Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis

James E. Smith, Robert L. Winkler,

MANAGEMENT SCIENCE
Vol. 52, No. 3, March 2006, pp. 311–322
ISSN 0025-1909 | EISSN 1526-5501 | 06 | 5203 | 0311

DOI 10.1287/mnsc.1050.0451

The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis

James E. Smith, Robert L. Winkler
Fuqua School of Business, Duke University, Durham, North Carolina 27708
{jes9@duke.edu, rwinkler@duke.edu}

Decision analysis produces measures of value such as expected net present values or expected utilities and ranks alternatives by these value estimates. Other optimization-based processes operate in a similar manner. With uncertainty and limited resources, an analysis is never perfect, so these value estimates are subject to error. We show that if we take these value estimates at face value and select accordingly, we should expect the value of the chosen alternative to be less than its estimate, even if the value estimates are unbiased. Thus, when comparing actual outcomes to value estimates, we should expect to be disappointed on average, not because of any inherent bias in the estimates themselves, but because of the optimization-based selection process. We call this phenomenon the optimizer's curse and argue that it is not well understood or appreciated in the decision analysis and management science communities. This curse may be a factor in creating skepticism in decision makers who review the results of an analysis. In this paper, we study the optimizer's curse and show that the resulting expected disappointment may be substantial. We then propose the use of Bayesian methods to adjust value estimates. These Bayesian methods can be viewed as disciplined skepticism and provide a method for avoiding this postdecision disappointment.

Key words: decision analysis; optimization; optimizer's curse; Bayesian models; postdecision surprise; disappointment
History: Accepted by David E. Bell, decision analysis; received January 17, 2005. This paper was with the authors 1 month for 1 revision.

The best laid schemes o' Mice an' Men,
Gang aft agley,
An' lea'e us nought but grief an' pain,
For promis'd joy!

—Robert Burns "To a Mouse: On turning her up in her nest, with the plough" 1785

1. Introduction

A team of decision analysts has just presented the results of a complex analysis to the executive responsible for making the decision. The analysts recommend making an innovative investment and claim that, although the investment is not without risks, it has a large positive expected net present value. The executive is inclined to follow the team's recommendation, but she recalls being somewhat disappointed after following such recommendations in the past. While the analysis seems fair and unbiased, she can't help but feel a bit skeptical. Is her skepticism justified?

In decision analysis applications, we typically identify a set of feasible alternatives, calculate the expected value or certainty equivalent of each alternative, and then choose or recommend choosing the alternative with the highest expected value. In this paper, we examine some of the implications of the fact that the values we calculate are estimates that are subject to random error. We show that, even if the value estimates are unbiased, the uncertainty in these estimates coupled with the optimization-based selection process leads the value estimates for the recommended action to be biased high. We call this bias the "optimizer's curse" and argue that this bias affects many claims of value added in decision analysis and in other optimization-based decision-making procedures. This curse is analogous to the winner's curse (Capen et al. 1971, Thaler 1992), but potentially affects all kinds of intelligent decision making—attempts to optimize based on imperfect estimates—not just competitive bidding problems.

We describe the optimizer's curse in §2 of this paper and show how it affects claims of value added by decision analysis. This phenomenon has been noted (but not named) in the finance literature by Brown (1974) and in the management literature by Harrison and March (1984), who label it postdecision surprise. However, it seems to be little known and underappreciated in the decision analysis and broader management sciences communities. In §3, we consider the question of what to do about the optimizer's curse. There we propose the use of standard Bayesian methods for modeling value estimates, showing that these methods can correct for the curse and, in so doing, may affect the recommendations derived from the analysis. The prescription calls for treating the decision-analysis-based value estimates like the noisy estimates that they are and mixing them with prior estimates of value, in essence treating the decision-analysis-based value estimates somewhat skeptically. In §4, we discuss related biases, including the winner's curse, and in §5, we offer some concluding comments.

2. The Optimizer's Curse

Suppose that a decision maker is considering n alternatives whose true values are denoted $\mu_1, \ldots, \mu_n$. We can think of these "true values" as representing the expected value or expected utility (whichever is the focus of the analysis) that would be found if we had unlimited resources—time, money, computational capabilities—at our disposal to conduct the analysis. A decision analysis study produces estimates $V_1, \ldots, V_n$ of the values of these alternatives. These estimates might be the result of, say, a \$50,000 consulting effort, whereas it might cost millions to calculate the true value to the decision maker.[1]

The standard decision analysis process ranks alternatives by their value estimates and recommends choosing the alternative i^* that has the highest estimated value V_{i^*}. Under uncertainty, the true value μ_{i^*} of a recommended alternative is typically never revealed. We can, however, view the realized value x_{i^*} as a random draw from a distribution with expected value μ_{i^*} and, following Harrison and March (1984), think of the difference $x_{i^*} - V_{i^*}$ between the realized value and value estimate as the postdecision surprise experienced by the decision maker. A positive surprise represents some degree of elation and a negative surprise represents disappointment. Averaging across many decisions, the average postdecision surprise $x_{i^*} - V_{i^*}$ will tend toward the average expected surprise, $\mathrm{E}[\mu_{i^*} - V_{i^*}]$.

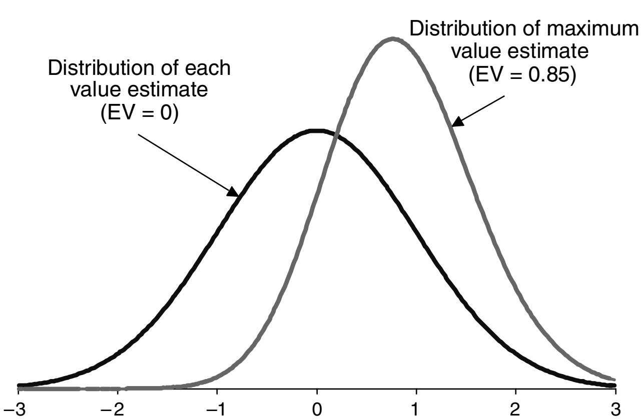
If the value estimates produced by a decision analysis are conditionally unbiased in that $\mathrm{E}[V_i \mid \mu_1, \ldots, \mu_n] = \mu_i$ for all i, then the estimated value of any alternative should lead to zero expected surprise, i.e., $\mathrm{E}[\mu_i - V_i] = 0$. However, if we consistently choose the alternative with highest estimated value, this selection process leads to a negative expected surprise, even if the value estimates are conditionally unbiased. Thus, a decision maker who consistently chooses alternatives based on her estimated values should expect to be disappointed on average, even if the individual value estimates are conditionally unbiased. We formalize this optimizer's curse in §2.3 after illustrating it with some examples.

2.1. Some Prototypical Examples

To illustrate the optimizer's curse in a simple setting, suppose that we evaluate three alternatives that all have true values (μ_i) of exactly zero. The value of each alternative is estimated and the estimates V_i are independent and normally distributed with mean equal to the true value of zero (they are conditionally unbiased) and a standard deviation of one. Selecting the highest value estimate then amounts to selecting the maximum of three draws from a standard normal distribution. The distribution of this maximal value estimate is easily determined by simulation or using results from order statistics and is displayed in Figure 1. The mean of this distribution is 0.85, so in this case, the expected disappointment, $\mathrm{E}[V_{i^*} - \mu_{i^*}]$, is 0.85. Because the results of this example are scale and location invariant, we can conclude that given three alternatives with identical true values and independent, identical, and normally distributed unbiased value estimates, the expected disappointment will be 85% of the standard deviation of the value estimates.

