An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables

J. EDWARD RUSSO
FRANCE LECLERC*

The nature of the choice process for commonly purchased nondurables was examined by tracking eye fixations in a laboratory simulation of supermarket shelving. The observed process contains three stages that were interpreted as (1) orientation, (2) evaluation, and (3) verification. Orientation consisted of an overview of the product display, although some initial screening out of alternatives also occurred. The evaluation stage, the longest by far, was dominated by direct comparisons between two or three alternative products. The last stage, devoted to verification of the tentatively chosen brand-size, mainly examined alternatives with few or no previous fixations. Greater familiarity with a product category led to a choice process that was shorter and that focused on fewer alternatives, but these effects were confined to the evaluation stage. The findings are fully compatible with the general view that the choice process is constructed to adapt to the immediate purchase environment.

Over the last two decades the dominant view of the purchase decision has shifted from classic utility maximization to a choice process that constructively adapts to the purchase environment. Payne (1976) proposed a two-stage choice process that was contingent on characteristics of the choice environment: an initial heuristic stage that efficiently eliminated all but a few alternatives, followed by a second stage devoted to a thorough consideration of the remaining set. Then, eschewing a fully preplanned strategy, Bettman (1979; Bettman and Zins 1977) characterized the choice process as constructive; a patchwork of subprocesses put together at the time of choice to adapt to the specific decision environment (see also Payne 1982; Svenson 1979). Finally, Hoyer (1984) and Olshavsky (1985) described choice "tactics" and heuristics developed adaptively over repeated choices in the same product category (see also Hoyer and Brown 1990; Leong 1993).

The transformation of the prevailing paradigm of consumer choice from planned, optimizing, and comprehensive to constructed, adaptive, and simplifying poses a formidable challenge to theory construction. Future theories will almost certainly be less parsimonious than utility maximization and more closely tied to actual choice processes. In order to build such theories, the choice process needs to be observed in enough detail over enough task environments to reveal its key regularities and adaptive mechanisms. As a first step toward this ultimate goal, we track the choice process in detail in one task environment. Further, we focus on a common claim about the overall structure of the choice process, namely, that it can be partitioned into distinct stages.

STAGES OF THE CHOICE PROCESS

The claim that the choice process can be partitioned into distinct stages or phases is common to many of the explanations for the descriptive failure of multivariate utility theories.1 Most commonly these models include two stages: an initial screening stage followed by a more thorough evaluation of the few remaining alternatives (Biehal and Chakravarti 1986; Lussier and Olshavsky 1979; Payne 1976; Wright and Barbour 1977). Alternatively, the choice process may not be characterized by qualitatively separable stages but rather by more or less continuous trends in processing operations, such as an early emphasis on screening that later shifts to a dominance of evaluation operations.

Given that nearly two decades have elapsed since the stage hypothesis was first proffered in the consumer

* J. Edward Russo is professor of marketing at the Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853-4201. France Leclerc is assistant professor of marketing at the Sloan School of Management, MIT, Cambridge, MA 02115. The authors gratefully acknowledge the comments of James R. Bettman and Eric J. Johnson on an earlier draft of this article. This research was supported by grant number MH 24880 from the National Institute of Mental Health to J.E.R.

1 Stages connote processes or activities that are qualitatively different from each other. Phases can share this meaning but also refer to repetitions of the same activity. For an example of the latter, see Klein and Yadv (1989). In what follows, we shall use only the narrower term, stages, which means time segments of the choice processes that are distinguished by their different subprocesses.

© 1994 by JOURNAL OF CONSUMER RESEARCH, Inc. • Vol. 21 • September 1994 All rights reserved. 0093-3301/95/2102-0042$2.00

Copyright © 1994. All rights reserved.
choice context, why has it not been resolved empirically? One problem has been that adaptivity during the process may blur the stage boundaries. For instance, in many choices it may be adaptive to interrupt an initial screening stage and begin a comprehensive comparison between two alternatives to preserve current contents of short-term memory (Russo and Rosen 1975).

A second barrier has been the need for data that are detailed enough to reveal stage boundaries if they exist. In the absence of such data, process-tracing analyses have typically segmented the process by time, such as by contrasting its first and second halves (see, e.g., Bettman and Park 1980). Methodologically, concurrent verbal protocols have been used to reveal important and consistent differences between halves. However, they have not been able to distinguish between a true stage structure and a choice process of more or less continuous trends over time. While keeping in mind that adaptivity may blur any stage boundaries, we use eye fixations in an attempt to provide the detail needed to distinguish between the stage and trend structures.

