Should Consumers Use the Halo to Form Product Evaluations?
A Theoretical and Experimental Study

Peter Boatwright
Ajay Kalra
Wei Zhang

November, 2004

Peter Boatwright is an Associate Professor of Marketing and Ajay Kalra is an Associate Professor of Marketing at the Tepper School of Business, Carnegie Mellon University, Pittsburgh PA 15213. Wei Zhang is Senior Manager at Amgen, Inc. The order of authorship is alphabetical.
Should Consumers Use the Halo to Form Product Evaluations?
A Theoretical and Experimental Study

Abstract

In purchase situations where attribute information is either missing or difficult to judge, a well-known “rule of thumb” consumers use to form evaluations is the halo effect. The psychology literature has widely considered the halo to reflect consumers’ inability to discriminate between different attributes and have therefore labeled it as the “halo error” or the “logical error.” The objective of this paper is to offer a rationale for the halo effect. We use a decision-theory framework to show that the halo is consistent with the goal of minimizing estimation risk. Contrary to conventional wisdom we demonstrate that a decision using the halo allows consumers to make better evaluations of an object as compared to not using the halo heuristic. Therefore, using the halo results in utility maximization and is indicative of rational behavior. We draw some implications for advertising strategies and perform experimental tests.
Introduction

That consumers make purchase decisions with incomplete information is a well-established fact. In some purchase situations consumers face an overload of product attribute information, leading to use of heuristics to simplify decision making (e.g., Payne, Bettman, and Johnson et al 1993). In other purchase scenarios some of the product attribute information may be missing or may be difficult to evaluate objectively. Examples of these situations include products or services that have attributes with experience or credence properties. Here, consumers are known to make inferences about the missing attribute utilizing information from the available attributes (e.g. Huber and McCann 1982). The process used to make inferences is characterized by the well-known halo effect.

The term halo effect refers to two broad effects. The first one referred to in the literature is the interdimensional similarity halo, where a person tends to rate an object similarly across different dimensions. In the marketing context this implies consumers will use an observable attribute to infer an unobservable one. The second effect refers to the general impression halo, where a person’s overall evaluation or impression leads her to evaluate all aspects of performance. The halo effect has a rich history in both the psychology as well as the business literatures and has been used to explain several phenomena such as attitude formation (Beckwith and Lehmann 1975) and technology licensing (Sine, Shane and Gregorio, 2003)

Psychologists have long believed the halo represents an example of how people deviate from optimal behavior. The literature has dubbed both the general impression halo and the interdimensional bias as the “halo error” or the “logical error” (Balzer and Sulsky 1992; Cooper 1981; Feldman 1986; Lance and Woehr 1986)). In fact, the literature also assumes that “removing the halo is both plausible and worthwhile.” (page 219, Murphy, Jako, and Anhalt 1993). In contrast to the literature, we offer an explanation for why this so-called nonoptimal behavior is in fact quite reasonable.
We focus our attention on the interdimensional bias as, from the marketing perspective, the processes leading to the formation of overall global evaluations are more relevant.\(^1\) We demonstrate that by using what we term the halo estimator, estimation risk is reduced. Our halo estimator is equivalent to the James-Stein estimator (Stein 1955; James and Stein 1960). Our use of the James-Stein estimator is as a representative of James-Stein-type estimators and even more generally as a representative of “shrinkage” estimators, for there are many alternatives that share similar properties. Shrinkage estimators have been particularly useful in limited data situations, such as for portfolio analysis in finance (Jorion, 1986) or for estimating household-specific model parameters in marketing (Rossi and Allenby 1993). The domain of our research is the inference of any attribute value that has experience or credence properties. We examine the impact of using the interdimensional halo process on changes in estimation risk, where estimation risk is the expected value of the estimation loss, and where loss is defined as the difference between the belief and true value of a product attribute. Our theory provides a motive for a consumer to use the halo effect when she makes inferences about unobserved or missing attribute information. Contrary to conventional wisdom, therefore, using the halo is not a logical “error” as subscribed to in the psychology literature.

By understanding the process through which the halo effect reduces estimation risk, the paper also obtains an understanding of how consumers use information to draw inferences. The process has important implications for firms designing advertising strategies. A major decision a firm focus is to choose the attributes that will be used in an advertising campaign where the firm exposes consumers to multiple messages. Firms often face the choice of exposing consumers to information about a few (or one) attributes of a brand many times (e.g., Maytag’s use of reliability) or about several attributes (Chevrolet’s use of comfort, safety, reliability, etc). Based

---

\(^1\) For the sake of completeness, however, we also demonstrate that the results hold for both interpretations. See the appendix for a discussion of the general impression halo.
on the model the paper provides suggestions as to which approach is more effective and the conditions under which these strategies should be used.

The rest of the paper is organized as follows. We first discuss the model assumptions, provide a rationale for the halo effect, and derive the propositions. Next, we provide a numerical example to illustrate the model. We then describe experimental tests and results. Finally, we discuss the conclusions and implications.

**Theoretical Framework**

In our model we examine a common scenario where a consumer purchases a product with multiple attributes. Before purchase a consumer can obtain information about a product’s attributes by multiple means: trial use, word-of-mouth, advertisements, and so on. Even trial use of a product will not always give a consumer a precise understanding of true performance levels of experience attributes, for some attribute values are more easily inferred than others. For example, a single usage of a particular brand of toothpaste can give a consumer quite precise knowledge about its taste. Even after being loyal to a brand of toothpaste, however, a consumer might not be accurate when inferring\(^2\) the cleaning power of that brand. This lack of precise knowledge applies to all experience attributes.