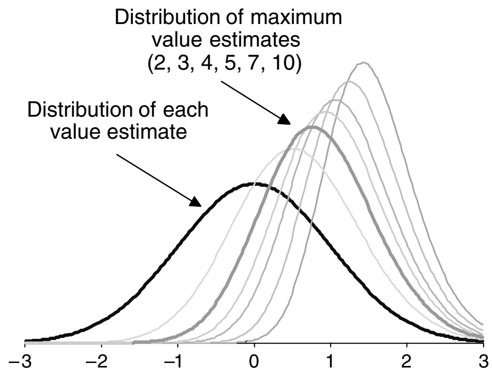
This expected disappointment increases with the number of alternatives considered. Continuing with the same distribution assumptions and varying the number of alternatives considered, we find the results shown in Figure 2. Here, we see that the distributions shift to the right as we increase the number of alternatives and the means increase at a diminishing rate. With four alternatives, the expected disappointment reaches 103% of the standard deviation of the value estimates, and with 10 it reaches 154% of the standard deviation of the value estimates.

Figure 1 **The Distribution of the Maximum of Three Standard Normal Value Estimates**



[1] Our use of "true values" is in the spirit of Matheson (1968), who refers to probabilities or values "given by a complete analysis." Tani (1978) objects to the use of "true" in this context, noting that this value is subjective and depends on the decision maker's state of information; he refers to "authentic probabilities" rather than "true probabilities." These concerns notwithstanding, the use of the term "true values" in this setting seems both natural and standard, having been used by Lindley et al. (1979) and Lindley (1986), among others.

Figure 2 **The Distribution of the Maximum of *n* Standard Normal Value Estimates**



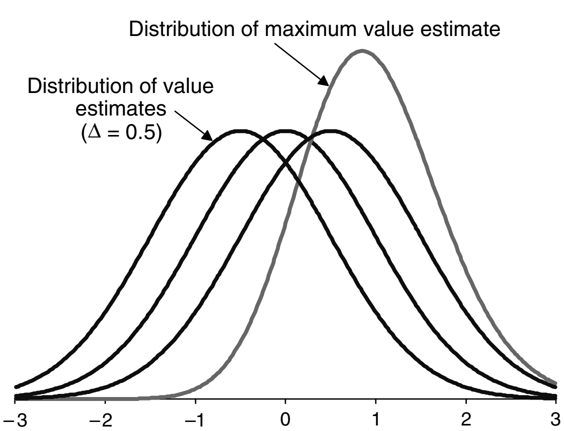
Number of alternatives	Expected disappointment
1	0.00
2	0.56
3	0.85
4	1.03
5	1.16
6	1.27
7	1.35
8	1.43
9	1.48
10	1.54

The case where the true values are equal is, in a sense, the worst possible case because the alternatives cannot be distinguished even with perfect value estimates. Figure 3 shows the results in the case of three alternatives, where the true values are separated by Δ: $\mu_i = -\Delta, 0, +\Delta$. The value estimates are again assumed to be unbiased with a standard deviation of one. In Figure 3, we see that the magnitude of the expected disappointment decreases with increasing separation. Intuitively, the greater the separation between the alternative that is truly the best and the other alternatives, the more likely it is that we will select the correct alternative. If we always select the truly optimal alternative, then the expected disappointment would be zero because its value estimate is unbiased.

We have assumed that the value estimates are independent in the above examples. In practice, however, the value estimates may be correlated, as the value estimates for different alternatives may share common elements. For example, in a study of different strategies to develop an R&D project, the value estimates may all share a common probability for technical success; errors in the estimate of this probability would have an impact on the values of all of the alternatives considered. Similarly, a study of alternative ways to develop an oil field may share a common estimate (or probability distribution) of the amount of oil in place. In practice then, we might expect value estimates to be positively correlated.

To illustrate the impact of correlation in value estimates, consider a simple discrete example with two alternatives that have equal true values and value estimates that are equally likely to be low or high by some fixed amount. This setup is illustrated in Table 1. If the two value estimates are independent, there is a 75% chance that we will observe a high value estimate for at least one alternative and overestimate the true value of the optimal alternative and a 25% chance of underestimating the true value; the value estimate of the selected alternative will thus overestimate the true value on average. If the two value estimates are perfectly positively correlated, there is a 50% chance of both estimates being high and a 50% chance of both being low, and we would have an estimate for the selected alternative that is equal to the true value on average. Thus, we should expect positive correlation to decrease the magnitude of the expected disappointment. Negative correlation, on the other hand, should increase expected

Figure 3 **The Distribution of Maximum Value Estimates with Separation Between Alternatives**



Δ	Expected disappointment	Probability of correct choice
0.0	0.85	0.33
0.2	0.66	0.42
0.4	0.51	0.50
0.6	0.39	0.59
0.8	0.30	0.66
1.0	0.22	0.73
1.2	0.17	0.78
1.4	0.12	0.83
1.6	0.10	0.87
1.8	0.07	0.90
2.0	0.05	0.92
2.2	0.03	0.94
2.4	0.02	0.95
2.6	0.01	0.97
2.8	0.01	0.98
3.0	0.00	0.98

Table 1 A Simple Discrete Example with Dependence

	Project 1 high	Project 1 low
Project 2 high	HH	LH
Project 2 low	HL	LL

disappointment, but negative correlation is less likely to hold in practice.

While the previous examples have assumed that the true values are fixed, the true values will be uncertain in practice. Just as we might expect value estimates to be positively correlated, we might expect the true values to be positively correlated for the same reasons. For example, uncertainty about a probability of technical success may lead to the true values for alternatives that depend on this probability being positively correlated. While positive correlation among the value estimates decreases expected disappointment, positive correlation among the μ_is tends to decrease the separation among the true values, which, as discussed earlier, increases expected disappointment. Table 2 shows how the expected disappointment varies with correlation in a setting where there are four alternatives and the true values have standard normal distributions with a common pairwise correlation that varies across rows in the table. The value estimates have a mean equal to the true mean (and are thus conditionally unbiased), a standard deviation of one, and a common pairwise correlation that varies across the columns in the table. In the no-correlation case, the expected disappointment is 73% of the common standard deviation of the value estimates and true values.[2] As expected, increasing the correlation among the V_is decreases the expected disappointment; increasing the correlation among the μ_is has the opposite effect. Moving along the diagonal in Table 2, we see that increasing both correlations simultaneously leads to a net decrease in expected disappointment. Even with modestly high degrees of correlation, say, with both correlations at 0.5 or 0.75, the expected disappointment remains substantial at 52% or 36% of the standard deviation of the value estimates and true values.

2.2. Claims of Value Added in Decision Analysis

In decision analysis practice, it is common to calculate and report the "value added" by an analysis. Value added is typically defined as the difference between the estimated value of the optimal alternative identified in the analysis and the estimated value of a default alternative (or current plan) that would have been chosen if no analysis were done. Even though the estimated value of each alternative may be unbiased, the value of the optimal alternative will be affected by the optimizer's curse and such claims of value added will also be affected.