If a stage structure is found, what subgoals characterize the different stages? Is one stage devoted to heuristic-based screening followed by a second the goal of which is comprehensive product comparisons? This is what the most common two-stage theories seem to suggest with at least some empirical support. For instance, Van Raaij (1977) reports that direct comparisons between two or three alternatives occur more frequently later in the decision process, while Lussier and Olshavsky (1979) claim that the second stage of the process is based on comparisons.

Finally, we consider the effect of consumer familiarity with the product category (Alba and Hutchison 1987). Shoppers more familiar with a product category are expected to be more focused in their search, able more quickly to eliminate undesired alternatives, and more adept at locating and settling on the chosen alternative.

METHOD

Empirical Rationale

To reveal the choice process in sufficient detail, we chose to track the sequence of eye fixations. Verbal protocols were also recorded, but, for several reasons discussed below, their additional contribution to the eye-fixation data was negligible. We simulated a supermarket environment in the laboratory using real product specimens (containers) with actual prices marked on them and duplicating key elements of actual shelf arrangements. Performing the study in a real supermarket was ruled out by the need to record eye fixations. The simulated supermarket shelves were placed against a one-way mirror so that eye fixations could be recorded unobtrusively.

Typical shoppers were recruited and were asked to make three purchase decisions. However, these pur-

### FIGURE 1

THE SHELF ARRANGEMENT FOR PEANUT BUTTER

<table>
<thead>
<tr>
<th></th>
<th>Jif</th>
<th>Jif</th>
<th>Jif</th>
<th>Jif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laura</td>
<td>$1.91</td>
<td>$1.42</td>
<td>$0.97</td>
<td>$0.99</td>
</tr>
<tr>
<td>Snapper</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 oz</td>
<td>$1.65</td>
<td>$0.70</td>
<td>$0.81</td>
<td>$0.59</td>
</tr>
<tr>
<td>Jif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 oz</td>
<td>$1.42</td>
<td>$0.96</td>
<td>$0.80</td>
<td>$0.59</td>
</tr>
<tr>
<td>Jif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 oz</td>
<td>$1.31</td>
<td>$0.85</td>
<td>$0.80</td>
<td>$0.59</td>
</tr>
<tr>
<td>Jif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 oz</td>
<td>$1.19</td>
<td>$0.85</td>
<td>$0.80</td>
<td>$0.59</td>
</tr>
<tr>
<td>Jif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 oz</td>
<td>$0.99</td>
<td>$0.85</td>
<td>$0.80</td>
<td>$0.59</td>
</tr>
</tbody>
</table>

Notes.—The three distances between the four containers on each shelf were each 51 cm (20 in.). The three vertical distances from the bottom of one shelf to the top of the one below were, from highest to lowest, 30.5 cm (12 in.), 33 cm (13 in.), and 40.6 cm (16 in). The bottom shelf was 83.7 cm (33 in.) above the floor. Nu Made and Real Roast were store brands.

chases were hypothetical, and shoppers knew that their purchases were being observed by researchers. Additional differences from shopping in an actual supermarket were subjects' awareness of being in a university laboratory and an unfamiliar shelf display.

Stimuli

Three commonly purchased product categories were studied: applesauce, ketchup, and peanut butter. Each was represented by 16 brand-sizes at the nearby supermarket that served as our prototype. These three product categories contained six, five, and six different brands, respectively, and ten, five, and seven different sizes, respectively. Note that these do not form a factorial matrix of 4 brands × 4 sizes but do reflect manufacturers' decisions on which sizes to produce and store executives' decisions on which brands and sizes to stock at the store that we used as our model. Four shelves were constructed; 16 containers, one specimen of each brand-size, were placed on the shelving in a pattern as similar as possible to that in the prototype store. The only display constraint was the use of a 4 × 4 array (four items on each of the four shelves), with enough separation among the 16 containers to enable errorless identification of any item being fixated. One such display is shown in Figure 1. Presenting multiple specimens of each brand and size, as on actual supermarket shelves, would have partially blocked the recording through the one-way mirror. Also, reducing the separation (empty shelf space) between containers would have led to errors in identifying which product (brand-size) was being fixated. The actual products used were purchased immediately prior to the two weeks of data collection.
supermarket's prices were marked on each container; no shelf tags were used.