Although a consumer does not have precise attribute information for many attributes, knowledge of product attributes improves with increasing information (from experience). For instance, a consumer will not be able to determine the exact life of a Duracell battery before she uses it. If this consumer has used one Duracell battery that lasted 130 minutes and one that lasted 110 minutes, however, she might infer the average lasting time of a Duracell battery is 120 minutes. In this context the consumer has experienced two observations of battery life, which we

---

\(^2\) The meaning of “infer” is different in the marketing and statistics literature. In the marketing literature the term is used when an individual draws an opinion on an unobserved attribute from attributes whose values are known. In statistics inferences are drawn on any attribute when information on that attribute has any uncertainty at all. In this paper we use the definition from statistics.
can label $x_1$ and $x_2$, where $x_1=130$ and $x_2=110$. Because the consumer is interested in the typical battery life (which we label $\theta$), she may estimate it using the simple average of $x_1$ and $x_2$.

The example about Duracell batteries characterizes our general framework, in which we assume information on all the product attributes relevant in this paper to be random variables ($x$), meaning a consumer has some degree of uncertainty regarding the typical (true population) value of each product attribute ($\theta$). The uncertainty of attribute values comes from at least three sources. First, variation in the production process could result in some randomness of attribute performance. Second, a particular product may perform differently in different conditions or at different times. A car’s fuel efficiency will change according to road conditions and weather. Thus, even for the same consumer, her perception about the same attribute of the same product can be quite different across time periods. Third, advertising is an important source for consumers to acquire product information. Since it is legal for manufacturers to use puffery, consumers may find it difficult to assess the degree of the overstatement of the claims made. All these variations in perception are caused by some uncontrolled or unobserved factors that create randomness in the perception of values of each product attribute. Technically, even attributes whose values are immediately obvious can be considered to have random outcomes, in that the variance of the random outcome could be close to zero. Our paper focuses on attributes with greater uncertainty—the more interesting case to marketers.

In sum the attributes we discuss can be characterized as experience attributes rather than search attributes. The importance of experience attributes is well established in the literature. Roberts and Urban (1988) find that “consumers generally make decisions with some uncertainty about the true value of attributes that they will obtain.” They also believe the existence of uncertainty about the true value of attributes is due to “inherent product variability” and “imperfect information.” They suggested that as a consumer gains more information about the
product, her beliefs about the mean value and the uncertainty of the product would be updated accordingly.

In the case of products with multiple attributes, all attribute values need to be inferred. Expanding our notation to allow for multiple attributes, let $x_{ij}$ represent the $j^{th}$ observation of attribute $i$, and let $\theta_i$ represent the true (unobserved) value of attribute $i$. Each observation is denoted as $x_{ij}, j = 1, 2, \ldots, N$, where $N$ is the number of observations. Then

$$x_{ij} = \theta_i + \epsilon_{ij}, \quad (1)$$

where $\epsilon_{ij}$ is a random variable that can be characterized by some distribution. Just as in our earlier example for a Duracell battery, one reasonable estimate of $\theta_i$ is simply $\bar{x}_i$, the arithmetic mean of observations of attribute $i$. So far the implication is that assessment of an attribute is made on experience and information about that attribute alone. The corollary is that information about this attribute will not be used to infer other attributes, unless of course the attributes are correlated. The literature on the halo effect, however, has observed that people use information on one attribute when inferring a second. This effect has been found to occur even when the two attributes are completely unrelated. Although the halo literature deems such behavior unreasonable, we show consumers are better off when using a halo effect. In what follows we define “better off.” To do so, we turn to the decision-theory literature where the focus is precisely on such definitions. Afterwards we show how consumers benefit by using the halo effect.

**Rational Consumer as Utility Maximizer**

We view the consumer as being rational in the sense that she minimizes her estimation risk. In the decision-theory notation of Robert (1994), given a decision space $D$ and true parameter space $\Theta$, a consumer’s utility function can be expressed as $U(\theta, \delta)$, where $\delta \in D$ and $\theta \in \Theta$. Once the utility function is constructed, the corresponding loss function is simply
negative utility, or $L(\theta, \delta) = -U(\theta, \delta)$. Maximizing utility is therefore equivalent to minimizing loss. Simply to be consistent with the notation of decision theory, we discuss our framework in terms of loss functions rather than utility functions. Note in particular that the consumer—not we as researchers—makes the inferences and therefore incurs the estimation loss. Our model shows that it is in the interest of consumers to act as Bayesians and that the consumer is never worse off by acting as a Bayesian. Or more generally, our model shows it is in a consumer’s interest to use a Stein-like estimator, even if she does not give a Bayesian interpretation to it.

We define loss in terms of the accuracy of a consumer's inference concerning the attributes of a product. A consumer bases her estimate of attribute $i$, $\delta_i(x)$, on product experience or other information about the product, where $x$ is a vector of observed values of a product’s attributes based on experience or product information. The consumer that accurately estimates the value of attribute $i$ ($\delta_i(x) = \theta_i$) incurs no loss, where $\theta_i$ is the true average value of $x_i$. Misperceptions on the part of the attribute lead to loss.

In particular, we assume a squared error loss function, or

1. The loss due to misperception increases with the square of the magnitude of the misperception.
2. Loss is additive across attributes.