Table 2 Expected Disappointment as a Function of Correlation in Value Estimates and in True Values

	Correlation among value estimates (V_is)				
	0.00	0.25	0.50	0.75	0.90
Correlation among true values (μ_is)					
0.00	0.73	0.59	0.41	0.22	0.09
0.25	0.78	0.64	0.45	0.25	0.12
0.50	0.84	0.69	0.52	0.29	0.12
0.75	0.92	0.77	0.58	0.36	0.18
0.90	0.98	0.84	0.67	0.43	0.23

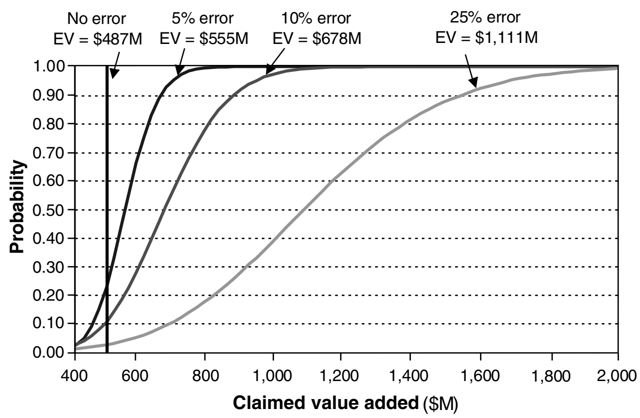
Clemen and Kwit (2001) considered 38 well-documented studies at Eastman Kodak from 1990–1999 and estimated the total value added by decision analysis at Kodak in this period. Lacking a preidentified default alternative in these analyses, they focused on an alternative measure of value added as the difference between the value (defined as the expected net present value) of the optimal alternative and the average estimated value for all of the alternatives considered in the study. This measure of value added will be nonnegative by definition and will also be affected by the optimizer's curse. The total value added in these 38 studies, using Clemen and Kwit's (2001) measure of value added, is $487 million.[3]

We can use Clemen and Kwit's (2001) data to perform some simulations to give a sense of the magnitude of the expected disappointment that might be experienced in practice. In our simulations, we take the Kodak value estimates to be true values and generate sample value estimates from these true data. Specifically, we take the true value (μ_i) for each alternative to be the value given by Kodak's decision analysis study and generate value estimates (V_i) for each alternative that are independent and normally distributed with mean μ_i and standard deviation equal to 5%, 10%, or 25% of the absolute value of μ_i. (We consider a correlated case below.) For each study, we then select the alternative with the maximum value and calculate the "claimed value added" as the difference between this maximum value and the average of the value estimates for all alternatives considered

[2] This is less than the expected disappointment in Figure 2 because the results in Figure 2 assume that all of the true values are identical. Here, the true values are uncertain and may be separated, thereby decreasing the expected disappointment.

[3] Clemen and Kwit (2001) also consider two other measures of value added, which are similarly affected by the optimizer's curse but will not be discussed here. They also extrapolate from this sample of 38 well-documented studies to estimate a total value added of approximately $1 billion for all studies done in this time frame.

Figure 4 Simulation Results Based on Clemen and Kwit's (2001) Kodak Data



in that study. With these assumptions, the true value added is the $487 million reported by Clemen and Kwit (2001).

The results of this simulation exercise are summarized in the cumulative probability distributions of Figure 4. With standard deviations of 5% in the value estimate, value added is overestimated approximately 75% of the time and the average claimed value added is $555 million, which overstates the true value added ($487 million) by 14%. With standard deviations of 10% and 25%, the value added is overestimated even more frequently and the average claimed value added jumps to $678 million and $1.111 billion, overstating the true value added by 39% and 128%, respectively. Thus, the optimizer's curse can have a substantial impact on estimates of value added.

To illustrate the effects of dependence, we also ran a simulation where the standard deviation of V_i is 10% of the absolute value of μ_i and the value estimates for each study were correlated with pairwise correlation coefficients equal to 0.5. (The value estimates are correlated within each study but are still assumed to be independent across studies.) The total claimed value in this case is $602 million (compared to $678 million in the independent case), overstating the true value added by 24% (compared to 39% in the independent case). Thus, positive dependence reduces the magnitude of the effect of the optimizer's curse, but it remains considerable.

2.3. Formalizing the Optimizer's Curse

Having illustrated the curse with some examples, we now formally state the result.

Proposition 1. *Let $V_1, \ldots, V_n$ be estimates of $\mu_1, \ldots, \mu_n$ that are conditionally unbiased in that $\mathrm{E}[V_i \mid \mu_1, \ldots, \mu_n] = \mu_i$ for all i. Let i^* denote the alternative with the maximal estimated value $V_{i^*} = \max\{V_1, \ldots, V_n\}$. Then,*

$$\mathrm{E}[\mu_{i^*} - V_{i^*}] \leq 0. \quad (1)$$

Moreover, if there is some chance of selecting the "wrong" alternative (i.e., μ_{i^*} not being maximal in $\{\mu_1, \ldots, \mu_n\}$), then $\mathrm{E}[\mu_{i^*} - V_{i^*}] < 0$.

Proof. Let us first consider a fixed set of true values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ with uncertain value estimates $\mathbf{V} = (V_1, \ldots, V_n)$. Let j^* denote the alternative with the maximum true value $\mu_{j^*} = \max\{\mu_1, \ldots, \mu_n\}$. With $\boldsymbol{\mu}$ fixed and $\mathbf{V}$ uncertain, j^* is a constant and i^* is a random variable. For any $\mathbf{V}$, we have

$$\mu_{i^*} - V_{i^*} \leq \mu_{j^*} - V_{i^*} \leq \mu_{j^*} - V_{j^*}. \quad (2)$$

The first inequality follows from the definition of j^* and the second from the definition of i^*. Taking expectations of (2) conditioned on $\boldsymbol{\mu}$ and integrating over the uncertainty regarding the value estimates (with distribution $\mathbf{V} \mid \boldsymbol{\mu}$), we have

$$\mathrm{E}[\mu_{i^*} - V_{i^*} \mid \boldsymbol{\mu}] \leq \mathrm{E}[\mu_{j^*} - V_{j^*} \mid \boldsymbol{\mu}] = 0, \quad (3)$$

with the equality following from our assumption that the value estimates are conditionally unbiased. Because $\mathrm{E}[\mu_{i^*} - V_{i^*} \mid \boldsymbol{\mu}] \leq 0$ for all $\boldsymbol{\mu}$, integrating over uncertain $\boldsymbol{\mu}$ yields $\mathrm{E}[\mu_{i^*} - V_{i^*}] \leq 0$ as stated in the proposition. If there is no chance of selecting the wrong alternative (i.e., $i^* = j^*$ with probability one), then the inequalities in (2), and hence in (3), all become equalities and $\mathrm{E}[\mu_{i^*} - V_{i^*}] = 0$. If a nonoptimal alternative is selected, then the first inequality in (2) will be strict. Thus, if there is some chance of this happening, then the inequality in (3) will be strict and $\mathrm{E}[\mu_{i^*} - V_{i^*}] < 0$. □

Thus, a decision maker who consistently chooses alternatives based on her estimated values should expect to be disappointed on average, even if the individual value estimates are conditionally unbiased. This optimizer's curse is quite general and does not rely on any of the specific assumptions (e.g., normal distributions) used in our illustrative examples.

3. What to Do About the Optimizer's Curse

The numerical examples of the previous section indicate that the effects of the optimizer's curse may be substantial. In this section, we consider the question of what to do about the curse: How should we adjust our value estimates to eliminate this effect? How should it affect decision making?