Subjects
Subjects were 47 primary household shoppers, all female, although that was not a requirement. They were recruited from local civic, religious, and social organizations in a large Western city that was selected to offer a wide range of age, education, and income. The mean number of years of formal education was 1.5 years above the citywide average, with the mean household income $2,200 below the citywide average. Subjects ranged in age from 18 years old, the minimum permitted, to above 70 years old. In return for each participant's time, a payment of $5 was made to her organization, and she received $2.

Task
Each subject was instructed to shop as she normally would in her own supermarket and to assume that the product category of interest was on her shopping list that day. Before making any choices, the names of the store brands used were explained to subjects, and they were asked to assume that our store brands were identical to the ones available at their usual supermarket. They were also required to generate a think-aloud (concurrent) verbal protocol and to signal the end of a trial by announcing their choice.

The 47 subjects and three product categories yielded 141 choices. Of these, three were lost because of failure of the video equipment recording the eye fixations, which left 138 trials as the main database.

Procedure
Subjects began with practice choices in seven non-experimental product categories. The alternative brand sizes were presented on a 61 × 61-cm (2 × 2-ft) poster board that schematically represented shelving and products in the same arrangement as in our prototype store but without all the individual containers. These choices were used to accommodate subjects to the laboratory environment and to practice the generation of a concurrent think-aloud verbal protocol.

Subjects also made two practice choices using actual products and shelving prior to the three experimental choices. These two preliminary choices were arranged on the same wall as the first experimental choice. A subject was led down the wall and asked to look not at the next product but rather at two marks on the floor that indicated where to place her feet. This facilitated the eye-movement recording by ensuring that a subject's face was not hidden behind one of the containers. When these marks were reached she was informed of the product category and was allowed to look up to the shelving and begin her selection process. The shelves for the practice and experimental choices were located such that the latter took place in front of the one-way mirror. Thus, subjects proceeded as if they were moving down a supermarket aisle, and the one-way mirror seemed to be an incidental aspect of the room. When debriefed afterward, only one subject claimed to have noticed the mirror and wondered about its role in the experiment, but even she did not suspect that it concealed a video camera. A questionnaire recorded each subject's purchase frequency and usually purchased brand, as well as several demographic characteristics. The entire session lasted between 30 and 45 minutes.

DATA ANALYSIS
Eye Fixations
The eye fixations from the 138 choice trials were coded for location (but not for duration) and were grouped into sequences that corresponded to possible stages. The duration of these sequences was calculated (the details of this coding procedure and other technical aspects of the fixation analysis are given in the Appendix). It should be noted, however, that estimating the durations of the more than 6,000 eye fixations from videotapes was impractical. Thus, the durations of individual fixations are not available for analysis. Instead, the durations of sequences of fixations will be used to measure the length of possible stages.

Raw eye fixations can be analyzed both individually and by aggregating them into subsequences that are believed to reflect processing units. The former is straightforward and typically yields conclusions about what objects were re-fixated most often, seen earliest, examined only once, and so forth. We used the second procedure, aggregation into a meaningful subprocess, to infer the presence of pair and triple comparisons. A direct comparison between two alternatives was identified by a fixation sequence of the form X-Y-X... and a comparison among three alternatives by X-Y-Z-X... Two tests validated the interpretation of these fixation patterns as comparison processes (see the Appendix).

Verbal Protocols
To our considerable disappointment the verbal protocols could not be used to augment any of the present analyses. First, they were relatively sparse, only 4.90 statements per choice (before the required announcement). Second, the protocols revealed much more about the product attributes being considered than about processing operations. Unfortunately, attributes reflect the content of the choice process rather than its structure, which is the focus of the present paper. Also note that sparse protocols are much more likely to report what product attribute was considered than which operation was used, because an attribute is usually described by a single word, while an operation often requires a com-
ple phrase. Thus, the sparseness of the protocols made it difficult to reliably distinguish, say, an information acquisition from an evaluation of an alternative. We attribute this to the partial automation of operations that is to be expected in repetitively performed tasks (Russo, Johnson, and Stephens 1989). Thus, rather than complicate the analysis of the structure of the choice process by introducing elements of content, we do not consider the verbal protocols further.