Under these assumptions the loss function can be written as

$$L(\theta, \delta(x)) = \sum_i (\delta_i(x) - \theta_i)^2.$$  (2)

Since $x$ is also a random variable, it has a density function $f(x | \theta)$. Estimation risk is defined as the average loss across the observation space,

$$R(\theta, \delta) = E_\theta \left( L(\theta, \delta(x)) \right) = \int_x L(\theta, \delta(x)) f(x | \theta) dx.$$  (3)
Whether squared error loss is appropriate for all types of product attributes it is not initially apparent. Product attributes can be classified into two general categories, an “ideal-point” category and a “more-the-better” category. For ideal-point attributes consumers avoid extreme attribute values. For instance, the amount of sugar in soda is an ideal-point attribute. As an example of a more-the-better attribute, consider the quietness of a dishwasher. The quieter the dishwasher, the higher the utility for consumers.

Squared error loss not only applies to ideal point attributes, but also to more-the-better attributes due to the price-quality tradeoff. Consider a consumer who has paid a price premium for a car with a navigation system or premium music system. After consumption the consumer might realize she paid more for the product than her use of it merits, a loss even for a more-the-better feature.

Also note that because the main purpose of this paper is to investigate the impact of the halo on the consumer’s utility function, we focus on that portion of the utility function that accounts for the halo, which is the estimation loss. The impact of the halo upon consumption utility involves additional assumptions regarding the distributions on individual specific attribute utilities. Since the identification and justification of the halo does not depend on these distributions, we study only estimation loss (judgment tasks and not choice tasks). Estimation loss has been explicitly acknowledged and investigated in multiple research areas (Bawa, Brown, and Klein 1979; Jorion 1986; Casella and Berger 1990; Murthi, Choi, and Desai 1997; Robert 1994). Recent statistical work on James-Stein-type estimators (such as composite estimators [Green and Strawderman 1991]) has shown some extensions to be superior even for an infinite sample size (Kim and White 2001).

Returning to a marketing framework, the consumer’s goal is to choose some decision heuristic \( \delta \) that allows her to translate product information \( x \) into estimates of the true values of
product attributes $\theta$. Decision heuristics that lead to less loss (lower risk, greater utility) are preferred to those that lead to greater loss (higher risk, less utility).

Extant literature on the halo effect proposes that for unrelated product attributes, consumers should solely use information on attribute $i$ when inferring attribute $i$. For instance, if observations on attribute $i$ are normally distributed, $x_i \sim N(\theta_i, 1)$ for $i = 1 \ldots M$, the literature proposes that $\delta(x) = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M)$. What the halo literature observes is that consumers do not use this decision heuristic, but instead treat uncorrelated attributes as if they are related. In other words, a consumer’s actual decision heuristic $\delta(x)$ is a function of not just $x_i$ but of all attributes $x_1, x_2, \ldots, x_M$. We show in what follows that a halo decision heuristic, where the estimate of attribute $i$ is a function of all attributes, leads to lower risk than would be achievable by estimating each attribute independently.

Assume that product attributes $i = 1 \ldots M$ are each normally distributed, $x_i \sim N(\theta_i, 1)$. The consumer observes one draw from each of the $M$ distributions and uses that data to estimate the true level of each attribute, $(\theta_1, \theta_2, \ldots, \theta_M)$. Given that the attributes are not at all related, neither being correlated nor having related means, one plausible estimator would be $\delta^*(x) = (x_1, x_2, \ldots, x_M)$. However, the estimator $\delta^*(x)$ is dominated in terms of risk (in decision theoretic terms, $\delta^*(x)$ is inadmissible for $M>2$) by the estimator $\delta^{JS}(x) = (\delta_1^{JS}(x), \delta_2^{JS}(x), \ldots, \delta_M^{JS}(x))$,

$$\delta_i^{JS}(x) = \mu_i + \left(1 - \frac{M - 2}{V}\right)(x_i - \mu_i)$$ (4)

where

$$V = \sum_{i=1}^{M} (x_i - \mu_i)^2,$$ (5)
$M$ is the number of product attributes, and $\mu_i$ is any initial guess of $\theta_i$. Equations (4) and (5) define the so-called James-Stein estimator. This estimator has been used in fields such as medical sciences, ecology, and finance. For instance, Jorion (1986) showed that this estimator yielded substantial gains in portfolio selection problems. There are other estimators besides the James-Stein (J.S.) that dominate $\delta^*(x)$. Our use of the J.S. estimator is simply as an example of a more general class of estimators that involve some kind of “shrinkage” towards a mean. Although the J.S. estimator is inadmissible itself, we use it in this research to illustrate the intuition of the halo, since it is well known with tractable derivations. In addition, it does not assume consumers to be Bayesians, an assumption required by some alternative shrinkage estimators. Also note that although the James-Stein estimator requires at least three product attributes, Bayesian estimators are defined for any number of attributes and would similarly lead to a halo effect.

Note in particular that a consumer's estimate of the true value of the $i^{th}$ attribute $\delta_i^{JS}(x)$ is a function of all product attributes and that this heuristic leads to lower risk for products with at least 3 attributes; that is, it is preferable to estimating each attribute independently. In other words, the halo effect is due to $V$ (equation 5). Also note that the distributional assumption on the attributes, $x_i \sim N(\theta_i,1)$, can be greatly generalized without changing the main result, and that estimating each attribute as a function of the others dominates independent assessment. In particular, an analogous result can be shown when (1) multiple draws of each attribute can be observed; (2) attributes differ in importance; (3) variance is not 1 but unknown; (4) attribute variances differ; (5) attributes can be correlated; and (6) $x_i$ follows some distribution other than the normal, provided the first and second moments of that distribution exist (Efron and Morris 1973). For instance, the “salient dimension” explanation (Lance, LaPointe, and Stewart 1994) of the halo corresponds to attributes having differential weights; such weights can easily be included in the model. We discuss the case of unit variance simply for ease of exposition. Also, we use
the James-Stein estimator as representative of a larger class of estimators that share its property of treating seemingly independent quantities as dependent.