The key to overcoming the optimizer's curse is conceptually quite simple: model the uncertainty in the value estimates explicitly and use Bayesian methods to interpret these value estimates. Specifically, we assign a prior distribution on the vector of true values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and describe the accuracy of the value estimates $\mathbf{V} = (V_1, \ldots, V_n)$ by a conditional distribution $\mathbf{V} \mid \boldsymbol{\mu}$. Then, rather than ranking alternatives based on the value estimates, after we have done the decision analysis and observed the value estimates **V**, we use Bayes' rule to determine the posterior distribution for $\boldsymbol{\mu} \mid \mathbf{V}$ and rank and choose among alternatives based on the posterior means, $\hat{v}_i = \mathrm{E}[\mu_i \mid \mathbf{V}]$ for $i = 1, \ldots, n$. For example, in the models developed later in this section, the posterior mean is a weighted average of the value estimate V_i and prior mean $\bar{\mu}_i$,

$$\hat{v}_i = \alpha_i V_i + (1-\alpha_i)\bar{\mu}_i, \tag{4}$$

where α_i $(0 < \alpha_i < 1)$ is a weight that depends on the relative accuracies of the prior estimate and value estimate. These posterior means combine the information provided by the value estimates V_i with the decision maker's prior information, and could be interpreted as treating the value estimates somewhat skeptically.

By revising the value estimates in this way, we obtain posterior value estimates $\hat{v}_i$ that do not exhibit the optimizer's curse, either conditionally (given any particular estimate of **V**) or unconditionally. We formalize this result as follows.

Proposition 2. *Let* $\mathbf{V} = (V_1, \ldots, V_n)$ *be estimates of* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, *let* $\hat{v}_i = \mathrm{E}[\mu_i \mid \mathbf{V}]$, *and let* i^* *be the alternative with the maximal posterior value estimate* $\hat{v}_{i^*} = \max\{\hat{v}_1, \ldots, \hat{v}_n\}$. *Then,* $\mathrm{E}[\mu_{i^*} - \hat{v}_{i^*} \mid \mathbf{V}] = \mathrm{E}[\mu_{i^*} - \hat{v}_{i^*}] = 0$.

Proof. We prove this result by first conditioning on **V** and integrating out uncertainty about $\boldsymbol{\mu}$ given **V**. For a given set of value estimates **V**, the alternative i^* with the maximum posterior value estimate $\hat{v}_{i^*}$ is fixed. The conditional expectation of $\mu_{i^*} - \hat{v}_{i^*}$ is

$$\begin{aligned}\mathrm{E}[\mu_{i^*} - \hat{v}_{i^*} \mid \mathbf{V}] &= \mathrm{E}[\mu_{i^*} - \mathrm{E}[\mu_{i^*} \mid \mathbf{V}] \mid \mathbf{V}] \\ &= \mathrm{E}[\mu_{i^*} \mid \mathbf{V}] - \mathrm{E}[\mu_{i^*} \mid \mathbf{V}] = 0.\end{aligned}$$

The first equality here follows from the definition of $\hat{v}_{i^*}$ and the second equality follows from the linearity of expectations and the definition of conditional expectations. Thus, for every value estimate **V**, $\mathrm{E}[\mu_{i^*} - \hat{v}_{i^*} \mid \mathbf{V}] = 0$. If we then integrate over uncertainty in the value estimates, we find $\mathrm{E}[\mu_{i^*} - \hat{v}_{i^*}] = 0$. □

Thus, the decision maker who interprets value estimates as a Bayesian skeptic should not be expected to be disappointed on average.

Before we consider specific examples of Bayesian models, it may be useful to provide some intuition about how the revised estimates overcome the optimizer's curse. For models with posterior means of the form of Equation (4), the expected disappointment associated with a given value estimate V_i is $\mathrm{E}[V_i - \mu_i \mid \mathbf{V}] = V_i - \hat{v}_i = (1-\alpha_i)(V_i - \bar{\mu}_i)$. If we evaluate n alternatives and choose the one with the highest value estimate (alternative i^*), the expected disappointment is $\mathrm{E}[V_{i^*} - \mu_{i^*} \mid \mathbf{V}] = V_{i^*} - \hat{v}_{i^*} = (1-\alpha_{i^*}) \cdot (V_{i^*} - \bar{\mu}_{i^*})$. Note that as the number of alternatives increases, we expect V_{i^*} to increase because of the order statistic effect illustrated in Figure 2, and the expected disappointment will increase accordingly even if α_i does not change with n. On the other hand, if we base our decision on the revised estimates, choosing the alternative j^* with the highest revised estimate, the expected disappointment is $\mathrm{E}[\hat{v}_{j^*} - \mu_{j^*} \mid \mathbf{V}] = \hat{v}_{j^*} - \hat{v}_{j^*} = 0$. The key issue here is proper conditioning. The unbiasedness of the value estimates V_i discussed in §1 is unbiasedness conditional on $\boldsymbol{\mu}$. In contrast, we might think of the revised estimates $\hat{v}_i$ as being unbiased conditional on **V**. At the time we optimize and make the decision, we know **V** but we do not know $\boldsymbol{\mu}$, so proper conditioning dictates that we work with distributions and estimates conditional on **V**.

3.1. A Multivariate Normal Model

We illustrate this Bayesian approach by considering some standard models that demonstrate features of the general problem while allowing simple calculations; similar models are discussed in Gelman et al. (1995) and Carlin and Louis (2000). Suppose that the prior distribution on the vector of true values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is multivariate normal with mean vector $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, \ldots, \bar{\mu}_n)$ and covariance matrix $\boldsymbol{\mathcal{M}}'$; we abbreviate this as $\boldsymbol{\mu} \sim \mathrm{N}(\bar{\boldsymbol{\mu}}, \boldsymbol{\mathcal{M}}')$.[4] Further, suppose that, given the true values $\boldsymbol{\mu}$, the value estimates $\mathbf{V} = (V_1, \ldots, V_n)$ are also multivariate normal, with $\mathbf{V} \mid \boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\mathcal{V}})$; these value estimates are conditionally unbiased in that their expected value is equal to the true value. Then, applying standard results for the multivariate normal distribution, we can find the following unconditional (predictive) distribution on **V** and posterior distribution for $\boldsymbol{\mu} \mid \mathbf{V}$:

$$\mathbf{V} \sim \mathrm{N}(\bar{\boldsymbol{\mu}}, \boldsymbol{\mathcal{M}}' + \boldsymbol{\mathcal{V}}) \quad \text{and} \tag{5a}$$

$$\boldsymbol{\mu} \mid \mathbf{V} \sim \mathrm{N}(\hat{\mathbf{v}}, \boldsymbol{\mathcal{M}}''), \quad \text{where} \tag{5b}$$

$$\boldsymbol{\alpha} = \boldsymbol{\mathcal{M}}[\boldsymbol{\mathcal{M}} + \boldsymbol{\mathcal{V}}]^{-1}, \tag{5c}$$

$$\hat{\mathbf{v}} = \boldsymbol{\alpha}\mathbf{V} + (\mathbf{I} - \boldsymbol{\alpha})\bar{\boldsymbol{\mu}}, \quad \text{and} \tag{5d}$$

$$\boldsymbol{\mathcal{M}}'' = (\mathbf{I} - \boldsymbol{\alpha})\boldsymbol{\mathcal{M}}'. \tag{5e}$$

Thus, the posterior mean for alternative i, $\hat{v}_i = \mathrm{E}[\mu_i \mid \mathbf{V}]$, given as the ith component of $\hat{\mathbf{v}}$, is a combination of the prior mean vector $\bar{\boldsymbol{\mu}}$ and observed value estimates **V** with the mixing described by the matrix $\boldsymbol{\alpha}$.

