The Cost of Thinking

Abstract

Because consumer research often faces the finite or quantal choice problem, a recent study developed a theory of choice that explicitly considers the difficulty in comparing diverse alternatives. The theory's objective was to offer a methodology for explicitly dealing with 'thinking costs' that uses both the notions of preferences over characteristics and probabilistic predictions of choice. A second objective was to examine the precise cost of using various simplifying decision rules as compared to a utility maximizing procedure, allowing a theoretical comparison of various simplifying strategies on a cost basis. The resulting model does have some restrictive assumptions. However, it is able to quantify decision-making costs and the likelihood of mistakes. The model may be applied in the areas of: 1. Product characteristics in an advertisement, 2. Information presentation, and 3. New product sales forecasting. However, if used on an individual level, the adoption of simplifying choice strategies can leave a consumer vulnerable to manipulation. For example, due to a lack of proper feedback of information, a consumer is kept from learning of mistakes, and the best product may never be found.
The Cost Of Thinking

STEVEN M. SHUGAN*

A theory and methodology are developed for explicitly considering the cost of comparing diverse choice alternatives. The theory allows (1) explicit analytical measures of the cost of using various simplified decision strategies, and (2) predictions regarding the distribution of mistakes a consumer is likely to make when reducing decision-making effort.

To the vast majority of mankind nothing is more agreeable than to escape the need for mental exertion. . . . To most people nothing is more troublesome than the effort of thinking (James Bryce, The American Commonwealth 1888).

The finite or quantal choice problem frequently occurs in consumer research (Bettman 1971; Blattberg and Sen 1976; Einhorn 1970; Fishbein and Ajzen 1972; Luce 1959; Marschak and Radner 1972; McFadden 1970; Tversky 1972; Tversky and Kahneman 1979). A consumer or decision maker faces a choice conflict in which the individual must select a choice from some set of alternatives (products, brands or generally choice objects). The consumer, after choosing one of the alternatives or products, derives satisfaction from the product represented by the product's utility (Farquhar 1977; Green and Wind 1973; Herstein and Milnor 1953) or affect (Fishbein and Ajzen 1972). Naturally, many theorists began by assuming individuals would choose their most preferred (optimal) product, thereby maximizing their utility. However, this approach ignored measurement errors and lacked insight for situations in which new alternatives are offered or old alternatives deleted.

Attempts to incorporate nonoptimal alternatives focused on probabilistic predictions of choice. For example, Luce's axiom (Luce 1959) or Clarke's rule (Clarke 1957) and its extensions (Morgan 1974) propose a mechanism where the probability of any product being chosen is a function not only of product preference, but also of the utilities of the nonoptimal products. Marketers began using information from the entire set of products to estimate choice probabilities. McFadden (1970) used statistical estimation implying Luce's assumption (for product addition and deletion) to determine underlying consumer preferences from choice data. Later, Hauser (1976) showed this probabilistic approach to be consistent with deterministic axioms of preference.

Unfortunately, the influence of nonoptimal alternatives is somewhat arbitrary. The actual consumer choice mechanism is not considered. Therefore, it is not difficult to construct choice situations that are inconsistent with Luce's axiom (Becker, DeGroot, and Marschak 1963; Debreu 1960; Tversky and Russo 1969).

One problem with Luce's axiom is its lack of consideration of product differences and similarities. In marketing terms, Luce's axiom implies that when a new brand is introduced, it derives its market share proportionally from all other brands regardless of substitutability. This deficiency was remedied by viewing preferences for characteristics as fundamental (Fishbein and Ajzen 1972; Keeney and Raiffa 1976; Lancaster 1966) rather than preferences for products. Tversky (1972) brilliantly combined the notion of characteristics (albeit binary) with a Luce-type mechanism.

The theory was now sufficiently rich to address problems of new products, deletion of products, and changing preferences for existing products. However, empirical examinations (Bass 1974; Bass, Pessemier, and Lehmann 1972; Hayes 1964; Payne 1976) showed that behavior was far more complex. When faced with a choice conflict, consumer perceptions were formed

---

* Steven M. Shugan is Assistant Professor of Marketing, Graduate School of Business, University of Chicago, Chicago, IL 60637. Parts of this research were conducted at the University of Rochester. The author wishes to thank Donald Lehmann, Dov Pekelman, Subrata Sen, John Hauser, and the anonymous reviewers for their many helpful comments.

---

1 The quantal choice, or "all-or-nothing response," refers to problems where responses can be expressed as "occurring" or "not occurring," e.g., whether an insect is dead or living. Problems such as "how much will he buy?" are not quantal choice problems. A large literature exists on statistical techniques for dealing with quantal response data.
by acquiring information on each product and then processing that information to arrive at an expected utility. Preference only partially influences choice by determining benefits. However, the determination has costs—rife information, numerous alternatives, time pressure, the consumer's limited information processing capabilities, and the general effort exerted to solve the problem. Generally, the net utility of finding the best product from one set of products may be different from the net utility of finding it as best from another set of products. That is, there may be a cost associated with the act of making a decision—the "cost of thinking."

The study of decision-making or thinking costs has been well accepted by many researchers (Coombs 1964; Dawes 1964; Simon 1957; Simon and Newell 1971) who have studied specific simplifying rules for processing information that purport to lower decision-making costs (Bettman 1977; Bettman and Kukkar 1977; Einhorn 1971; Slovic, Fischhoff, and Lichtenstein 1970; Wright and Barbour 1977) and their applications to marketing (Lehmann 1977; Russo 1977; Wright 1975). These rules often search for a "satisfactory" alternative rather than an optimal one, and, hence, the process is referred to as "satisficing" (Simon 1957). This experimental research has provided substantial understanding of consumer behavior. However, the costs associated with these rules have yet to be rigorously defined and measured. To adequately understand, model, predict, and possibly influence the consumer choice process, we must build a theoretical foundation for "thinking costs." We must quantify thinking costs, determine a unit of measurement, and explore how that measurement varies across choice conflicts.