The key notion here (for the whole class of estimators, although we discuss only the James-Stein estimator) is that people should treat independent attributes as if they are dependent. The intuition behind the result can be presented using an illustration using baseball (or cricket). If the goal is to predict a specific player’s batting average in the upcoming season, the previous year’s batting average for that individual serves as the best estimate. If the goal is to predict batting averages for members of a team for the upcoming season, the previous year’s averages are no longer the best estimates. Rather, estimates for each individual team member should be influenced by the batting averages of the other players on the team (via an estimator like the James-Stein estimator.) Conceptually, the James-Stein estimator capitalizes on the similarity of the data-generating process, namely, that batting averages tend to share a central tendency. If there is high variation among players’ averages, the halo estimator will yield practically the same results as the more intuitive one, last year’s batting averages. In the extreme case of very high variation, the two estimators are identical. It is important to note that the halo estimator on average out-performs the alternative estimator used in the illustration, thereby justifying its use.

In a marketing context household scanner panel data are used to determine price elasticities of individual households. Just as the total panel would be utilized to estimate the individual elasticity of a new household for which no purchase history is observed, the Stein estimator suggests that the total panel would improve estimates for households where purchase histories exist but are limited. The benefits of the James-Stein estimator also apply to theoretical constructs that are not directly measured e.g. IQ (Yung 1999). When predicting the IQ of a student in a class, one can obtain better IQ estimates for that student by using IQ scores of everyone else in the class.
Propositions

From the model above we can derive the following propositions.

Proposition 1: A utility maximizing consumer will use her knowledge about all product attributes to infer the value of individual attributes.

This proposition directly follows from what we have stated above, in that the risk of \( \delta^{JS} (x) \) is strictly less than that of \( \delta^* (x) \) as indicated in equations (6) and (7), and that \( \delta^{JS}_i (x) \) is a function of all attributes and not simply data \( x_i \):

\[
R\left( \delta^{JS}_i (X), \theta_i \right) = \int \sum_{i=1}^{M} \left( \delta^{JS}_i (X) - \theta_i \right)^2 f (X_i | \theta_i) dX_i \\
\leq M - \frac{(M - 2)^2}{M - 2 + \sum (\theta_i - \mu_i)^2} \quad (6) \\
< M
\]

\[
R\left( \delta^*_i (X), \theta_i \right) = \int \sum_{i=1}^{M} \left( \delta^*_i (X) - \theta_i \right)^2 f (X_i | \theta_i) dX_i \\
= M. \quad (7)
\]

For details on the derivation of equations (6) and (7), please see Efron and Morris (1975).

Therefore, by using the halo a consumer actually is able to reduce her estimation risk. The use of the halo is optimal for the utility maximizing (loss minimizing) consumer. Rather than showing existence of a halo, we show a motive, a theory as to why it should exist. Psychologists have had a consensus opinion that the halo effect is not rational and it reflects consumers’ inability to accurately evaluate the attributes of products (Cooper 1981; Feldman 1986). Here, we find that even rational consumers would use the halo.

Proposition 2: The extent of the effect of the halo depends on the level of noise of information across the attributes. When the variance across the attributes gets larger, the effect of the halo becomes weaker. Similarly, the influence of other attributes on the inference on the \( i^{th} \) attribute increases as the variance across the attributes decreases.

This proposition is directly from Equation (4). Since \( M \) must be 3 or greater, \( M-2 \) is at least 1. As \( V \) approaches infinity the James-Stein estimator becomes a function of \( x_i \) alone, so there is no
halo. The implications of equation 6 are similar. The greater the variance of attributes (here measured by dispersion in $\theta_i$), the smaller the gain for using the halo estimator, *ceteris paribus*, because the risk of the James-Stein estimator approaches the risk of the alternative.

*Proposition 3:* The strength of the halo effect also depends on the number of product attributes. The more attributes that are used, the stronger the halo. Similarly, the influence of other attributes on the inference on the $i^{th}$ attribute increases as the number of total attributes increases.

This proposition is also directly from Equation (4). Although we could consider the impact of changes to $M$ holding $V$ constant to show this result, it also would be reasonable in some contexts to hold $V/M$ constant, since $V$ is a summation of $M$ quantities. Consider the weight on $\mu_i$, since a larger weight on $\mu_i$ reflects a greater halo effect. Rewriting $(M-2)/V$ as $\left(1-\frac{2}{M}\right)/c$, where $c=V/M$ is held constant, this quantity converges to $1/c$ as $M$ grows to infinity. For small values of $M$ (such as $M=3$), this quantity is smaller $(1/(3c))$; therefore, smaller values of $M$ yield smaller halo effects. A similar calculation for equation 6 will likewise reveal that the larger the $M$, the greater the gain to using the halo estimator, *ceteris paribus*.

**A General Framework with Information about Overall Quality**

In the above framework we have not considered price, warranties, or other similar information about overall quality. Consider multiple similar products competing in an efficient market, all of which have the same price but are otherwise not identical. In such a category one would expect valued attributes to be negatively correlated. If two cars are identically priced and one dominates on all observed attributes, it is reasonable to infer that the other would excel on the unobserved attributes. If the car that dominates on observed attributes has a price that far exceeds the other, however, one might infer that it dominates on the unobserved attributes as well. Therefore, the assumptions about the price (or indeed any other attributes that are cues of overall quality) are quite important to discuss, since the direction of the expected correlation of attributes depends upon them.
One simple way then to introduce this type of information is

\[ q \sim N\left( \alpha + \beta \sum \theta_i, \tau^2 \right), \]

where \( q \) is observed information on quality, the sum of the \( \theta_i \)'s is the actual total quality\(^3\), and the fact that the expected value of \( q \) is a linear function of quality reflects the assumption that the information about quality and the actual quality are correlated. The noisiness of the quality information \( q \) is captured by the variance \( \tau^2 \).