IDENTIFICATION OF STAGES

To identify possible stages, we searched for patterns of eye fixations. This search was entirely data driven, because there was little theory or prior empirical work to guide it. A conservative approach was taken; we searched for obvious patterns that appeared more or less consistently in many trials. We were able to identify with confidence only one such pattern, a sequence of eye fixations without any refixation of a previously seen alternative. As simple as it may sound, the stage distinctions that follow are all based on this single pattern.

Stage 1

The initial sequence of eye fixations was often characterized by a serial examination of adjacent products in an unbroken path without refixating one already seen. This type of sequence might represent one of two different processes, screening or orientation. Screening is an initial consideration of the available alternatives with the intent of eliminating most of the 16 that are available while retaining the more promising choice candidates for consideration in a second stage. Recall that screening is the first stage of the standard two-stage theory. It should require no more than one fixation for each alternative. Indeed, familiarity with brands and sizes, combined with either an expectation or a partial viewing of the shelf layout (e.g., smallest sizes on the top shelf), enables complete screening without complete viewing. Some of the 16 alternatives might be eliminated (i.e., screened out) on the basis of their presumed location, such as ignoring the entire bottom shelf because that is usually where the largest sizes are.

Orienting is acquiring information about available brands and sizes and about their physical layout. Such an initial overview enables shoppers to restrict subsequent processing to a smaller candidate set, such as the larger sizes on the bottom shelves. Thus, orienting helps shoppers to prepare for the more focused processing that follows.

The difference between screening and orientation is the extent to which this stage is used for an initial round of eliminations versus an organizing overview that precedes final eliminations. More fixations and longer mean fixation durations point to the screening interpretation of stage 1, while orienting requires relatively fewer and briefer fixations. We attempt to discriminate between these two interpretations in the data analyses that follow. First, we explain how this stage was identified (or operationally defined).

Stage 1 was defined as all fixations up to but not including the first refixation. The latter is likely to indicate a reconsideration rather than an initial examination. When a comparison such as X-Y-X-Y overlapped the stage 1-stage 2 boundary, it was partitioned between stages in proportion to the assignment of fixations. Thus, in the initial fixation sequence U-V-W-X-Y-X-Y, the first five fixations constitute stage 1, and 50 percent (two of four fixations) of the pair comparison between X and Y were assigned to each of stages 1 and 2.

Stage 3

We also found that sometimes the final fixations prior to the announced choice consisted of a string of single fixations that resembled the patterns observed in stage 1. This may have reflected a kind of review in which a tentative choice has been made and a last look at some alternatives finalizes that selection just prior to its announcement.

To identify stage 3, we again used the first refixation criterion, this time by working backward from the announcement fixation. That is, reversing the actual order of the sequence of fixations, which is the same as counting backward from the announcement fixation, the first refixation that was encountered defined the end of stage 2.

Stage 2

So far, we have identified an initial segment of the eye fixation sequence as stage 1 and a corresponding sequence immediately preceding the announcement of the chosen brand-size as stage 3. What lies between them is stage 2. Although this intermediate segment may seem precarious, and may be reduced to zero if stages 1 and 3 are long, the opposite actually occurs (as will be seen shortly).

There were three modifications of this criterion. First, a refixation was ignored if it occurred when backing out of one of the four corner locations (six trials). Second, an initial statement in the verbal protocol that the subject was searching for her usual brand reduced stage 1 to zero fixations for seven trials (because the usual brand was the first fixation and, presumably, began a different stage of processing) and increased stage 1 from zero to three fixations in two other cases. Finally, we were concerned that whenever more than one-half of all preannouncement fixations were to be allocated to stage 1, this assignment might be invalid. When this happened (eight trials), we imposed the stricter criterion (beyond no refixations) that the fixations occur at adjacent locations. These modifications reflect a compromise between attempting to identify a coherent initial process and applying a simple criterion, in the case, all fixations up to the first refixation.

The only modification to this criterion was the assignment of all fixations in comparisons to stage 2. Thus, if a trial terminated with the sequence X-Y-X-Y-Z-W, the middle two fixations were assigned to stage 2 and not to stage 3, as they would have been if the criterion of the first backward refixation had been strictly followed.
Stage 2 seems to be where most of the real work of evaluation takes place. It is devoted to deliberate, effortful evaluation, especially of the alternatives that are considered more seriously.

Stage 3b

The fixation during which the subject announced the chosen brand-size always terminated stage 3 (or stage 2, when no fixations were assigned to stage 3). Yet, in over 50 percent of the choices, there was at least one additional eye fixation. These postannouncement fixations were assigned to a separate stage. Unlike the boundaries between stages 1 and 2 and between stages 2 and 3, this last one is defined independently of the eye fixations.