3.2. The Independent Normal Case

If we assume that the true values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ are independent and the value estimates $\mathbf{V} = (V_1, \ldots, V_n)$ are conditionally independent given $\boldsymbol{\mu}$, then the covariance matrices $\boldsymbol{\mathcal{M}}'$ and $\boldsymbol{\mathcal{V}}$ are diagonal and the

[4] We adopt the convention that all vectors are column vectors unless indicated otherwise.

Bayesian updating process decouples into independent processes for each alternative. Specifically, suppose that $\mu_i \sim \mathrm{N}(\bar{\mu}_i, \sigma^2_{\mu_i})$ and $V_i \mid \mu_i \sim \mathrm{N}(\mu_i, \sigma^2_{V_i})$. Then, the posterior distribution $\mu_i \mid V_i \sim \mathrm{N}(\hat{v}_i, \sigma^2_{\mu_i \mid V_i})$, where

$$\alpha_i = \frac{1}{1+\sigma^2_{V_i}/\sigma^2_{\mu_i}}, \tag{6a}$$

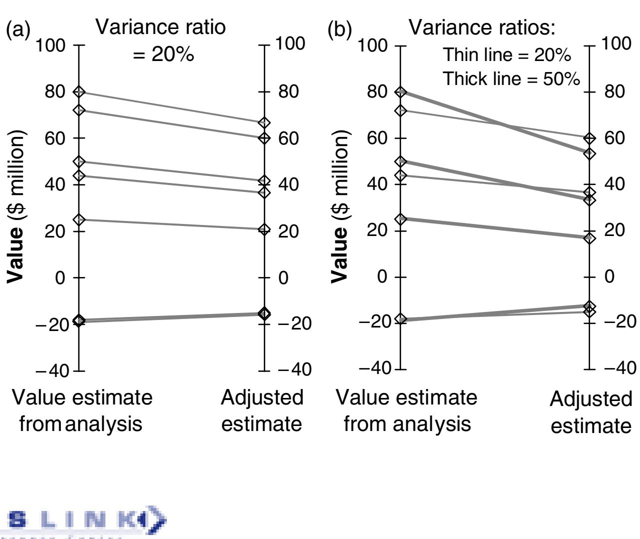
$$\hat{v}_i = \alpha_i V_i + (1-\alpha_i)\bar{\mu}_i, \quad \text{and} \tag{6b}$$

$$\sigma^2_{\mu_i \mid V_i} = (1-\alpha_i)\sigma^2_{\mu_i}. \tag{6c}$$

Thus, the posterior value estimate $\hat{v}_i = \mathrm{E}[\mu_i \mid \mathbf{V}]$ depends on the prior mean $\bar{\mu}_i$ and the variance ratio $\sigma^2_{V_i}/\sigma^2_{\mu_i}$ that describes the relative accuracy of the estimation process. If the estimation process is very accurate, $\sigma^2_{V_i}$ will be small compared to $\sigma^2_{\mu_i}$ and α_i will approach one. In this case, the posterior mean $\hat{v}_i$ will approach the value estimate V_i. With less accurate estimation processes, the posterior mean $\hat{v}_i$ will be a convex combination of V_i and $\bar{\mu}_i$, and thus will be shrunk toward the prior mean $\bar{\mu}_i$.

We can demonstrate this simple model by applying it to a specific study performed at Kodak in 1999, identified in Clemen and Kwit (2001) as study 99-9. Figures 5(a) and (b) show the results. In Figure 5(a), we assume that all of the seven alternatives have a common variance ratio ($\sigma^2_{V_i}/\sigma^2_{\mu_i}$) of 20% and a common prior mean $\bar{\mu}_i = 0$. (The bottom two alternatives have very similar values.) The value estimates for each alternative (V_i) are shown on the left side of the figure and the revised value estimates ($\hat{v}_i$) for the same alternatives are shown on the right, with each line connecting the two estimates for a particular alternative. Thus, the seven lines in each figure represent the seven alternatives. In this case, the value estimates are all shrunk toward the prior mean with a common weight $\alpha_i = \alpha = 0.8333$ on each value estimate and $1-\alpha = 0.1667$ on the prior mean. The value of the recommended alternative is shrunk from \$80.0 to \$66.7 million and the value added by the analysis (the difference between the maximal value and the average for all alternatives) is reduced from \$46.6 to \$38.8 million.

Figure 5 **Shrinkage Estimates with Common Variance Ratios (a) and Different Variance Ratios (b)**



While the ranking of alternatives is not changed if the alternatives all have the same prior mean and variance ratio, shrinkage may lead to different rankings when the variance ratios differ across alternatives. In particular, alternatives whose values are hard to estimate may be passed over by alternatives with lower, but more accurate, value estimates. This is demonstrated in Figure 5(b). Here, we use the same value estimates and prior mean as in Figure 5(a), but allow the different alternatives to have variance ratios of either 20% or 50%, as noted in Figure 5(b). In this case, the alternative with the highest value estimate V_i is not preferred when ranked by posterior means: the difficulty in accurately estimating its value causes the estimate to be treated more skeptically and shrunk further toward the prior mean. Examining Equation (6b), it is easy to see that the rankings of alternatives may also be affected by differences in prior means.

3.3. Identical Covariance Structures

As indicated in §2.1, in many applications we might expect the value estimates for the different alternatives to be correlated because the alternatives share some common elements, and the true values might be correlated also for the same reasons. There we mentioned the example of a study considering alternative ways to pursue an R&D project, where a probability of technical success and the estimate of this probability are relevant for the different alternatives being considered. If the value estimates and true values depend on such underlying factors in the same way, we might expect the covariance matrices for the value estimates and true values ($\boldsymbol{\mathcal{V}}$ and $\boldsymbol{\mathcal{M}}'$) to be similar in some sense.

The Bayesian updating formulas simplify considerably if we assume that the two covariance matrices are identical up to a change of scale, $\boldsymbol{\mathcal{V}} = \gamma\boldsymbol{\mathcal{M}}'$. Here, we can think of the value estimation process (with covariance matrix $\boldsymbol{\mathcal{V}}$) preserving the same covariance structure as the true values, but with the uncertainty changed by a variance ratio of γ. In this case, the mixing matrix $\boldsymbol{\alpha} = \boldsymbol{\mathcal{M}}'[\boldsymbol{\mathcal{M}}' + \boldsymbol{\mathcal{V}}]^{-1} = \boldsymbol{\mathcal{M}}'[\boldsymbol{\mathcal{M}}' + \gamma\boldsymbol{\mathcal{M}}']^{-1} = (1+\gamma)^{-1}\mathbf{I}$, where $\mathbf{I}$ is the n-by-n identity matrix. The posterior means are given by

$$\hat{v}_i = \alpha_i V_i + (1-\alpha_i)\bar{\mu}_i, \tag{7}$$

and the posterior covariance matrix is $\boldsymbol{\mathcal{M}}'' = (1-\alpha)\cdot\boldsymbol{\mathcal{M}}'$, where $\alpha_i = \alpha = (1+\gamma)^{-1}$ for all i. Comparing this α_i with Equation (6a), we see that the variance ratio γ plays exactly the same role in terms of α_i as in the independent case, and the posterior means are determined separately in exactly the same manner in Equations (6b) and (7). Thus, when the two covariance matrices share a common structure, the dependencies among the V_is and among the μ_is "cancel" in terms of their effects on the revised value estimates in the sense that for a given $\mathbf{V}$, we wind up with the same revised estimates that would be given in the independent case. For example, the results of Figure 5(a) could have been generated by a model where the value estimates and true values are correlated and $\boldsymbol{\mathcal{V}} = \gamma\boldsymbol{\mathcal{M}}'$, with $\gamma = 20\%$.[5]

3.4. A Hierarchical Model

The two previous examples of Bayesian models rely heavily on the assessment of the prior mean for each alternative. In some cases, a decision maker may be able to specify these means without much difficulty. In other cases, however, such assessments may be difficult and the decision maker may want to use the observed value estimates themselves to estimate the mean of the μ_is. A natural Bayesian way to do this is to treat the mean true value itself as uncertain and use the observed value estimates $\mathbf{V} = (V_1, \ldots, V_n)$ to update the prior distribution of this mean true value. We now describe a hierarchical model that illustrates these features.