The task of the consumer is, as before, to infer attribute levels \( \theta_i \) from available data, which now include information on overall quality. We use a Bayesian analysis to solve for the posterior distribution of the \( \theta_i \)'s, recognizing that the James-Stein estimator and a full Bayesian decision theoretic solution will yield similar results. The details of the derivation and posterior equations are provided in the Appendix.

As one would expect, the noisiness of the quality information is of utmost importance. As \( \tau^2 \) becomes small (equation A1), the inference about the unobserved attribute is based upon that difference in the quality indicator and the observed attributes. In other words, there is negative correlation across the observed and unobserved attributes, all else equal. This result shows that when two products are identically priced and one dominates on all observed attributes, it is reasonable to infer that the other would excel on the unobserved attributes.

At the other extreme, when \( \tau^2 \) is large, the situation described by equations 1 through 5 applies. This result can be seen in equation A4, where an empirical Bayes or James-Stein estimator would estimate \( \theta_i \) in a manner consistent with the halo effect. Put another way, one would expect the halo effect to be strongest in situations for which quality information is a noisy indicator of overall product quality.

\(^3\) Note also an implicit assumption about the scaling of the \( \theta_i \)'s, such that the total quality is not a weighted average that allows for different scaling but a simple summation, an assumption we adopt here simply for convenience.
Numerical Example

In this section we provide a numerical example to illustrate the benefits of the James-Stein estimator. We begin by demonstrating that the estimation risk using the halo is lower as compared to not using the halo. We also show that the results hold when the attributes are correlated. Finally, we show that (i) when the number of attributes increases and (ii) when the variance across attributes decreases, the risk difference between using the halo and not using the halo increases.

Risk Comparison of Halo Effect and Non-Halo Effect

Our simulation considers a product with five attributes. The true value of each attribute $i$, $\theta_i$, is a random draw from a normal distribution with mean 1.0 and variance $\tau^2$, where $\tau^2$ ranges from 0.5 to 3.0. Each observation $x_{ij}$ is a random draw from a normal distribution with mean $\theta_i$ and variance 1. The estimator with the halo is calculated using Equation (4) and the non-halo estimator is calculated using the sample mean. Figure 1 shows the risk graphs of the halo and non-halo estimators for different $\tau^2$. This figure shows that the estimation risk of the halo effect is always lower than the non-halo estimator, and the gain increases as $\tau^2$ decreases.

Although our derivations assume independent attributes, the results hold for correlated attributes as well. To show this using a numerical example, we assume two of the five attributes to be correlated with each other, whereas all other attribute pairs are uncorrelated. We vary the degree of correlation of the attribute pair, comparing the risk of the halo estimator with the non-halo estimator. The results are shown in Figures 2 and 3. Figure 2 demonstrates that if two attributes are negatively correlated with each other, the halo estimator always has lower estimation risk compared to the non-halo estimator. The estimation risk of the halo estimator remains the same with different degrees of negative correlations. Figure 3 demonstrates the risk comparison of halo and non-halo estimators when two attributes are positively correlated. The
risk of the halo estimator is always lower than the non-halo estimator and the difference increases when the degree of correlation decreases.

**Risk Comparison for Different Variance and Number of Attributes**

Figure 4 shows the risk difference between our halo estimator and non-halo estimator varies with the number of attributes and with the variance across attributes. From Figure 4, which shows the risk difference for products with 6 and 12 uncorrelated attributes, we see that when the number of attributes increases, the risk difference increases. Also, as the variance decreases the risk difference increases.

As stated in the propositions, the effect of the halo will strengthen when the number of attributes increases or when the variance across attributes decreases. More precisely, there is an increasing gain in using the halo estimator relative to the non-halo estimator when the number of attributes increases or when the variance across attributes decreases.

**Experiments**

Consider the case where consumers receive information about product attributes in an advertisement. Given the same cost of a campaign, firms have a choice of either communicating information on a few (or single) attributes repeatedly or exposing consumers to multiple attribute information. Also assume that consumers, prior estimate of attribute values is that the attribute is average, whereas advertised information presents attributes as above average.

The implications of Propositions 2 and 3 lead to the following two hypotheses:

- \( H_1: \) Multi-attribute information with small variance leads to a stronger halo than does exposure to multiple attributes with high variance.
- \( H_2: \) Multi-attribute information with small variance leads to a stronger halo than does exposure to single-attribute information.

Experiment 1
The purpose of this experiment is to investigate how consumers’ inferences differ when the consumers are exposed to multi-attribute information where the variance differs. According to Hypothesis 1 an inference based on information from multi-attributes with low variance will lead to more positive inferences relative to information from multiple-attributes where the variance is high. Hypothesis 1 is tested using a single-factor between-subjects design with three levels.