This was an unexpected stage. Why would shoppers continue to examine the product display when they have completed and announced their choice? An answer was suggested by one shopper who, after this postannouncement examination, asked whether she could change her mind and select a different brand-size (which she was allowed to do). This suggests additional checking or verification. Indeed, if our interpretation of stage 3 as verification is correct, it is difficult to imagine another interpretation for this postannouncement stage besides further, although possibly different, verification. For this reason we number this last stage as stage 3b and renumber stage 3 as stage 3a.

RESULTS

Main Process Characteristics

We first describe the main characteristics of the choice process as revealed by eye fixations and differentiated by stage. Table 1 displays the most important characteristics of the choice process revealed by eye fixations. These characteristics are organized into several categories and are reported separately for each stage. The important points to observe in the data in Table 1 are (1) whether there are significant differences across the stages and (2) whether those differences accord with the stage interpretations given above.

Presence of Stages. Note immediately that stages 1, 3a, and 3b are not always present. All three are optional. Of these, however, it is only stages 3a and 3b, the two verification stages, that subjects omitted from more than 10 percent of the choices. Even here, at least one of these last two stages was present in 114 (83 percent) of the 138 trials. How are the occurrences of stages 3a and 3b related? First, are they independent? Alternatively,
does one imply that the other is more likely, because verification may be needed more for some choices than for others? Finally, does one imply the absence of the other, because by performing the same function they compensate for each other? The data support the last of the three possibilities, compensation. The probability of a nonzero stage 3b is .67 when stage 3a is missing but drops to .47 when stage 3a is present. Fisher's exact test (two-tailed) rejects the independence of stages 3a and 3b (p < .05). Thus, they exhibit a compensatory pattern in which the absence of one verification stage means the other is more likely to occur.

Total Work and Effort. The next two measures, time and the number of fixations, give a picture of the relative size of each stage. Stage 2 dominates the choice process by both measures. Even in the conservative case, when all three other stages are present (which is only 30 of 138 trials), it accounts for 63 percent of the total time and 54 percent of the fixations. For each of these two measures and for the subsequent eight process characteristics, a one-way ANOVA and Bonferroni tests of pairwise differences were performed. For both total time and fixations, only stage 2 was significantly different from the other stages.

One of the most telling characteristics in distinguishing among the identified stages is the mean fixation duration.6 The fixations in stages 1 and 3b are relatively brief and are significantly shorter than the durations of stages 2 and 3a. Furthermore, stage 2 exhibits an average duration that is significantly longer than that of stage 3a and that is over twice as long as that of either of the other stages. By contrast, the largest duration difference in the choice tasks studied by Russo and Rosen (1975) and Russo and Dosh (1983) was only 23 percent. That maximum is smaller than the minimum difference among the three pairs of adjacent stages in the present study. The large differences in duration also explain why stage 1 encompasses 20 percent of the fixations (5.42 out of 27.13) but only 12 percent of the total time (3.44 seconds out of 29.65 seconds).

The data for the two measures of total work, time and the number of fixations, are in straightforward agreement with the stage interpretations. For instance, the two verification stages are shorter than the preceding evaluation—a concordant, if unsurprising, finding. Much more interesting and more powerfully confirmatory are the mean durations. If stage 2 is devoted to evaluative inferences, as we claim, it is quite reasonable to expect it to have the longest eye fixations (Rayner 1978). Both orienting or screening in stage 1 and verification in stage 3 should be accomplished with quicker examinations of alternatives. This expectation is strongly supported. What is less expected is the longer duration of the fixations in stage 3a, which is two-thirds longer than those of stage 3b.

Breadth of Search. Both the number of different alternatives (brand-sizes) fixated and the proportion of brands fixated (counting any size) tell a similar story.7 Although stage 2 contains on average nearly three times as many fixations as stage 1, only 84 percent more alternatives and 34 percent more brands are fixated in stage 2 than in stage 1. Indeed, of the 11.59 alternatives eventually fixated, half (5.84) have been seen by the end of stage 1. This pattern of visual coverage of the alternatives reinforces the breadth of stage 1, and the contrasting depth and focus of stage 2. Note that the number of different alternatives fixated should be distinguished from the number of different fixations, because the latter includes all re-fixations on the same alternative.8 Since re-fixations were effectively prohibited by the definitions of stages 1 and 3a, only their relative paucity in stage 3b is noteworthy as being fully compatible with its interpretation as a final verification.