This paper develops a theory of choice that explicitly considers the difficulty in comparing diverse alternatives. The objective is to provide a methodology and development for explicitly dealing with "thinking costs" that use both the notions of preferences over characteristics and probabilistic predictions of choice. Specifically, one objective is to define a measurable (i.e., well-defined and calculable) unit of thought, and then use it to quantify and estimate the cost of utility maximization. Another objective is to examine the precise cost of using various simplifying decision rules as compared to a utility maximizing procedure, allowing a theoretical comparison of various simplifying strategies on a cost basis.

Of course, simplifying decision rules may lead to less than optimal alternatives, which could be called mistakes. An important objective of this paper is to examine these mistakes and how a reduction in thinking costs often leads to a reduction in expected benefits. It will be shown that under certain specific conditions Luce's axiom describes these mistakes, and under more general conditions Tversky's mechanism describes them.

BRIEF REVIEW OF CHOICE THEORY

Strategies that intend to save decision-making costs by simplifying the choice process include conjunctive, disjunctive (or maximax), minimax, and lexicographic strategies, all generally referred to as noncompensatory. That is, one characteristic cannot compensate for a deficiency in another. The conjunctive rule states: any product not meeting a minimum cutoff level on any characteristic is eliminated. The Federal Drug Administration uses a conjunctive strategy in issuing standards (e.g., purity, weight, age) that all ethical drugs must meet. Wright (1975) cites a variation of this strategy, i.e., choosing the product that meets any of the cutoffs.

The disjunctive strategy or rule is a maximax strategy. Products are compared on their best characteristic. The product with the highest rating on its best characteristic is chosen. The minimax strategy suggests products should be judged on their weakest characteristic, and the one with the strongest weakest characteristic should be selected. For example, an electric circuit may be chosen on the basis of its weakest component because once that component fails, the circuit fails. Finally, the lexicographic strategy first ranks the characteristics in order of importance and then selects the product rated best on the most important characteristic. If two or more products rate equally, the next most important characteristic is used as a tie breaker. For example, the winner of a chess tournament is the person winning the most games. However, if two or more people tie on this criterion, a measure of the quality of the opponents may be used as a tie breaker. The lexicographic rule proper will not be dealt with in detail, because a generalization of the Tversky mechanism contains the essential lexicographic elements and represents a much richer model of the consumer decision process.

Note that each of the preceding examples of simplifying strategy usage was justified on the outcome it provided rather than the savings of decision-making costs. The strategies, in fact, determined the best product. It is much more difficult to compare the strategies on the ease of their use than on their potential to select the best alternatives. The first step involves experimentation on the relative ease of adopting various simplifying decision rules. However, a theoretical framework is needed to formally compare the potential savings in decision-making costs for the various strategies.

THE CONFUSION INDEX

Development of a "Thinking Cost"

Let us start by considering the cost of a utility-maximizing model. It will then be possible to determine how simplifying decision rules reduce that cost. Sup-
pose a consumer wishes to choose the best (most preferred) product from several products. For the moment, assume this decision is occurring for the first time. If there are $M$ alternative products under consideration, the consumer must make $M - 1$ comparisons to determine the most preferred, i.e., eliminate $M - 1$ products. For example, to discover Lysol is the best of four household cleaners, the consumer could compare Lysol with the other three, making $M - 1$ comparisons. If there were a fixed cost, $f$, per comparison, the total cost of the decision could be computed by multiplying that cost by $M - 1$. Precisely,

$$\text{difficulty of choice} = (M - 1)f,$$  \hspace{1cm} (1)

where,

- $f$ = the cost of comparing two products, and
- $M$ = the number of alternative products considered.

Equation 1 can be written as,

$$\text{general difficulty of a choice} = mf^r,$$  \hspace{1cm} (2)

where,

- $m$ = the number of product comparisons, and
- $f^r$ = the average difficulty or cost of comparing two products.

Equation 2 provides a method for computing “thinking costs.” However, some products may be harder to compare with each other than other pairs of products. Therefore, a method for computing the cost of comparing two products, $f$, is required.

The Difficulty of Comparing Two Products

Assume the consumer's preferences are determined (directly or as a cue) by the product's characteristics. Then, a consumer who wishes to choose between two products may proceed by comparing the two products on their characteristics. These comparisons may be viewed as aspect (Tversky 1972), or attribute, comparisons. For example, a consumer choosing between two household cleaning products may first compare them on ammonia content. Second, a comparison on drying speed may take place. Next, the products could be compared on the attractiveness of their respective colors. These comparisons could then proceed until all characteristics of the product are exhausted, uniquely defining each product. The two products may be compared on a multitude of characteristics before a choice is made.

Assume that associated with each characteristic comparison is a fixed cost, a unit of comparison effort. The products must be evaluated on the characteristic and their differences assessed. It is then reasonable to assume that the more comparisons necessary to make a choice, the more difficult the choice. If the choice can be made after comparing the products on one characteristic only, the choice is relatively easy. If, however, several hundred characteristic comparisons are required, the choice can be considered relatively difficult. This paper measures the thinking cost associated with a choice by positing that $f$ is monotonically related to the number of characteristic comparisons made. That is, more difficult decisions require more characteristic comparisons.