Suppose that the value estimates are independent and multivariate normal with identical accuracy: $\mathbf{V} \mid \boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma_V^2\mathbf{I})$. The true values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ are uncertain and drawn independently from a multivariate normal distribution: $\boldsymbol{\mu} \mid \bar{\mu} \sim \mathrm{N}(\bar{\mu}\mathbf{1}, \sigma_\mu^2\mathbf{I})$, where $\bar{\mu}$ is a scalar and $\mathbf{1}$ is an n-vector of ones, $(1, \ldots, 1)$. Finally, at the top level of the hierarchy, suppose that the mean true value, $\bar{\mu}$, is uncertain and has a univariate normal prior distribution: $\bar{\mu} \sim \mathrm{N}(\mu_0, \sigma_0^2)$. Thus, in this model, it is as if we draw true values at random from a distribution whose mean is uncertain. We then observe independent estimates of these true values and use these estimates to update our estimates of the individual true values $(\mu_1, \ldots, \mu_n)$ and the mean true value $\bar{\mu}$.

This hierarchical model fits into the two-level model developed in §3.1 by taking $\bar{\boldsymbol{\mu}} = \mu_0\mathbf{1}$, $\boldsymbol{\mathcal{V}} = \sigma_V^2\mathbf{I}$, and $\boldsymbol{\mathcal{M}}' = \sigma_\mu^2\mathbf{I} + \sigma_0^2\mathbf{1}\mathbf{1}^{\mathrm{T}}$. With these special assumptions, we can find an analytic form for the mixing matrix $\boldsymbol{\alpha}$ (Equation (5c)) and the posterior means $\hat{\mathbf{v}} = \boldsymbol{\alpha}\mathbf{V} + (\mathbf{I} - \boldsymbol{\alpha})\bar{\boldsymbol{\mu}}$. Specifically, $\hat{v}_i = \mathrm{E}[\mu_i \mid \mathbf{V}]$ becomes a weighted combination of the value estimate V_i, the prior mean μ_0, and the average value estimate $\bar{V} = \sum_{i=1}^n V_i/n$:

$$\hat{v}_i = w_1 V_i + w_2\mu_0 + w_3\bar{V}, \tag{8}$$

with weights

$$w_1 = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_V^2} = \frac{1}{1 + \sigma_V^2/\sigma_\mu^2}, \tag{8a}$$

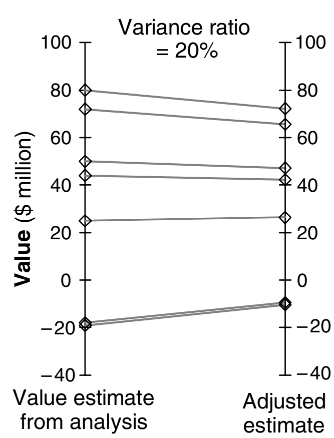
$$w_2 = \frac{\sigma_V^2}{n\sigma_0^2 + \sigma_\mu^2 + \sigma_V^2}, \quad \text{and} \tag{8b}$$

$$w_3 = \frac{n\sigma_0^2\sigma_V^2}{(\sigma_\mu^2 + \sigma_V^2)(n\sigma_0^2 + \sigma_\mu^2 + \sigma_V^2)}. \tag{8c}$$

With a little algebra, we can see that these weights sum to one. Note that the weight w_1 on the value estimate V_i is of the same form as the weights α_i in the two previous models (Equations (6a) and (7)), with the remaining weight $1 - w_1$ split between the prior mean μ_0 and the average of the value estimates $\bar{V}$. The weight w_1 on the value estimate does not depend on the uncertainty in the prior mean (σ_0^2), but the allocation of the remaining weight to the prior mean μ_0 and average value estimate $\bar{V}$ depends on this uncertainty. As our uncertainty about the mean true value decreases (i.e., $\sigma_0^2 \to 0$), the weight on the average value estimate $\bar{V}$ approaches 0. On the other hand, as our uncertainty about the true mean increases (i.e., $\sigma_0^2 \to \infty$), the weight on the prior mean μ_0 approaches 0. In this latter case, the posterior mean $\hat{v}_i$ reduces to a weighted combination of the value estimate V_i and the observed average value $\bar{V}$, exactly like the independent case in §2.2, but with the average value $\bar{V}$ appearing in place of the prior mean $\bar{\mu}_i$.

We can illustrate this hierarchical model by applying it to the same example considered in Figure 5(a). In Figure 6, we show the results in the limiting case, where we assume little prior information about

Figure 6 Shrinkage Estimates with a Hierarchical Model



[5] While the adjustments to the value estimates do not vary with the degree of correlation when they share a common correlation structure, the overall expected disappointment may vary because it depends on the joint distribution of true values and value estimates. For example, the diagonal cases in Table 2 have common covariance structures and the overall expected disappointment varies with the degree of correlation.

the mean true value ($\sigma_0^2 \to \infty$) and consider a variance ratio of 20%, as in Figure 5(a). In this case, the posterior mean for each alternative is a weighted average of the value estimate for that alternative and the average value estimate for all alternatives: specifically $\hat{v}_i = 0.833V_i + (1-0.833)\overline{V}$, where $\overline{V} = \$33.4$ million with these alternatives. The shrinkage in this case is then similar to the shrinkage in the example in Figure 5(a), but with the value estimates shrunk toward a mean of \$33.4 million instead of the prior mean of 0 assumed earlier. For example, the maximum value estimate of \$80 million is shrunk to \$72.4 million in this case rather than \$66.7 million in the earlier example. The value added by the analysis is \$38.8 million rather than the \$46.2 million calculated in Clemen and Kwit (2001). Shrinkage with this model does not lead to any changes in the ranking of alternatives, because we have assumed the variances are equal for all alternatives. If we used a more complex hierarchical model with differing variances, we might find changes in rankings like those shown in Figure 5(b).

3.5. Assessment Issues

While the models considered in §§3.2–3.4 are introduced primarily as examples that demonstrate Bayesian procedures for adjusting value estimates, the fact that these models allow simple calculations may make them useful as rough approximations in practice. The rules for adjusting value estimates given by these models all reduce to a revised estimate that is a weighted average of the form

$$\hat{v}_i = \alpha_i V_i + (1-\alpha_i)\bar{\mu}_i, \tag{9}$$

where in the hierarchical case, the prior mean $\bar{\mu}_i$ is replaced by a mix of the average value estimate $\overline{V}$ and the prior mean μ_0. In each of these models, $\alpha_i = (1+\gamma_i)^{-1}$, where γ_i is a variance ratio equal to $\sigma_{V_i}^2/\sigma_{\mu_i}^2$ or the matrix equivalent of that in §3.2.