**Subjects and Procedure**

One hundred and one subjects at a major eastern university participated in the experiment in order to obtain class credit. They were told that the objective of the study was to investigate how consumers evaluate products. The subject was asked to read a report about a hypothetical product. In the report, ratings of five attributes labeled Attribute A to Attribute E were provided using a 100-point scale. The subjects were told that they had to infer the value of an attribute “F”. To ensure that the subjects did not focus on any specific attribute, they were instructed that the attribute F was not correlated to any of the other attributes. The three conditions were low variance, medium variance, and high variance. The values of the attributes were constructed so that the extreme values were identical across the three conditions. Thus, in the low-variance condition, the attribute values were \{60, 80, 80, 85, 95\}; in the medium-variance condition, the values were \{60, 75, 84, 86, 95\}; and in the high-variance condition were \{60, 70, 85, 90, 95\}. The mean in all conditions is 80 and the upper and lower values are anchored identically by 65 and 90. We use the same upper and lower values to control for alternative process that may be used for making inferences such as anchoring and adjustment. At the end of the task, two questions assessed whether the subjects perceived attribute F to be correlated to any of the attributes. Subjects were asked to indicate if Attribute F was “not correlated at all/highly correlated” or “not related at all/highly related” to other attributes.
Results

We first checked whether subjects had used any of the given information on a specific attribute to make an inference on the missing attribute. The responses on the correlation question indicated that the subjects had not.

**Dependent Measure:** The inferred attribute F was measured using a 100-point scale. The mean and standard deviations of the dependent measure at these three conditions are reported in Table 1. According to Hypothesis 1 consumers will infer higher values on a missing attribute when exposed to multi-attribute low variance information as compared to multi-attribute information with high-variance. As predicted, the inferences are much lower in the high-variance condition than in the low-variance condition \(F_{1,92} = 7.92\ p < .001\). Also, the inferences are significantly lower in the medium-variance condition than in the low-variance condition \(F_{1,92} = 2.85\ p < .05\). The results of this experiment indicate that after controlling for other potential decision processes such as anchoring and adjustment biases, Hypothesis 1 is supported.

Experiment 2

The objective of Experiment 2 is to test Hypothesis 2. Hypothesis 2 states that multi-attribute information will lead to more positive beliefs about an inferred attribute than an inference based on information about a single attribute. This effect only would appear when the variance is small in the multi-attribute case, i.e. when the halo would be stronger, for there are opposing effects. Holding variance across attributes constant, more attributes strengthen the halo. Holding the number of attributes constant, lower variance results strengthens the halo. So relative to the single attribute condition, which has no variance, the results for the multi-attribute conditions depend on the degree of variance.

Sixty-four students participated in the study. We use airlines as the product category for this experiment. Subjects were told that the objective of the study was to investigate how
consumers evaluate airline services. The four attributes selected were based on several surveys that have identified them as important: check-in efficiency, baggage handling, flight-attendant service, and departure/arrival record. Check-in efficiency, flight-attendant service, and baggage handling were used as the observed attributes. The subjects were asked to make inferences about departure/arrival record.

In the single-attribute condition, only the rating of 72 for flight-attendant service was provided. There were two multi-attribute conditions: in the low variance condition, the three ratings for check-in efficiency, baggage handling, and flight-attendant service were 72, 80, and 64 respectively. In the high variance multi-attribute condition, the ratings for check-in efficiency, baggage handling, and flight-attendant service were 96, 56, and 64 respectively. The mean ratings across the three attributes were 72 in each condition. After reading the report subjects were asked to provide their predicted ratings for the departure/arrival record of Alpha Airlines. Subjects’ expertise about airline service was also recorded with two self-reported measure using 9-point scales ($\alpha = .90$).

Results

**Dependent Measure.** The inferred departure/arrival record was measured using a 100-point scale. The mean and standard deviations of the dependent measure at these three conditions are reported in Table 2. According to Hypothesis 2 the inferred value of an attribute will be higher when consumers are exposed to multi-attribute information with low variance. Thus, subjects will make more positive inferences about an unknown attribute after receiving information on multiple attributes compared to when they are provided with information about only one attribute, provided the multiple attribute information has low variance and the two sets of attributes (single and multiple) have the same mean. The results of the experiment show that, as expected, the inferred missing attribute is significantly higher in the multi-attribute low-variance condition than in the single-attribute condition ($F_{1,60} = 4.96, p < .05$). We have no a
priori hypothesis on the difference in inference between the single-attribute and high-variance condition. The results show that the single-attribute condition leads to significantly more positive inferences than the multi-attribute high-variance attributes ($F_{1,60} = 8.26\; p < .01$). The variance in the stimuli appears to be high enough to weaken the halo.

**Conclusion**

A long-held view in the psychology literature has been that the use of the halo to form evaluations results in misjudgments. The objective of this paper was to demonstrate that the use of the halo, in fact, results in the reduction of estimation risk as compared to not using the halo. Thus, it is rational for a utility-maximizing consumer to use the halo when faced with a situation where information is either incomplete or difficult to judge. The improvement through the use of the halo in forming evaluations of missing or imprecise information increases when the number of observed attributes increases. The use of the halo is also more effective when the variance across the observed attributes is relatively low as compared to when the variance is relatively high.

Although the halo has been demonstrated several times, our experiment shows that the halo exists even when subjects are informed that the unobserved and the observed are independent. The experiment also provides evidence that the use of multiple-attribute information is more effective than the use of single-attribute information, but only when the variation across attributes is relatively low. Also, the use of low-variance multi-attribute information leads to more positive inferences than when the variance across multi-attribute information is high. The results from the numerical example also indicate that the use of halo results in lower estimation risk when there is either positive or negative correlation between attribute levels.
We have made several assumptions in our analysis that temper the results. For instance, we ignore any potential costs of acquiring information. We assume that information is exogenously available, that consumers make their choices given the available information. By necessity, we assume non-trivial costs of acquiring additional information on the “true” value of an attribute, costs which apply to experience and credence goods. If information on the true value were costless, and consumers were easily able to judge the true value, there would be no estimation risk and thus no halo effect. Also, the results apply only to products for which consumers evaluate at least three attributes. Furthermore, although consumers are never worse off by using the halo effect relative to assuming independence of attributes, the halo yields the greatest benefit for products with high variance across attributes.