This representation of the consumer choice process would be void of implications without a methodology to classify choice situations with respect to the number of comparisons necessary to resolve the conflict. Fortunately, this representation can be interpreted as a sampling problem. The consumer can be viewed as sampling product pair differences by characteristic. For example, consider the household cleaning product comparison. The consumer first compares the products on ammonia content. This comparison is basically sampling from the population of product differences. The sample chosen has one observation—difference in ammonia content. Again, the second comparison on drying speed can be viewed as an observation on drying speed difference. The sample now contains a third observation—color attractiveness difference. At each point in the sampling, the consumer can infer the true difference (i.e., the true preference) between the products. Given a positive inferred difference, the former product will be chosen. Given a negative inferred difference, the latter product will be chosen. Finally, given a difference close to zero, the consumer will remain uncertain and continue sampling, comparing the products on yet another characteristic.

The crucial question becomes: How many product difference comparisons need to be made so that the consumer will feel sufficiently confident to make a decision, i.e., choose the product judged superior on characteristics observed thus far? This number determines $f$, and, hence, the difficulty of the choice. Given some fairly unrestrictive assumptions (DeGroot 1970), sampling theory would dictate the following three factors as influencing the expected number of characteristic comparisons necessary to make the choice. If $z_r$ is the difference in utility between the products on attribute $r$, the three factors are:

1. The true difference in mean utility (average relative preference) between the two products. This is the expected value of $z_r$, $r$ probabilistically chosen, denoted $E(z_r)$.
2. The confidence level at which the decision must be made, denoted $\alpha$. This value is the probability of not making a mistake. \hspace{1cm} \text{\footnote{The $\alpha$ level has a strong relationship to the psychological theory of involvement.}}
3. The variability in the characteristic difference between the two products. This is the variance of $z_r$, $r$ probabilistically chosen, denoted $\text{var}(z_r)$.

$$z_r = x_{1r} - x_{2r},$$

where $x_{1r}$ is the characteristic of product 1 on attribute $r$, and $x_{2r}$ is the characteristic of product 2 on attribute $r$. The sample chosen has one observation—difference in mean utility between the two products. This is the variance of $z_r$, $r$ probabilistically chosen, denoted $\text{var}(z_r)$. \hspace{1cm} \text{\footnote{The $\alpha$ level has a strong relationship to the psychological theory of involvement.}}
The first factor is inversely related to the difficulty of the choice. If the true difference in utility between the two products is large, holding factors 2 and 3 constant, few characteristic comparisons will be required and the choice is easy. However, if the true difference in utility is small, many comparisons will be required to determine this small difference, hence the choice is more difficult. Hendrick, Mills, and Kiesler (1968) appear to have contradictory experimental results; however, other factors were not kept constant.

The second factor is directly related to the difficulty of the choice. The consumer may infer, at any time, which is the more preferred product and then choose that product. But the consumer would then be choosing a product before all possible characteristics have been compared. Hence, the consumer risks a mistake. Requiring more confidence implies a lower acceptable risk, which requires more comparisons and, hence, a more difficult decision. For example, in choosing sticks of gum, the consequence of making a mistake is relatively unimportant. Here, the consumer may only consider one product characteristic for comparison, perhaps brand name. However, in choosing a house, the consequence of a mistake may be very costly and more comparisons are required.

Note that $\alpha$ is exogenous to the model. The confidence level, $\alpha$, reflects how this choice interacts with other decisions and the expected difficulty of the decision at hand. The resources allocated, including thinking effort, to any choice will depend on opportunities made available by other choices. Here, $\alpha$ reflects and captures the effect of all outside choices on the choice at hand. Future research using Bayesian analysis may specify a loss function based on some global optimization.

The third and final factor is inversely related to the difficulty of the choice. As the variability in product characteristic differences increases (actually, the differences in utility), holding average relative preference constant, the number of comparisons necessary to make a choice at a given confidence level increases. This increase, in turn, heightens the difficulty of the decision. For example, in choosing two household cleaning products, a consumer would find the comparison relatively easy if one product uniformly dominated the other product on all characteristics (color, amount of suds, abrasiveness, etc.), that is, zero characteristic difference variability. Conversely, the consumer would find the comparison relatively difficult if the two household cleaners were not only very different on all characteristics, but also superior on an equal number of characteristics.

The preceding discussion can be formalized for preciseness, as follows:

$$N = \text{number of characteristics (e.g., ammonia content or color attractiveness for household cleaners) that uniquely identify the choice alternatives.}$$

$$X_{ij} = \text{the level of the } i\text{th characteristic for choice alternative } j \text{ (e.g., the ammonia content for Windex).}$$

$$U_i(X_j) = \text{satisfaction or utility derived from the } i\text{th characteristic for alternative } j \text{ abbreviated } U_j.$$  

$$U_i = \text{actual satisfaction or utility derived from the selection of alternative } j.$$  

Further, for simplicity, assume$^3$

$$U_i = \sum_{i=1}^{N} U_{ij}.$$  

Suppose a consumer must choose between product $j$ and product $k$ (for example, between Lysol and Windex). The consumer proceeds to compare the products on a series of characteristics. For each characteristic $r$, the utilities for both products, $U_{ij}$ and $U_{ik}$, respectively, are observed and the difference, $z_r = U_{ij} - U_{ik}$, is obtained. If $n$ characteristics are examined, the consumer will choose product $j$ if $\sum_{r=1}^{n} z_r > 0$ and product $k$ if $\sum_{r=1}^{n} z_r < 0$. (The case where $\sum_{r=1}^{n} z_r = 0$ will be discussed later.)