To apply Equation (9), we need to assess a prior mean and a variance ratio. We believe that many decision makers would be comfortable assessing a prior mean for a given alternative before observing the results of an analysis. Ideally, these assessments would be made before the analysis is begun. In practice, however, the alternatives under consideration often evolve during the analysis, and it may be difficult to get a truly "prior" assessment. Nevertheless, these assessments might be made before seeing the final value estimates from the analysis or might be made as a hypothetical exercise after the analysis, e.g., by asking questions like: Before you saw the results of this analysis, what would you have estimated the value to be?

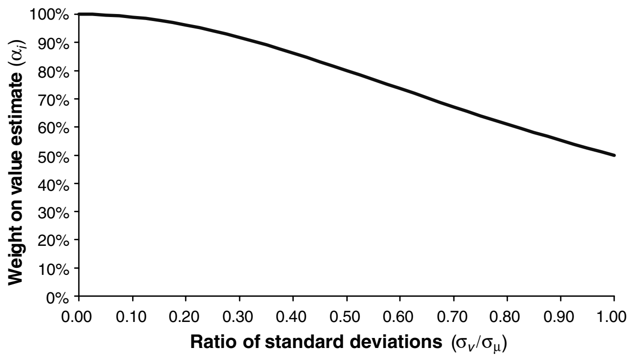
To assess the variance ratio, we could assess σ_{V_i} and σ_{μ_i} and calculate the ratio. Our sense is that σ_{μ_i} may often be fairly straightforward to assess: as a decision maker assesses a prior mean, we could prompt for a range (e.g., the 10th and 90th percentiles) that could be used to determine σ_{μ_i}. Asking about a range could increase the comfort level of the decision maker about assessing the mean if he is quite uncertain about the value of the alternative. We could also ask hypothetical questions after the analysis like: Before you saw the results of this analysis, how uncertain were you about the value?

Assessing σ_{V_i} requires contemplating questions such as: If the true value of an alternative is x, what range of values would you expect to see resulting from the analysis? When teaching complex decision analysis cases, we have been struck by the range of value estimates given by different student teams. Sometimes the differences in estimates reflect modeling errors, and other times the differences reflect reasonable variation in interpretations of facts in the case. The range of student answers will depend on the ambiguity in the case, how good the students are at analysis, and how much time and effort the students put into the case. The assessment of σ_{V_i} in professional applications will depend on similar considerations, and we would expect these assessments and ratios to vary significantly across different applications.

Alternatively, rather than assessing the accuracy of the analysis σ_{V_i} and calculating the variance ratio, we might think about assessing α_i, the weight on the value estimate for alternative i, by comparing the uncertainty before and after the analysis. Noting that the posterior variances are given by $\sigma_{\mu_i|V_i}^2 = (1-\alpha_i)\sigma_{\mu_i}^2$ (or the matrix equivalent in §3.3), we can think about α_i as the fraction of the prior uncertainty about the value of alternative i (measured as a variance) eliminated by doing the analysis. We can estimate α_i by assessing the prior variance $\sigma_{\mu_i}^2$ (as discussed before) and assessing the posterior variance $\sigma_{\mu_i|V_i}^2$. We might assess $\sigma_{\mu_i|V_i}^2$ after the analysis by asking questions such as: If you had another year and unlimited resources for additional analysis, how much might the estimate change? If we prompt for a range (e.g., the 10th and 90th percentiles) for the potential change, we could use this to estimate $\sigma_{\mu_i|V_i}^2$ and then calculate α_i.

Figure 7 shows a plot of the weight α_i on the value estimate V_i as a function of the ratio of standard deviations, σ_{V_i} and σ_{μ_i}. Here, we see that if the value estimates are quite accurate compared to the prior estimate ($\sigma_{V_i}/\sigma_{\mu_i}$ is near zero), the weight on the value estimate is very near one. The weights initially decrease slowly as this ratio increases. While it is difficult to speculate about what these values will be in practice, suppose, for example, that the ratios $\sigma_{V_i}/\sigma_{\mu_i}$ are in the 20%–50% range. This would imply that when revising value estimates, we should put

Figure 7 Relationship Between the Weight on the Value Estimate and Accuracy Assessments



weights α_i of 95%–80% on the value estimate from the analysis and weights $1 - \alpha_i$ of 5%–20% on the prior estimate of value. Thinking in terms of variance reduction, this would correspond to analysis reducing the uncertainty about the true value by 80%–95%. We suggest these numbers only as a representative range, and encourage others to think carefully through these issues and develop their own assessments in the context of the particular problem at hand.

4. Related Biases and "Curses"

We believe that the optimizer's curse is at best underappreciated and at worst unrecognized in decision analysis and management science. As mentioned earlier, the phenomenon has been noted, though not studied in detail, in other settings. In the finance literature, Brown (1974) considers the context of revenue and cost estimates in capital budgeting. He observes that a project is more likely to be accepted if its revenues have been overestimated and its costs underestimated, and that the selection process may thus introduce a bias of overestimating the value of accepted projects. In the organizational behavior literature, Harrison and March (1984) label this phenomenon postdecision surprise or postdecision disappointment, demonstrate it with a simple normal model, and discuss organizational implications of the phenomenon. Neither Brown (1974) nor Harrison and March (1984) present a general result like our Proposition 1 or offer constructive advice about what to do about the curse as we do in §3. Harrison and March (1984, p. 38) conclude: "Intelligent decision making with unbiased estimation of future costs and benefits implies a distribution of postdecision surprises that is biased in the direction of disappointment. Thus, a society that defines intelligent choice as normatively appropriate for individuals and organizations imposes a structural pressure toward postdecision unhappiness." Brown (1974) concludes by calling for additional research about how to overcome the curse.

As noted in the introduction, the optimizer's curse is analogous to the winner's curse (Capen et al. 1971, Thaler 1992), which refers to the tendency for the highest bidder in an auction with common or interdependent values to have overestimated the value of the item being sold. The underlying argument is similar to that of the optimizer's curse, with overestimating the value increasing the chance of winning the auction. A Bayesian analysis of the situation, looking in advance at the expected value of the item to a bidder given that the bidder wins, indicates that bidders should hedge their bids by bidding less than their value estimates (Winkler and Brooks 1980). However, the competitive situation of the auction is very different from the situation considered here and requires the use of game-theoretic reasoning. The analysis of these auctions typically requires strong common knowledge assumptions and, for tractability, often focuses on symmetric equilibria (e.g., Krishna 2002). The optimizer's curse is a much more prevalent phenomenon that, as Harrison and March (1984) argue, affects all kinds of intelligent decision making, not just competitive bidding problems.

There is also a connection between the optimizer's curse and the survivorship bias effect noted in the finance literature (e.g., Brown et al. 1992). For example, the average performance of mutual funds is inflated because poorly performing funds are closed. Decision analysis and other optimization processes, with the choice of only one alternative from a set of possibilities, represent the extreme form of survivorship bias, with only one survivor. A publication bias or reporting bias in which only favorable results are reported yields results similar in nature to the survivorship bias, as do some forms of sampling bias (e.g., sampling plans that lead to systematic underrepresentation of lower income households in surveys of family expenditures).

The optimizer's curse is also related to regression to the mean, a phenomenon that is often discussed when regression is introduced in basic statistics courses. For example, high performers on one test are likely to perform less well on subsequent tests. Thus, their expected performance on subsequent tests is less than their performance on the first test, analogous to the expected true value of an alternative being less than its estimated value. Harrison and March (1984) describe postdecision disappointment as a form of regression to the mean. The notion of shrinkage estimators, such as those developed in §3, has connections with regression to the mean as well as with empirical Bayes and hierarchical Bayes procedures (e.g, Gelman et al. 1995, Carlin and Louis 2000). The models in §3 result in shrinkage from the value estimates toward a prior mean, toward the mean of all of the value estimates, or toward some convex combination of those two.