Focus of Search. The next two measures capture the extent to which the fixations are efficiently located. The first of these is the proportion of fixations on the brand-size that was ultimately chosen. The greater the proportion of such fixations, the more efficiently focused are the fixations in that stage. Table 1 shows large differences across stages, which is confirmed by the ANOVA results. However, note that the chance base rate of the proportion of fixations on the chosen alternative differs across stages. In particular, by definition, stages 1 and 3a contain no re-fixations, which constrains the proportion of fixations on the chosen alternative (out of the 16 available). Thus, an additional test must be performed to compare each stage’s observed proportion of fixations on the chosen alternative with its corresponding chance base rate (computed individually for each trial). We exclude stage 3a from such a test, since, by definition, it was always terminated by the choice announcement and, therefore, was likely also to end with a fixation on the chosen alternative. Similarly, in the test for stage 2 we excluded the 72 trials when stage 2 was terminated by the choice announcement.

---

6We remind readers that durations of individual fixations were not available, only the durations of stages. The mean duration for each stage was computed from the total time and number of fixations for that stage. Thus, the sample sizes for the mean durations reported in Table 1 are the numbers of trials with a nonzero stage, not the total number of individual eye fixations in a stage.

7The proportion, rather than the total number, of fixated brands was computed, because the number of available brands was not equal across the product categories. Recall that the total number of displayed brands was six for applesauce, five for ketchup, and six for peanut butter.

8One caveat must be stated. In spite of the careful separation of alternatives on the shelves, we cannot completely rule out peripheral visual processing (i.e., taking in some information about a brand and a size peripherally while fixating some other alternative). If it occurred at all, however, peripheral processing should have been limited to gross visual characteristics such as size, certain kinds of packaging (e.g., canned vs. bottled applesauce) and brands (because manufacturers typically strive for unique, easily recognized labels).
Thus, for stage 1, for stage 2 as modified, and for stage 3b we performed a two-tailed t-test of the difference between the observed and chance rates, after the variance of both had been stabilized by an arcane transformation. These tests revealed that the chosen alternative was fixated more often in stage 1 (p = .06), reliably more often in stage 2 (p < .0001), and less often in stage 3b, but without statistical reliability (p < .20).

An upward trend in the proportion of eye fixations on the chosen alternative over stages 1 and 2 is to be expected as the choice process increasingly focuses on the brand-size finally chosen. However, the small proportion of fixations on the chosen alternative for stage 3b (.08) indicates that the verifications that occur here seldom involve the chosen alternative or at least seldom require its fixation.7 This result distinguished stage 3b from the other stages in the study and is uniform across all four products. It cannot be used to differentiate between the two verification stages, because the high proportion of fixations on the chosen alternative in stage 3a may be an artifact of the instruction to look at the product.

Single eliminations are fixations that are both the first and last look at an alternative. In other words, they identify the alternatives that are eliminated in a single fixation. Table 1 reports that 27 percent of the alternatives viewed in stage 1 are never fixated again, which shows that some eliminations, namely, quick ones requiring only a single look, occur quite early in the choice process. The rate of single eliminations is lowest for stage 2. This conforms to its evaluation role and provides further evidence that stage 2 concentrates on the more serious candidates for purchase. The increase in single eliminations in stages 3a and 3b means that some never-before-seen alternatives are considered in these stages and are quickly eliminated. This is consistent with their verification role.

*Other Characteristics: Acquisitions and Eliminations.* The proportions of fixations devoted to acquisition (the first fixation on an alternative) and to elimination (the last fixation) are shown in Table 1. Two aspects of these data seem noteworthy; the high proportion of acquisitions at the end of the choice (stages 3a and 3b) and the proportion of eliminations at the beginning (stage 1).

The acquisition rates of stages 3a and 3b amount to roughly a quarter to a third of all their fixations. They are roughly equivalent to fixating one never-before-seen alternative during stage 3a and fixating another one-and-a-half such alternatives in stage 3b. Thus, a sub-

---

7It is worth noting here that a one-to-one correspondence between the alternative being fixated and the one being considered cannot always be presumed. It is possible that the most frequent violation occurs when two products are being considered, with one fixated and the other held in short-term memory. Although we believe that stage 3b's large drop in fixations on the chosen alternative validly reflects a change in the nature of the processing, we cannot rule out the possibility that the chosen alternative is being accessed from memory.