The consumer must, then, choose the number of characteristics to observe (that is, $n$).$^4$ This $n$ will depend on the willingness of the consumer to make a mistake, which is represented by $\alpha$. The consumer requires $n$ to be large enough so that the probability of not making a mistake is less than $\alpha$. Precisely, the consumer requires both:

Condition A: $P(z_n > 0|U_j - U_k < 0) < 1 - \alpha$, and

Condition B: $P(z_n < 0|U_j - U_k > 0) < 1 - \alpha$, where $P(\cdot | \cdot)$ is a conditional probability function. $z_n$ is the sample mean with sample size $n$, i.e., $(1/n) \sum_{i=1}^{n} z_n$, and $\alpha$ is the confidence level ($0 < \alpha < 1$).

To minimize thinking cost, the consumer will select the minimum number of characteristic comparisons, $n^*$, so that the confidence level is maintained. Precisely,

$$n^* = \text{minimum } n \text{ so that Conditions A and B hold.}$$

As stated earlier, $n^*$ is a function of $E(z)$, $\text{var}(z)$, and $\alpha$. Now, the variance of $z$ can be interpreted as the perceptual difficulty in comparing the two products, which can be analyzed by breaking the $\text{var}(z)$ into its

---

$^3$ This assumption does not require a linear utility function, only additive separability over characteristics (see Farquhar 1977). Note that the assumption only concerns consumer preferences, and makes no direct assumption about information processing.

$^4$ Sequential sampling simply requires the use of standard and well-developed dynamic programming (for example, see Blackwell 1965 or Wetherill 1975) with special emphasis on optimal stopping (DeGroot 1970). However, the mathematics would soon become quite tedious and might obscure the insights of the subsequent development.
COST OF THINKING

FIGURE A
POSITIVE COVARIANCE — A SIMPLE CHOICE

\[ \text{cov} \left[ U_A, U_B \right] = 1.56 \]

![Diagram](utility-comparison_A_B)

\[ \text{Utility} \]

Components. Hence,

\[ \text{var}(z) = \text{var}(U_j - U_k), \]

\[ = \text{var}(U_j) + \text{var}(U_k) - 2\text{cov}(U_j, U_k), \]

where \( \text{cov}(U_j, U_k) \) is the covariance between \( U_j \) and \( U_k \).

The first term, \( \text{var}(U_j) \), can be interpreted as the lack of a halo effect (Beckwith and Lehmann 1975; Thorndike, 1920) for product \( j \). For example, if product \( j \) has a strong halo effect, i.e., it is perceived similarly on all characteristics, this term will be small. Hence, other factors constant, the larger the halo effect on product \( j \), the easier the choice becomes. The next term, \( \text{var}(U_k) \), can be analogously interpreted as the lack of a halo effect surrounding product \( k \). Again, the larger the halo, the easier the choice. The final term, \( \text{cov}(U_j, U_k) \), represents the perceptual similarity of product \( j \) and product \( k \). It is inversely related to the cost of thinking. This term will tend to be large if the two products vary similarly, that is, if products \( j \) and \( k \) are both rated highly on the same attributes.

Consider Figure A. Here, product A and product B are compared on three characteristics. They vary similarly, i.e., A is high on the same characteristics as B. Product A is superior on all characteristics, the covariance is positive, and the thinking cost is small.

Now consider Figure B, which illustrates two products, C and D. The utility of Product C is the same as Product A, i.e., ten, and the utility of Product D is the same as Product B, i.e., seven. Again, the difference in preference is three units. The respective variances are also identical. However, the covariance is now negative. Product C is superior on only one characteristic. The consumer must trade off the third characteristic with the first and second characteristics. The result is a more complex and difficult decision. Hence, the covariance term represents the perceptual complexity inherent in the product differences.

Summarizing, the cost of thinking is directly proportional to the perceptual complexity in comparing the products (halo effect and difference effect), and is inversely related to both the difference in preference between the products and the confidence at which the choice must be made. These three factors influence the expected number of comparisons necessary to make a particular choice and, thus, the expected difficulty of comparing two products. These three factors also determine a bound for the number of necessary comparisons and, hence, the potential difficulty; the actual difficulty of comparing the two products will be probabilistic and depend on the characteristic selection. This potential difficulty is a bound on the number of characteristic comparisons necessary to achieve confidence level \( \alpha \). That bound follows for a binary choice and is the sample size sufficient to achieve the desired confidence level, i.e., meet Conditions A and B:

\[ f_p = \frac{\text{var}(z)}{(1 - \alpha)E[z]^2}, \]

where \( f_p \) = the potential cost of comparing two products. This quantity is an upper bound on \( f \) as proven by Theorem 1:

Theorem 1: If \( f_p \) comparisons are made then the probability of making a correct choice is at least \( \alpha \).

Equation 3 is an upper bound on the minimum number of comparisons, \( n^* \), and not the actual \( n \). When the exact distribution of \( z \) is given, \( f \) can be computed directly. However, \( f_p \) may vary monotonically with \( f \) (i.e., still representing relative difficulty) and, there-

\[ \text{Proofs of theorems are given in Appendices, which are available from the author.} \]
fore, \( f_p \) provides a method for approximating thinking costs with a closed form expression. The methodology for approaching thinking costs as proposed by this paper is not dependent on using this particular surrogate for \( f \). Three comments in this connection are appropriate.