These results have important implications for researchers investigating decision processes and also for marketers designing communication strategies. Prior research on advertising strategies has suggested that a cosmetic variation strategy (repetition of the same attribute with minor creative variations) is more effective when motivation to process the advertisement is low as compared to when it is high but that a substantive variation strategy (using multiple attributes) results in more positive attitudes when the motivation to process the advertisement is high as compared to when it is low. Haugtvedt and colleagues (1994) find that there were no differences in attitudes when a multiple-attribute versus a repeated single-attribute strategy was used. Although this paper does not use advertisements or examine repetition, the results have broad implications for research in advertising. The implication from this analysis is that the use of more attributes in an advertisement campaign would be more effective than using fewer attributes repeatedly. This implication only holds if the consumers perceive all attributes used for the brand advertising as important and if the attribute claims are believed. If the creative strategy used in the campaign is unable to persuade consumers to perceive any specific attribute positively then
the campaign is not likely to be effective. In such situations it may be more effective to use fewer attributes or only those attributes consumers are likely to recognize favorably.
References


Huber, Joel, and John McCann, 1982 “The Impact of Inferential Beliefs on Product Evaluations” Journal of Marketing Research; 324-333.


Kim, Tae-Hwan and Halbert White, 2001 “James-Stein-Type Estimators in Large Samples with Application to the Least Absolute Deviations Estimator” Journal of the American Statistical Association 96 (454); 697-705.


Yung, Kwong Hiu 1999 “Explaining the Stein Paradox” unpublished.
Figure 1. Risk Comparison of Halo and Non-Halo Estimators

![Graph showing the risk comparison between Halo and Non-Halo Estimators for different variance values.]

Figure 2. Risk Comparison of Non-halo and Halo Estimators for Negatively Correlated Attributes

![Graph showing the risk comparison between Non-halo and Halo Estimators for different correlation values.]
Figure 3. Risk Comparison of Non-halo and Halo Estimator for Positive Correlated Attributes

Figure 4. Risk Difference under Different Variance $\tau^2$ and Number of Attributes
Table 1: Experiment 1 Results

<table>
<thead>
<tr>
<th></th>
<th>Multiple-Attribute Low Variance</th>
<th>Multiple-Attribute Medium Variance</th>
<th>Multiple-Attribute High Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferred Ratings</td>
<td>76.84 (6.73)</td>
<td>72.19 (12.70)</td>
<td>69.21 (11.99)</td>
</tr>
</tbody>
</table>

Table 2: Experiment 2 Results

<table>
<thead>
<tr>
<th></th>
<th>Multiple-Attribute Low-Variance</th>
<th>Single-Attribute</th>
<th>Multiple-Attribute High-Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferred Rating</td>
<td>73.87 (10.39)</td>
<td>66.09 (9.18)</td>
<td>54.84 (15.58)</td>
</tr>
</tbody>
</table>
Appendix

Prior
Let \( x_{ij} \) represent the \( j^{th} \) observation of attribute \( i \), and let \( \theta_i \) represent the true (unobserved) value of attribute \( i \). Also assume \( x_{ij} \sim N(\theta_i, \sigma^2) \).

For simplicity, we assume that the variances are the same across attributes. A product has \( m+1 \) attributes. We assume the priors for \( \theta_i, \ i = 1, \ldots, m+1 \) where \( \theta_i \sim N(\bar{\theta}, v^2_\theta) \). Or equivalently,

\[
\bar{\theta} | \bar{\theta}, v^2_\theta \sim N(A_2 \bar{\theta}, C_2),
\]

where \( \bar{\theta} \) is defined as \( (\theta_1, \theta_2, \ldots, \theta_{m+1})^\prime \),

\[
A_2 = \begin{pmatrix} 
1 \\
1 \\
\vdots \\
1
\end{pmatrix},
\]

and

\[
C_2 = I_{(m+1) \times (m+1)} v^2_\theta.
\]

Likelihood function
Let the overall quality indicator \( q \) be distributed

\[
q \sim N(\alpha + \beta \sum \theta_i, \tau^2).
\]

Assume that no information on the first attribute is observed, as before.

Let

\[
y = \begin{pmatrix} 
q - \alpha \\
x_2 \\
\vdots \\
x_{m+1}
\end{pmatrix}
\]
The likelihood for \((q, x_2, ..., x_{m+1})\) is

\[
L(y|\theta) = f(q-\alpha, x_2, ..., x_{m+1}|\theta),
\]

And \((q-\alpha, x_2, ..., x_{m+1}|\theta)\) follows multivariate normal distribution.

