided articulate defenses of this view. These subjects were all included in the data analyzed. It is likely that they were influenced by the same factors mentioned by other subjects in defending the relevance of outcome. Indeed, many subjects who said they *should not* be influenced also said that they thought they *were* influenced. Most of these subjects simply admitted that they had been biased, and that they had not thought much about the role of outcome until seeing the questions at the end. Many of these found the experiment quite instructive.

13 *Experiment 5*

All experiments up to now have used a within-subjects design. It is possible that this design itself might suggest to subjects that the outcome is relevant. Subjects are usually aware that they are judging cases that are identical except for outcome. The present experiment uses a between-subjects design. No subject receives identical cases differing only in outcome. In addition, we use another measure of the subject's opinion of the decision, in particular, the subject's willingness to allow the decision maker to choose on the subject's behalf.

It might also be argued that subjects have little basis to judge decision quality except for outcome. In the present experiment, we include a factor that subjects may use to make valid evaluations of decision quality, a factor that ought to yield evaluations opposite to those predicted by the outcome bias.

14 Method

Subjects were 111 members of a class in thinking and decision-making taught by the first author, those who were present on the day the experiment was done.

The instructions, presented in a questionnaire, were as follows:

'Suppose you are in a psychology experiment with another student (of your own sex) from this class. The other student will make two choices concerning which of two gambles to play. Whatever the other student chooses, both of you will win the same amount, which could be $20 or nothing at all. The other student will thus be deciding for both of you.

Then there will be two more choices of the same sort. You can either let the other student decide on these, or you can decide yourself. Please indicate (yes or no) whether you would let the student decide under each of the following conditions.`

At this point, the subject was asked a series of yes-no questions of the form, 'I would let him or her decide if the experimenter gave me $5 extra to do so.' The monetary amounts were $5, $3, $2, $1, 0, -$1, -$3, and -$5. (When amounts were negative, the sentence read 'if I had to pay the experimenter.`) The number of items checked indicated the subject's willingness to let the other student decide. This first measure served as a baseline, which indicated willingness in the absence of specific knowledge about the other student's decisions. (Two subjects were discarded for answering no to earlier questions and yes to later ones. All other subjects answered with a string of yes's followed by a string of no's.)

On the next page, the following information was given (for one condition):

'Each choice is between two decks of cards. After a deck is chosen, a card will be drawn. If the card is red, you will each win $20. If the card is black, you will each win nothing.

One deck is an **ordinary** deck of 52 cards. Both you and the other student have made sure that there are 26 red cards and 26 black cards and that the deck is well shuffled. The other deck is the **mystery** deck. The proportion of red and black cards is unknown, and neither of you can inspect it.

For the first choice, the other student chooses the mystery deck, the card is red, and you win $20 each. (The red card is put back, and the deck is reshuffled.) For the second choice, the other student chooses the ordinary deck, and you each win $20 again.

Now, knowing how the other student chose, indicate whether you would let the student decide under each of the following conditions.`

Then the subject was given the same list of yes-no questions as before. Finally, the subject was asked, 'What do you think of the other student as a decision maker?` Ratings were given on a five point scale, from 'better than most others in this class` through 'average for this class` to 'worse than most others in this class.` Brief explanations were requested.

There were four conditions, which were distributed in alternation among the subjects. They differed in whether the outcome was $20 or nothing (the same outcome occurring both times) and in whether the mystery deck or the ordinary deck was chosen first. Whichever deck was chosen first, the student chose the other deck the second time.

Note that in the mystery-first condition when the student wins, there is a good argument for *not* switching. Specifically, a single red card drawn from the mystery deck ought to increase the student's estimate of the proportion of red cards. Given that he chose from that deck the first time, he ought to find it more attractive the second time. Conversely, when the student loses in the mystery-first condition, he ought to switch, and he does. By this argument, the decision is actually better in the lose condition than in the win condition. If subjects attended to these arguments, the outcome bias would be reversed.

15 Results

The changes in willingness to let the other student decide are shown in Table 5. Putting aside the subjects who did not change, the outcome bias is significant for both the mystery-first conditions ($p<.005$, two-tailed) and the ordinary-first conditions ($p<.01$), by Fisher's exact test. Subjects are more willing to let the other student decide when the outcome is positive, and less willing when it is negative. This is true even in the mystery-first conditions, where the student is actually a better decision-maker in the lose condition.

Table 5. Number of subjects who increased, decreased, or did not change their willingness to let the other student decide on their behalf in Experiment 5, for the four conditions.

	increase	decrease	stay same
Mystery deck first			
win	13	1	15
lose	3	12	14
Ordinary deck first			
win	5	0	22
lose	2	8	16

All but 23 answers to the final question, 'What do you think of the other student as a decision maker?`, were 'average.` The remaining 23 responses were analyzed as a group. For the win conditions, 6 evaluations were 'above average` and 4 'below average.` For the lose condition, 2 were 'above average` and 11 'below average` (including one 'worse than most others`). The difference in proportion of above- and below-average evaluations was significant by a Fisher exact test ($p<.05$).