First, note that \( f_p \) defined by Equation 3 can be written in terms of the utilities of product \( j \) and product \( k \) as follows:

\[
f_p = \frac{\sigma_j^2 + \sigma_k^2 - 2\sigma_{jk}}{(1 - \alpha)(\mu_j - \mu_k)^2},
\]

where,
\[
\mu_j = (1/N) \sum_{i=1}^{N} U_{ij},
\]
\[
\mu_k = (1/N) \sum_{i=1}^{N} U_{ik},
\]
\[
\sigma_j^2 = (1/N) \sum_{i=1}^{N} (U_{ij} - \mu_j)^2,
\]
\[
\sigma_k^2 = (1/N) \sum_{i=1}^{N} (U_{ik} - \mu_k)^2,
\]
\[
\sigma_{jk} = (1/N) \sum_{i=1}^{N} (U_{ij} - \mu_j)(U_{ik} - \mu_k).
\]

Second, \( f_p \) is scale invariant. The utilities can be subject to any linear transformation and \( f_p \) remains the same. Hence, this \( f_p \) is consistent with the use of utility functions unique to a linear transformation, as derived from most axiom systems (Herstein and Milnor 1953; Keeney and Raiffa 1976). Further, the utilities can be measured with standard techniques, such as conjoint analysis (Green and Rao 1971; Green and Srinivasan 1977; Luce and Tukey 1964) or its extensions (Hauser and Shugan 1980).

Third, \( f_p \) is either infinite or undefined as the difference in utilities approaches zero while the actual \( f \) will approach the total number of characteristics. This means that when two products have exactly the same utility, it is impossible to determine which product is superior.\(^7\)

Comparing Multiple Products

If \( M \) products are considered, Equation 1 gives the cost of thinking. Letting \( f_i \) be the comparison cost for the \( i \)th comparison, Equation 1 can be rewritten as follows:

\[
c = \sum_{i=1}^{M-1} f_i = (M-1)\bar{f},
\]

where,
\[
c = \text{the cost or effort needed to make the choice},
\]
\[
\bar{f} = \text{the average binary comparison cost}.
\]

In general, the distribution of \( z \) is unknown. In this case, the average comparison cost, \( \bar{f} \), can be replaced by the average potential cost, \( f_p \). Note, the exact order in which the products are compared may affect both \( f \) and \( f_p \). Many ordering criteria are possible.\(^8\) In this paper, the criterion chosen will be the minimum \( f_p \) over all nonerrored\(^9\) orders. Hence, assume that nonoptimal products are optimally eliminated. Finally, define \( c_p \) as the potential difficulty or "thinking cost" termed the confusion index, formulated as follows:

\[
c_p = (M - 1)f_p^*,
\]

where,
\[
f_p^* = \text{the average cost per comparison given the optimal comparison order},
\]
\[
M = \text{the number of alternatives (} M - 1 \text{ is the number of comparisons).}
\]

A Numerical Example

Consider a choice involving three household cleaners differing on four characteristics. Table 1 indicates a consumer's utility associated with each product by characteristic. These utility values can be obtained, for example, from part worths in conjoint analysis (Tversky 1967) or by multiplying importance by belief in a linear compensatory model. Theorem 1 can be used to determine the difficulty of each binary comparison (for example, let \( \alpha = 0.5 \)).

The cost of deciding between Windex and Lysol can be computed as follows:

\[
\mu_{\text{Windex}} = 3.75 \quad \sigma_{\text{Windex}} = 14.19 \quad \sigma_{\text{Windex,Lysol}} = 10.44
\]
\[
\mu_{\text{Lysol}} = 4.75 \quad \sigma_{\text{Lysol}} = 8.19
\]

and
\[
c_p = \frac{[14.19 + 8.19 - 2(10.44)]}{(1 - 0.5)(3.75 - 4.75)^2} = 3.0.
\]

It follows that the cost between Windex and Ajax is:
\[
c_p = 54.0,
\]
and the cost between Lysol and Ajax is:
\[
c_p = 2.0.
\]

\(^7\) Perhaps the cost of comparisons can no longer be viewed as fixed when many comparisons must occur. In that case, the cost must be marginally increasing or \( \alpha \) must be decreased.

\(^8\) The expected cost rather than lowest cost is a usual and intuitively appealing measure. However, there are at least two reasons for using lowest cost. (1) Consider a lab psychologist who desires to measure the difficulty of a maze. That psychologist could use the number of necessary turns, for inches traveled, assuming the best possible path is taken, or the psychologist could use the number of necessary turns given a random path. The random path may be a poor measure because some sequences would obviously be eliminated by very simple learning. (2) The mean (expected value) may be a potential measure when \( f_p \) is used. This drawback comes from \( f_p \) being a potential rather than mean cost. Hence, some values of \( f_p \) will drastically overstate \( f \).

\(^9\) These orders assume binary choices are correctly resolved.
TABLE 1
UTILITIES BY CHARACTERISTIC BY HOUSEHOLD CLEANER

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Product</th>
<th>Drying speed</th>
<th>Color attractiveness</th>
<th>Polishing</th>
<th>Scent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windex</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Lysol</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Ajax</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Binary Comparisons. Previously, $c_p$ was argued to have the same monotonic properties as $c$. Thus, if the actual cost, $c$, were computed for each product comparison, the respective costs would be 1.7, 3.2, and 0.8, rather than 3.0, 5.4, and 2.0, yielding the same rank order.

The rank order depicts the relative difficulty of the respective binary choices. It becomes clear that choosing between Lysol and Ajax is relatively easy ($c_p = 2.0$). Lysol is, after all, superior on virtually every characteristic. However, choosing between Windex and Ajax is relatively difficult ($c_p = 54.0$). The inferior product for this consumer, Ajax, is superior on 3 of 4 attributes, requiring numerous tradeoffs to identify Windex as superior.

Windex versus Lysol versus Ajax. The lowest comparison cost would be achieved by first comparing Lysol and Ajax, eliminating Ajax and then comparing Lysol and Windex, eliminating Windex. The total comparison cost would then be 5.0. This sequential elimination can be pictured as a tournament, with the binary choices representing matches. The minimum $f_p$ can be thought of as the “lowest cost tournament,” namely:

- choose between Lysol and Ajax,
- eliminate Ajax as inferior,
- choose between Lysol and Windex,
- select Lysol as superior.