This process for adjusting value estimates may seem somewhat similar to ambiguity aversion or uncertainty aversion (Ellsberg 1961) in that it penalizes good alternatives for uncertainty in the value estimates, with the magnitude of the penalty increasing with the uncertainty in the estimate. However, the shrinkage process here will adjust low values upward toward the prior mean. In contrast, ambiguity aversion leads only to values being adjusted down for uncertainty in the value estimates.

Bell (1985) considers the implications of aversion to disappointment for decision making and decision modeling under uncertainty, where he defines disappointment as "a psychological reaction caused by comparing the actual outcome of a lottery to one's prior expectations" (p. 1). When an analysis has been performed, the value estimate V_{i*} of a chosen alternative serves as a natural reference point to which the true value is compared when it is learned. To the extent that decision makers associate some disutility with disappointment, as Bell (1985) suggests, the inflation of value estimates caused by the optimizer's curse would increase this disappointment and decrease the decision maker's expected and experienced utility.

Finally, our work has connections to the decision analysis literature on using experts. Our advice in §3 calls for explicitly modeling the results of analysis as uncertain and suggests the use of Bayesian techniques for interpreting these results. In essence, we recommend viewing the result of an analysis as being analogous to an expert report and treating it in much the same way as Morris (1974) recommends. While this Bayesian approach to interpreting decision analysis results has been considered by Nickerson and Boyd (1980) and Lindley (1986), these authors did not note the optimizer's curse. The expert-use literature may provide suggestions for developing Bayesian models analogous to those discussed in §3 for addressing the curse.

5. Conclusions

The primary goal of this paper is to make the decision analysis and management science communities aware of the optimizer's curse, and to help people understand how the curse affects the results of an analysis and how it can be addressed. The curse may be summarized as follows: If decision makers take the value estimates resulting from an analysis at face value and select according to these estimates, then they should expect to be disappointed on average, not because of any inherent bias in the estimates themselves, but because of the selection process itself. The numerical examples of §2 suggest that this expected disappointment may be significant in practice. The expected disappointment will be even greater, if, as is often suspected, the value estimates themselves are biased high.

It would be interesting, but we suspect quite difficult, to document the optimizer's curse using field data. It has proven difficult to document the winner's curse by using field data (see, e.g., Thaler 1992 and Porter 1995), and we think that it would be at least as difficult to document the optimizer's curse in this way. First, few firms keep careful records documenting their decision-analysis-based value estimates (V_i) and, unlike bids in auctions, these estimates are not public information. Second, even if we had data on the value estimates, as with the winner's curse, it may be quite difficult to determine the corresponding actual values (x_{i*}) for individual project decisions and to isolate the effects of the curse from other confounding factors. In the absence of reliable field data, it could be interesting to study the curse in a controlled laboratory experiment in which subjects would be asked to estimate values for complex alternatives, and then asked to choose one of these alternatives. The winner's curse has been found repeatedly in such laboratory settings (see, e.g., Kagel and Levin 2002).

The key to overcoming the optimizer's curse is conceptually very simple: treat the results of the analysis as uncertain and combine these results with prior estimates of value using Bayes' rule before choosing an alternative. This process formally recognizes the uncertainty in value estimates and corrects for the bias that is built into the optimization process by adjusting high estimated values downward. To adjust values properly, we need to understand the degree of uncertainty in these estimates and in the true values. Although it may be difficult in practice to formulate and assess sophisticated models that describe the uncertainty in true values and value estimates given by a complex analysis, the models we develop in §3 could perhaps be used to adjust value estimates or components of these estimates approximately in these settings.

Analysts are frequently frustrated by having their results treated skeptically by clients and decision makers: the analysts work hard to be objective and unbiased in their appraisals only to find their values and recommendations discounted by the decision makers. This "discounting" may manifest itself by the decision maker insisting on using an excessively high hurdle or discount rate or by the decision maker exhibiting what may appear to be excessive risk aversion, ambiguity aversion, or disappointment aversion. The optimizer's curse suggests that such skepticism may well be appropriate. The skeptical view of experienced decision makers may in fact be (in part) a learned response to the effects of the curse, resulting from informal comparisons of estimates of value to actual outcomes over time. The Bayesian methods for adjusting value estimates can be viewed as a disciplined method for discounting the results of an analysis in an attempt to avoid postdecision disappointment. They require the decision maker to think carefully about her prior estimates of value and the accuracy of the value estimates, and to properly integrate her prior opinions into the analysis.

In summary, returning to the executive mentioned in the opening paragraph of this paper: Yes, she does have reason to be skeptical of the results of the decision analysis. To arrive at values and recommendations she trusts, she should get involved in the analysis to be sure that it properly includes her opinions and knowledge and overcomes the optimizer's curse.

Acknowledgments

The authors thank Bob Clemen for sharing the data from Clemen and Kwit (2001) and Clemen, David Bell, the associate editor, and the referees for helpful comments.

References

Bell, D. E. 1985. Disappointment in decision making under uncertainty. *Oper. Res.* **33** 1–27.

Brown, K. C. 1974. A note on the apparent bias of net revenue estimates for capital investment projects. *J. Finance* **29** 1215–1216.

Brown, S. J., W. Goetzmann, R. J. Ibbotson, S. A. Ross. 1992. Survivorship bias in performance studies. *Rev. Financial Stud.* **5** 553–580.

Capen, E. C., R. V. Clapp, W. M. Campbell. 1971. Competitive bidding in high-risk situations. *J. Petroleum Tech.* **23** 641–653.

Carlin, B. P., T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Chapman and Hall/CRC, Boca Raton, FL.

Clemen, R. T., R. C. Kwit. 2001. The value of decision analysis at Eastman Kodak Company, 1990–1999. *Interfaces* **31** 74–92.

Ellsberg, D. 1961. Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.* **75** 643–669.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Rubin. 1995. *Bayesian Data Analysis*. Chapman and Hall, London, UK.

Harrison, J. R., J. G. March. 1984. Decision making and postdecision surprises. *Admin. Sci. Quart.* **29** 26–42.

Kagel, J. H., D. Levin. 2002. *Common Value Auctions and the Winner's Curse*. Princeton University Press, Princeton, NJ.

Krishna, V. 2002. *Auction Theory*. Academic Press, San Diego, CA.

Lindley, D. V. 1986. The reconciliation of decision analyses. *Oper. Res.* **34** 289–295.

Lindley, D. V., A. Tversky, R. V. Brown. 1979. On the reconciliation of probability assessments. *J. Roy. Statist. Soc. A* **142** 146–180.

Matheson, J. E. 1968. The economic value of analysis and computation. *IEEE Trans. Systems Sci. Cybernetics* **SSC-4** 325–332.

Morris, P. A. 1974. Decision analysis expert use. *Management Sci.* **20** 1233–1241.

Nickerson, R. C., D. W. Boyd. 1980. The use and value of models in decision analysis. *Oper. Res.* **28** 139–155.

Porter, R. H. 1995. The role of information in U.S. offshore oil and gas lease auctions. *Econometrica* **63** 1–27.

Tani, S. N. 1978. A perspective on modeling in decision analysis. *Management Sci.* **24** 1500–1506.

Thaler, R. H. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Free Press, New York.

Winkler, R. L., D. G. Brooks. 1980. Competitive bidding with dependent value estimates. *Oper. Res.* **28** 603–613.