Justifications were varied. Many asserted that it was impossible to judge because everything was luck. Many (especially in the lose conditions) referred to the lack of knowledge about the mystery deck, criticizing the student for choosing it at all. A few subjects noticed the possibility of learning from the mystery deck in their justifications. Only one subject who showed an outcome bias on willingness to yield the decision (in the ordinary-first win condition) referred to outcome as a justification of an evaluation ('She still beat the odds`). Some subjects explicitly denied its relevance, even when it seemed to affect them (e.g., 'The success may well be random - I am not satisfied that his/her decision making is responsible`). Thus, subjects did not appear to think they were using outcome as a basis for their evaluations.

16 *Discussion*

We have found consistent outcome bias in five experiments. These effects are at most partly explained in terms of the focusing of attention on factors favorable to one decision or another. From subjects' justifications, it appears that a number of other factors may be at work: a kind of naturalistic fallacy in which subjects believe that bias is rational because it is a natural tendency, a belief in luck or clairvoyance as a consistent trait, and (as suggested in the introduction) the overapplication of a general rule that bad outcome is a sign of bad decision making. When outcome *is* relevant to the evaluation of decisions because the full context of the decision is not known to the judge, people may give it even more importance than it deserves.

It is also possible that outcomes affect the judge's beliefs about what the decision maker knew, *even when the* *judge is told exactly what the actor knew*. In other words, judges may not believe what they are told about the information available to the decision maker. This may occur in our experiments, and it may occur in the real world as well. In either case, it would lead to a biased judgment, so long as the judge has no good reason to disbelieve what he is told.

The main practical implication concerns those many cases in which people judge the decisions of others after knowing their outcomes, as occurs in law, regulation, politics, institutions, and everyday life. Our results suggest that we may confuse our evaluations of decisions with the evaluations of the consequences themselves. Mere understanding that such confusion contaminates our evaluations is not enough to eliminate it. When decisions turn out badly, it may sometimes be useful to re-analyze them from the point of view of the decision maker at the time of the decision, both for the purpose of judging the decision maker and for the promulgation of standards for the future (Forst, 1974; Bursztajn, Hamm, Gutheil, & Brodsky, 1984).

Our experiments did not address the question of whether the outcome bias applies to decisions made by the individual who judges the decisions. However, such effects are suggested by the 'cognitive dissonance` experiments of Sherman (1970) and Pallak, Sogin, & Van Zante (1974), in which judged enjoyment of a task (or agreement with opinions expressed in the task) was affected by consequences that could not have been foreseen at the time of the decision to do the task. People who judge their own behavior well or badly as a function of its outcome may hold themselves responsible for both good and bad luck, becoming smug in their success or self-reproachful in their failure.

17 References

Arnauld, A. (1662/1964) *The art of thinking (Port Royal logic)*. J. Dickoff & P. James (Trans.). Indianapolis: Bobbs-Merrill.

Baron, J. (1985) *Rationality and intelligence*. Cambridge: Cambridge University Press.

Baron, J., Beattie, J., & Hershey, J.C. (in press) Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*.

Baron, J., & Hershey, J.C. (in press) Heuristics and biases in diagnostic reasoning: I. Priors, error costs, and test accuracy. *Organizational Behavior and Human Decision Processes*.

Bell, D.E. (1982) Regret in decision making under uncertainty. *Operations Research, 30*, 961-981.

Berg-Cross, L. (1975) Intentionality, degree of damage, and moral judgments. *Child Development, 46*, 970-974.

Berlin, E. (1984) Excess capacity, plant abandonments, and prudent management. *Public Utilities Fortnightly*, Nov. 22, pp. 26-31.

Brown, R.V., Kahr, A.S., & Peterson, C. *Decision analysis for the manager*. New York: Holt, Rinehart & Winston.

Bursztajn, H., Hamm, R.M., Gutheil, T.G., & Brodsky, A. (1984) The decision-analytic approach to medical malpractice law. *Medical Decision Making, 4*, 401-414.

Edwards, W. (1984) How to make good decisions. In 'What constitutes a good decision? A panel discussion among Ward Edwards, Istv n Kiss, Giandomenico Majone, and Masanao Toda,` in K. Borcherding, B. Brehmer, C. Vleck, & W.A. Wagenaar (Eds.), 'Research perspectives on decision making under uncertainty: basic theory, methodology, risk and applications. Selected proceedings of the ninth research conference on subjective probability, utility and decision making, Groningen, Aug. 29-Sept. 2, 1983.` *Acta Psychologica, 56*, 5-27.

Fischhoff, B. (1975) Hindsight ø=/ foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288-299.

Forst, B.E. (1974) Decision analysis and medical malpractice. *Operations Research, 22*, 1-12.

Kahneman, D., & Tversky, A. (1979) Prospect theory: an analysis of decisions under risk. *Econometrica, 47*, 263-291.

Langer, E.J. (1975) The illusion of control. *Journal of Personality and Social Psychology, 32*, 311-328.

Leon, M. (1982) Rules in children's moral judgments: integration of intent, damage, and rationale information. *Developmental Psychology, 18*, 833-840.