The tournament is illustrated in Figure C. Lysol is first matched against Ajax and found superior, with an associated cost of 2.0. Lysol is then compared to Windex and is found superior with an associated cost of 3.0. The total “thinking cost” of choosing Lysol from these three household cleaners is $2.0 + 3.0 = 5.0$.

There are many possible tournaments. The one actually used will depend on numerous factors, such as the expected cost of the tournament, as previously discussed. Again, note that the tournament represents only the expected or potential difficulty of the decision. The actual cost is random and will depend on the luck or skill with which the products are compared.

IMPLICATIONS FOR CHOICE BEHAVIOR

Optimality of Simplified Rules of Behavior

Tversky (1972), Coombs (1964), and Dawes (1964) have proposed simplifying decision rules that disregard information in an attempt to simplify the choice process. Wright (1975) offers a taxonomy of strategies and emphasizes the implication of these strategies to marketing research. Bettman (1971) and others (Payne 1976) have investigated the choice structure from an information processing viewpoint. Einhorn (1970) showed the mathematical relationship of several simplifying rules to linear compensatory models (as limiting cases). However, he emphasizes that “future research should concern the conditions, whether within the individual or in the task” under which these rules should apply. The confusion index previously developed provides a measure of the cost of these simplifying strategies.

A Conjunctive Rule. The conjunctive rule first assigns a minimal acceptable level to each characteristic. For example, if ammonia content is a characteristic of household cleaners, a minimal acceptable level may be 15 percent ammonia. Tversky (1972) extends this concept to define an aspect. Here, if the cleaner has 15 percent or more ammonia content, it is said to have ammonia. Essentially a step-function type of utility is assumed. However, each aspect could represent a level of the characteristic.
To determine when a conjunctive strategy is less costly than a compensatory model, the confusion index must be computed. Consider the choice between five products (A through E), on four characteristics evaluated, as shown in Table 2. To measure the difficulty of this decision, compute $c_p$ for the lowest cost tournament for which matches are correctly resolved. These matches, and associated costs, are as follows (the superior product labeled with an asterisk):

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$c_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B* vs. D</td>
<td>4.7</td>
</tr>
<tr>
<td>B vs. E*</td>
<td>13.7</td>
</tr>
<tr>
<td>A vs. C*</td>
<td>27</td>
</tr>
<tr>
<td>B vs. C*</td>
<td>66</td>
</tr>
</tbody>
</table>

The total cost of using a compensatory model is, therefore, 111.4.

Next, the conjunctive strategy cost can be evaluated. To operationalize the conjunctive strategy, a minimal acceptable level of 3.0 will be assigned to each characteristic. If a characteristic does not meet the assigned level, the consumer utility for that characteristic will be given a very large negative value, called $-M$. Then, when a difference comparison occurs on that characteristic, the product not meeting the reasonable level is essentially eliminated.

Table 3 illustrates how Table 2 is transformed by a conjunctive strategy. Costs were computed by letting $M$ approach infinity. The total cost of the conjunctive strategy is 76.0. Hence, the conjunctive strategy for this particular choice conflict not only involves less cost, but yields the same outcome (Product C) as the compensatory strategy. Also, the lowest cost tournament structure for the conjunctive strategy is, by coincidence, the same as the compensatory strategy.

An interesting feature of the conjunctive model is that its extreme case would be very costly. In this case, each and every characteristic must be observed to ensure each meets its predetermined minimum level. Therefore, the conjunctive rule was slightly redefined. When comparing two products' characteristics using a conjunctive strategy, the product meeting the fewest levels was eliminated. Then, in a sequential framework characteristic differences are observed until either one product does not meet the reasonable level or until the consumer is confident both meet all levels, in which case the product with the larger inferred utility is chosen. Therefore, the conjunctive rule implies a product must meet the minimum level on a reasonable number of characteristics. Theorem 2 indicates the cost of using a conjunctive rule.

**Theorem 2**: The "thinking cost" of eliminating a product (compared to any satisfactory product) with a conjunctive rule is given by:

$$C_{\text{conjunctive}} = \frac{(1/k) (N - 1)}{(1 - \alpha)},$$

where,

- $k =$ the number of characteristics on which the product misses the standard level for that characteristic,
- $N =$ the total number of characteristics, and
- $\alpha =$ the confidence level for the decision.

When the number of characteristics, $N$, becomes large, the "cost of thinking" associated with the conjunctive rule increases. Wright (1975) has found this result supported experimentally. Now, as the number of characteristics not meeting the minimum level (i.e., $k$) increases, the "cost of thinking" decreases. For when the product has numerous characteristics at levels below the minimum, it takes few characteristics comparisons to reject the product as inferior. (Because the potential difficulty is being computed, all characteristics are used when computing $C$ even though the consumer will not necessarily observe all characteristics.)

**A Disjunctive Rule (Maximax)**. A disjunctive rule employs a comparison of each product on its best characteristic. For example, if Product A excels on Characteristic 2 and Product B excels on Characteristic 4, then only these two characteristics are used to dictate the choice.

Using a disjunctive strategy, Table 2 is transformed into Table 4. Only the product's best characteristic is used, and all other characteristics are set to zero. Evaluating the lowest cost tournament for the disjunctive strategy yields a total cost of 178.0. Thus, the adoption of a disjunctive rule not only leads to a larger


“The thinking cost” than either a compensatory strategy or a conjunctive strategy, but also leads, in this case, to the selection of a less than optimal product (i.e., Product A). Theorem 3 indicates the cost of using a disjunctive strategy.