Therefore

\[
y \sim N(A_1 \theta, C_1),
\]

where

\[
A_1 = \begin{pmatrix}
\beta & \ldots & \beta \\
0 & 1 \\
1
\end{pmatrix},
\]

and

\[
C_1 = \begin{pmatrix}
\sigma^2
\end{pmatrix}
\]

Posterior

\[
p(\theta|y) \propto p(\theta) \cdot L(y|\theta) 
\propto \text{Normal} \times \text{Normal}
\]

\[
p(\theta|y) \sim N(Dd, D)
\]

where

\[
D^{-1} = A_1 C_1^{-1} A_1 + C_2^{-1}
\]

and

\[
d = A_1 C_1^{-1} y + C_2^{-1} A_2 \theta.
\]
To put some of these matrices back into original parameters:

\[
A_i C_{i-1} = \begin{pmatrix}
\beta & 0 \\
\beta & 1 \\
\vdots & \ddots \\
\beta & 1
\end{pmatrix}
\begin{pmatrix}
1/\tau^2 \\
1/\sigma^2 \\
\vdots \\
1/\sigma^2
\end{pmatrix}
\begin{pmatrix}
\frac{\beta}{\tau^2} & 0 & \cdots & 0 \\
\frac{\beta}{\tau^2} & \frac{1}{\sigma^2} & \cdots & 0 \\
\frac{\beta}{\tau^2} & 0 & \frac{1}{\sigma^2} \\
\frac{\beta}{\tau^2} & 0 & 0 & \frac{1}{\sigma^2}
\end{pmatrix}
\]

And

\[
A_i C_{i-1} A + C_{i-2} = \begin{pmatrix}
\frac{\beta^2}{\tau^2} & \frac{\beta^2}{\tau^2} & \cdots & \frac{\beta^2}{\tau^2} \\
\frac{\beta^2}{\tau^2} & \frac{\beta^2}{\tau^2} + \frac{1}{\sigma^2} & \cdots & \frac{\beta^2}{\tau^2} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{\beta^2}{\tau^2} & \cdots & \frac{\beta^2}{\tau^2} & \frac{\beta^2}{\tau^2} + \frac{1}{\sigma^2}
\end{pmatrix}
\begin{pmatrix}
1/\nu_0^2 \\
1/\nu_0^2 \\
\vdots \\
1/\nu_0^2
\end{pmatrix}
\begin{pmatrix}
\frac{\beta^2}{\tau^2} + \frac{1}{\nu_0^2} & \frac{\beta^2}{\tau^2} & \cdots & \frac{\beta^2}{\tau^2} \\
\frac{\beta^2}{\tau^2} & \frac{\beta^2}{\tau^2} + \frac{1}{\sigma^2} + \frac{1}{\nu_0^2} & \cdots & \frac{\beta^2}{\tau^2} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{\beta^2}{\tau^2} & \cdots & \frac{\beta^2}{\tau^2} & \frac{\beta^2}{\tau^2} + \frac{1}{\sigma^2} + \frac{1}{\nu_0^2}
\end{pmatrix}
\]
As a special case, suppose there are 3 attributes, or \( m = 2 \), then

\[
D = \left( A_1 C^{-1} A + C_2^{-1} \right)^{-1}
\]

\[
= \begin{pmatrix}
\frac{\beta}{\tau^2} / (1 + 1/\nu^2_\theta) & \frac{\beta}{\tau^2} & \frac{\beta}{\tau^2} \\
\frac{\beta}{\tau^2} & \frac{\beta}{\tau^2 + 1/\sigma^2 + 1/\nu^2_\theta} & \frac{\beta}{\tau^2} \\
\frac{\beta}{\tau^2} & \frac{\beta}{\tau^2} & \frac{\beta}{\tau^2 + 1/\sigma^2 + 1/\nu^2_\theta}
\end{pmatrix}
\]

\[
d = \begin{pmatrix}
\beta / \tau^2 & 0 & 0 \\
\beta / \tau^2 & 1/\sigma^2 & 0 \\
\beta / \tau^2 & 0 & 1/\sigma^2
\end{pmatrix}
\begin{pmatrix}
q - \alpha \\
x_2 \\
x_3
\end{pmatrix}
+ \begin{pmatrix}
1/\nu^2_\theta \\
1/\nu^2_\theta \\
1/\nu^2_\theta
\end{pmatrix}
\begin{pmatrix}
1 \\
1 \\
1 \mathrm{(m+1)x1}
\end{pmatrix}
\]

\[
Dd = \begin{pmatrix}
\frac{v^\beta(\alpha - (x_2 + x_3)) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2) (v_\beta + \sigma^2)^3} \\
\frac{v^\beta(\alpha - (x_2 + x_3)) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2) (v_\beta + \sigma^2)^3} \\
\frac{v^\beta(\alpha - (x_2 + x_3)) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2) (v_\beta + \sigma^2)^3}
\end{pmatrix}
\]

As \( \tau \to 0 \), the posterior mean is

\[
\begin{pmatrix}
\frac{v^\beta(q - \alpha) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2)} \\
\frac{v^\beta(q - \alpha) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2)} \\
\frac{v^\beta(q - \alpha) / \beta^2 + \beta (q - \alpha) / \beta (v_\beta + \sigma^2)}{v_\beta^2 (v_\beta + 3\sigma^2)}
\end{pmatrix}, \quad (A1)
\]

all the items with \( \bar{\theta} \) disappear.
As \( v_\theta \to 0 \), then the posterior mean is
\[
\begin{pmatrix}
  \theta \\
  \bar{\theta} \\
  \bar{\theta}
\end{pmatrix}.
\]  

(A2)

As \( \sigma \to 0 \), then the posterior mean is
\[
\begin{pmatrix}
  \frac{v_\theta \beta (\theta - \alpha - (x_2 + x_3) \beta) + \tau^2}{v_\theta \beta^2 + \tau^2} \\
  x_2 \\
  x_3
\end{pmatrix}.
\]

(A3)

As \( \sigma \to 0 \) and \( \tau \to \infty \), then the posterior mean is
\[
\begin{pmatrix}
  \bar{\theta} \\
  x_2 \\
  x_3
\end{pmatrix}.
\]

(A4)

In this last equation, \( x_2 \) and \( x_3 \) would offer information about \( \bar{\theta} \) in an empirical Bayes context analogous to the James-Stein estimator.