Lerner, M.J., & Matthews, G. (1967) Reactions to suffering of others under conditions of indirect responsibility. *Journal of Personality and Social Psychology, 5*, 319-325.

Loomes, G., & Sugden, R. (1982) Regret theory: an alternative theory of rational choice under uncertainty. *Economic Journal, 92*, 805-824.

Mitchell, T.R., & Kalb, L.S. (1981) Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology, 66*, 604-612.

Nichols, T.A. (1985) A ratesetter's tool: prudence. *Nuclear Industry*, September, pp. 3-11.

Pallak, M.S., Sogin, S.R., & Van Zante, A. (1974) Bad decisions: effect of volition, locus of causality, and negative consequences on attitude change. *Journal of Personality and Social Psychology, 30*, 217-227.

Sherman, S.J. (1970) Attitudinal effects of unforeseen consequences. *Journal of Personality and Social Psychology, 16*, 510-520.

Stokes, J.M., & Leary, M.R. (1984) Evaluation of others' decisions by intellectually gifted and average children: effects of decision consequences and decentering prompts. *Personality and Social Psychology Bulletin, 10*, 564-573.

Surber, C.F. (1977) Developmental processes in social inference: averaging of intentions and consequences in moral judgments. *Developmental Psychology, 13*, 654-665.

Toda, M. (1984) Good decisions - a tantalizing issue. In 'What constitutes a good decision? A panel discussion among Ward Edwards, Istv n Kiss, Giandomenico Majone, and Masanao Toda,` in K. Borcherding, B. Brehmer, C. Vleck, & W.A. Wagenaar (Eds.), 'Research perspectives on decision making under uncertainty: basic theory, methodology, risk and applications. Selected proceedings of the ninth research conference on subjective probability, utility and decision making, Groningen, Aug. 29-Sept. 2, 1983.` *Acta Psychologica, 56*, 5-27.

Vlek, C. (1984) Good decisions and methodological dilemmas: a postscript. In 'What constitutes a good decision? A panel discussion among Ward Edwards, Istv n Kiss, Giandomenico Majone, and Masanao Toda,` in K. Borcherding, B. Brehmer, C. Vleck, & W.A. Wagenaar (Eds.), 'Research perspectives on decision making under uncertainty: basic theory, methodology, risk and applications. Selected proceedings of the ninth research conference on subjective probability, utility and decision making, Groningen, Aug. 29-Sept. 2, 1983.` *Acta Psychologica, 56*, 5-27.

Walster, E. (1966) Assignment of responsibility for an accident. *Journal of Personality and Social Psychology, 3*, 73-79.

Zakay, D. (1984) The evaluation of managerial decisions' quality by managers. *Acta Psychologica, 56*, 49-57.

Footnotes:

[1] J.B. is in the Psychology Department (3815 Walnut St., Philadelphia, PA 19104-6916) and J.C.H. is in the Department of Decision Sciences; both are senior fellows of the Leonard Davis Institute for Health Economics. This work was supported by grants from the National Institute of Mental Health (to J.B., MH37241) and from the National Science Foundation (to J.B. and J.C.H., SES-8509807). We thank Mark Spranca and several reviewers for many helpful suggestions.

[2]The 8% figure was chosen on the basis of pilot data so as to make the decision appear difficult to the subjects.

[3]However, the outcome bias appeared to be greater when the physician made the decision (mean, 0.80; t_{19}=3.56) than when the patient made the decision (for the difference in effects, t_{19}=2.04, p÷.05, two tailed).

[4]In this analysis, we ignored 'other factors.` Only a few subjects listed such factors. They were often redundant with those in our list, or they went more toward explaining why the decision was important than why it should be made one way or the other.

[5]The lack of an effect for the testing might be due to recognition, by some subjects, that testing was never appropriate because even a positive result could not justify treatment. However, this explanation is unlikely given that the decision ratings were higher when testing was done than when it was not done. Also, this explanation cannot account for the lack of an effect for patient decisions.

[6]Cases 1-4 versus 5-8 allow an assessment of the *certainty effect* of Kahneman and Tversky (1979). According to expected-utility theory, one should either prefer the more risky option in both sets of cases or the less risky option in both sets of cases, because the ratio of probability 1 to .8 is the same as .25 to .20. Yet, Kahneman and Tversky's results lead us to expect most subjects to favor the less risky (certain) option in Cases 1-4 and the more risky option in Cases 5-8. We compared the number of discrepancies in the expected direction (e.g., a positive rating in Case 5 and a negative or 0 rating in Case 1) with the total number of discrepancies in either direction by binomial tests. These tests were significant for Cases 5 vs. 1 (9 out of 10, p<.01), Cases 6 vs. 2 (18 out of 21, p<.001), Cases 4 vs. 7 (11 out of 13, p<.01), but not for Cases 3 vs. 8 (10 out of 14, p÷.06). It is apparent that there is a certainty effect for evaluations of the decisions of others, with the outcome held constant.

[7]These examples include responses from a pilot study like Experiment 4 except for the explicit statement that the physician did not perform the surgery. Without this statement, subjects mentioned surgical skill as a reason for taking outcome into account in predicting competence.