Theorem 3: For the disjunctive strategy, the “thinking cost” involved in comparing a product with the null alternative is given by:

\[ C_{\text{Disjunctive}} = \frac{(N - 1)}{(1 - \alpha)}, \]

where,

\[ N = \text{total number of characteristics}, \] and
\[ \alpha = \text{the confidence level}. \]

\( C \) is a function of \( N \) because the consumer must find the best characteristic. This expression reveals the conjunctive strategy always to have less or equal cost than the disjunctive strategy for comparing a product against the null alternative (product elimination).

It also can be shown, using the confusion index, that a decision is easier using disjunctive strategy for comparing two products with the same best characteristic, than when each is best on a different characteristic.

This analysis assumes the consumer knows the best characteristic when it is found. This assumption should be revised in future research to include only the maximum of the sample as the “best known characteristic.”

The Maximin Rule. Maximin strategy compares the products on their weakest characteristic. The strategy dictates the selection of the product with the highest value on its weakest characteristic.

Table 5 reflects Table 2 transformed for a Maximin strategy. The total cost of a Maximin strategy for this set of products is 160.0. For this example, the strategy is, therefore, more costly to implement than either the conjunctive or compensatory strategies, and is less costly than the disjunctive strategy. However, the Maximin strategy does determine the optimal product for this example. Note again that the Maximin strategy is redefined along the same lines as the conjunctive strategy. Theorem 4 indicates the relative cost of the Maximin strategy.

Theorem 4: For the Maximin strategy, the “thinking cost” involved in comparing a product against the null alternative is given by:

\[ C_{\text{Maximin}} = \frac{(N - 1)}{(1 - \alpha)}. \]

Also, when using a Maximin strategy, the cost of comparing two products is smaller if they have the same worst characteristic rather than different worst characteristics.

Errors and Mistakes

Thus far only costs were considered. However, simplified decision rules reduce expected benefits by allowing mistakes. The previous section assumed that consumers made enough characteristic comparisons so that the chance of error was less than \( \alpha \) for each product elimination. It is then possible to state the probability of an error, i.e., not choosing the best product. That probability is one minus the probability of choosing the best product computed by Equation 6.

\[ P(\text{choosing best product}) = \frac{1}{(\alpha M)(\alpha(1 - \alpha)^{M-1} + (1 - \alpha) - (1 - \alpha)^M)}, \]

where \( M = \text{the number of products} \) and \( \alpha = \text{the confidence level}. \)

Note that for \( \alpha = 0.5 \), \( P(\text{choosing best product}) \) is \( 1/M \).

Until now, some confidence level was set and enough characteristic comparisons were made to achieve that confidence level, the number of necessary comparisons depending on the products. The more comparisons necessary to make the choice, the more costly was the choice. Assume the consumer sets the cost of thinking rather than the confidence level. Here the consumer is assumed to allocate the appropriate thought to the decision to be consistent with an expected utility maximization. For example, let the consumer set the cost of thinking at the lowest possible level, i.e., one. Then, given two products, \( A \) and \( B \). Theorem 5 gives the probability the consumer will choose Product \( A \) over Product \( B \), denoted \( P(A \text{ over } B) \), and takes the form of Luce’s axiom (with ordinary utility functions substituted for Luce’s scale values) when (1) consumers minimize thinking costs, (2) the utility functions take just two values (e.g., 0 or 1), and (3) the products have no characteristics in common.

Theorem 5: If the individual seeks to minimize thinking costs, the utility functions are defined over aspects
Choosing Among Unlimited Numbers of Products

Recalling Equation 2, the expected cost of making a decision was $m\bar{f}$. However, in general, the number of product comparisons may be less than the total number of products. For example, even if there are 30 products, one product could be found superior to five others after four comparisons. The consumer may now stop and feel satisfied (confident) with the decision. In this case, not only is $f$ a random variable, but so is $m$.

Now if $m$ and $f$ are independent, then we could use Equation 3 to obtain a potential difficulty measure when the number of product comparisons is variable. However, determining the optimal number of products has been well studied in the decision analysis literature (Blackwell 1965; Wald 1947) and in psychology (Pollack 1970), and used to formulate a theory of information and search (Nelson 1970; Stigler 1961; Wilde 1977). Therefore, the problem of determining $m$ independent of $f$ will not be addressed here. If the average or expected number of products $m$, denoted $\bar{m}$, could be determined, the expected decision difficulty would be $\bar{m}\bar{f}$. Then, $\bar{m}$ would decrease the decision difficulty when the consumer felt future product comparisons would lead to no further improvements, which is consistent with experimental results (Hendrick, Mills, and Kiesler 1968; Kiesler 1966).

Often, $m$ and $f$ are not independent. Then, as the choice process proceeds, the average comparison cost changes. One would expect comparison costs to increase as set size decreases, because $E(f)$ decreases. Hence, the relative costs of different simplifying rules change as the choice proceeds, as has been found experimentally by Payne (1976) and Wright and Barbour (1977).

APPLICATIONS AND IMPLICATIONS

Being able to quantify decision-making costs and the likelihood of mistakes, given reduced decision-making effort, has numerous implications, as the following three examples indicate.

Product Characteristics in an Advertisement

The determination of how many product characteristics should be included in an ad has been more of an art than a science (Kotler 1978). The methodology previously discussed allows the potential for ascertaining how many product characteristics should be included in an ad. For example, suppose consumers sample ads rather than read them in their entirety. Equation 3 implies that when advertised brands are in product categories with high characteristic variability, the ad should mention as many of the brand's favorable characteristics as possible. When the characteristic variability is low, the company should stress only the
brand’s best characteristics. Similarly, Equation 3 states that high-priced (relative to total income) products requiring large confidence levels should include more characteristics than products with smaller consumer confidence levels, keeping variability constant. Finally, if future research could measure the confidence level ($\alpha$) at which consumers approach particular product category decisions, the best number of characteristics to advertise to maximize purchase probabilities could be determined.

Information Presentation

In some situations decisions should be made easier, in others more difficult (Russo 1977). For example, in arranging a data base, control panel, or a mail-order catalog, decisions should be made easy. Hence, items should be grouped to minimize average characteristic utility differences (other factors held constant). Also, in some situations changing the difficulty of the decision could change the respective choice probabilities. For example, school cafeterias may want to influence children’s meal selections.

New Product Sales Forecasting

New products are generally thought to compete most with “similar” products. These “similar” products are thought to attract people desiring the same characteristics as the new product. However, this phenomenon may require some time. In the short-run when test marketing occurs and ultimate product success is predicted, the consumer may still be gathering information about the new product. Therefore, the purchase probabilities may reflect thinking costs leading to partial product evaluation. The new brand may, in the short-run, receive its market share from competitive products that are easy to compare to the new brand rather than from competitive products for which the effort of comparison is greater, even though long-run shares may be quite different.

Testable Hypotheses

Numerous studies show brand identification can cause a uniformity of perception across attributes (Allison and Uhl 1964). This effect manifests itself by creating strong halo effects about brands (Beckwith and Lehmann 1975). This empirical finding indicates that brand name identification will decrease $var(z)$. By Equation 3, $f_c$ would be decreased, and by Equation 4 the “cost of thinking” is decreased. That is, less information need be sought to maintain the same confidence, $\alpha$. This implication has had some experimental verification (Jacoby, Szybillo, and Busato-Schach 1977). However, Theorem 1 defines precise implications that could be empirically tested. For example, $f_c$ could be computed and compared against “stated difficulty,” “time spent,” and other empirical measures.

The cost of thinking can be reduced by (1) memory, (2) summary statistics, and (3) probabilistic sampling. Thinking costs will be large when the confidence $\alpha$ is large (e.g., for large ticket items relative to total income) and when the characteristic utility variability is large (e.g., products that serve very different markets). In these cases, the consumer may try to reduce costs through memory.

The cost of future decision making can be reduced by remembering large characteristic differences. Just as attribute variability allows greater differentiation (van Raaij 1977), remembering some $z$ can reduce $c_p$. Further, poor memory may encourage adoption of simplifying decision rules. The cost of decision making can also be reduced by gathering summary statistics. Hence, large thinking costs with common tastes, i.e., same $U(-)$, will lead to awards given to best products, certifications, and branding. Finally, if key attributes, those with large variability, are known the consumer can selectively sample characteristics, engaging in probabilistic sampling. Thus, large thinking costs will lead to activities for finding key discriminatory attributes. For example, a consumer buying a boat for the first time may first seek a book about “what to look for in a boat,” rather than information on particular boats. Experience may be defined by knowing which attributes have high variability.

LIMITATIONS

In attempting to model and abstract consumer behavior, some restrictive assumptions are often required. This paper provides no exception. Fortunately, quantification has made these assumptions less obscure. For example, the use of single stage sampling rather than sequential sampling was an obvious limitation of the current development. Another limitation is the assumption of a fixed cost per comparison. Although this is a handy assumption, it is clear that this cost should increase as comparisons are made. If “thought” is a limited resource, its use should meet with increasing marginal opportunity costs. This limitation is related to the necessity of requiring $\alpha$ to be determined outside the model.

A third limitation is the static nature of the model. Clearly, the real problem of interest would be the dynamic model. Current research in that area has led to some interesting theories (Lehmann 1977). Although current research on memory (Johnson and Russo 1978) is consistent with this paper, dynamic extensions would require the inclusion of memory and learning. In the dynamic case, it may no longer be appropriate to assume an arbitrary consumer randomly selects characteristics, but instead it may be more appropriate to assume an a priori vector of probabilities. In fact, the behavior of this probability vector over time might...
hold the key to the most powerful applications of the development. Also, the evaluation costs of the characteristics may change in the dynamic case because of consumer memory.

**SUMMARY AND CONCLUSION**

A way was provided to explicitly model and measure the cost of thinking. A fundamental unit of thought was defined, which measures the potential difficulty of a decision by examining the characteristic utilities of the alternatives.

With this framework for exploring thinking, a confusion index was derived as a measurable bound on the expected number of necessary units of thinking required to make a choice. It was then possible to determine the relative costs of specific simplifying decision rules as a function of the alternatives. Formulas for computing costs of various decision rules were then derived.

On an individual level, adoption of simplifying choice strategies can leave a consumer vulnerable to manipulation. Choice conflicts can be changed (for example, by the inclusion of nonoptimal and therefore irrelevant products) to lead the consumer to select an inferior product. Hence, by manipulating the choice setting, some degree of control can be exercised over the consumer.

Einhorn and Hogarth (1978) note this effect may occur over an extended period of time. Lack of a proper feedback of information keeps the consumer from learning of the mistakes. Thus, mistakes recur and the best product is never found. Choice rules can, then, have implications for advertising copy decisions, in-store display design, strategies for launching new products, and pricing decisions.\(^\text{12}\)

[Received March 1978. Revised February 1980.]

**REFERENCES**


——, and Shugan, Steven M. (1980), "Intensity Measures

\(^\text{12}\) A relevant yet unanswered question is the effect of simplifying rules on an aggregate market. Perhaps, if consumers continually made mistakes, someone would have the incentive to inform them of their mistakes. Whether a firm could manipulate the behavior of one individual without interference from other individuals remains an unknown.