---

\[copyright \text{© 1994. All rights reserved.}\]
comparisons include the brand-size eventually chosen at a rate significantly greater than chance.

Differences across Product Categories?

Does the pattern of results just described hold uniformly over all three product categories (ketchup, ketchup, and peanut butter), or are these results limited to only one or two categories? A two-way ANOVA was performed, with stage and product category as the factors, for each of the 10 characteristics in Table 1 (excluding thin). Only one of the 10 product main effects was significant at \( p < .05 \), the proportion of different brands fixed. No interaction between product and stage was significant. The paucity of significant effects suggests that the one observed may be a false alarm. Certainly, the overall pattern of results across the ten characteristics convincingly points to uniformity of the stage differences over the three product categories.

Differences within Stage 2

The identification of the stages left a large undifferentiated stage 2, on average 58 percent of all fixations and 70 percent of the total time. We could find no way to use the eye fixations to further partition this stage into discrete substages. This is not to say that there were no differences in fixations across stage 2, but rather that they took the form of trends rather than exhibiting any clear discontinuities.

To reveal these trends, stage 2 was divided into four segments as equal as possible in the number of eye fixations (because individual fixation durations were not calculated, any partition of a stage had to be based on fixations and not on time). The last seven process characteristics reported in Table 1 were calculated for each quartile and are reported in Table 2. The first four measures from Table 1 were dropped because their analysis (1) made no sense (number of trials with a nonzero stage), (2) could only reflect the equal division of stage 2 into four quartiles (exactly for number of fixations and approximately for total time), or (3) was impossible (individual fixation durations were not available). Each of the remaining seven measures in Table 2 was subjected to a one-way ANOVA that tested for significant differences among quartiles. The resulting F-statistics are provided in Table 2. All seven characteristics yield a significant value of \( F \) except one, the proportion of different brands fixed (\( F(3, 512) = 2.25, p = .082 \)).

The observed trends show no surprises. They are monotonic, or nearly so, with one exception. There is a low initial value for the proportion of fixations on single eliminations (.13). We note that the reason this latter value is low is not because there are few acquisitions (the next row of Table 2 indicates the opposite). Instead, we speculate that it might be due to the need for an initial period of time in stage 2 to form a basis for evaluation that can then support the elimination of newly encountered alternatives. The six (nearly) monotonic trends, with the proportion of fixations on single eliminations excluded, tell a simple story. The two measures of breadth of search decrease over time, while the single monotone measure of focus of search increases. Similarly, acquisitions decline while eliminations increase, and the use of direct

\[ \text{Table 2: Process Characteristics by Quartiles of Stage 2} \]

<table>
<thead>
<tr>
<th>Process characteristic</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breadth of search:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of different alternatives fixed</td>
<td>4.09</td>
<td>3.56</td>
<td>3.62</td>
<td>3.16</td>
<td>4.16**</td>
</tr>
<tr>
<td>Number of different brands fixed</td>
<td>.50</td>
<td>.47</td>
<td>.46</td>
<td>.43</td>
<td>2.25</td>
</tr>
<tr>
<td>Focus of search:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of fixations on the chosen alternative</td>
<td>.16</td>
<td>.20</td>
<td>.26</td>
<td>.38</td>
<td>15.61***</td>
</tr>
<tr>
<td>Proportion of fixations on single eliminations</td>
<td>.13</td>
<td>.20</td>
<td>.16</td>
<td>.08</td>
<td>4.79**</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of fixations on acquisitions</td>
<td>.44</td>
<td>.39</td>
<td>.33</td>
<td>.15</td>
<td>6.37***</td>
</tr>
<tr>
<td>Proportion of fixations on eliminations</td>
<td>.26</td>
<td>.36</td>
<td>.39</td>
<td>.52</td>
<td>16.19***</td>
</tr>
<tr>
<td>Proportion of fixations in comparisons</td>
<td>.40</td>
<td>.48</td>
<td>.54</td>
<td>.74</td>
<td>17.75***</td>
</tr>
</tbody>
</table>

* \( p < .05 \).
** \( p < .01 \).
*** \( p < .001 \).

\[ \text{The proportion of brands fixed was larger for ketchup (69) than for applesauce (61) or for peanut butter (60). We could find no substantive cause for this pattern but do note that, while each of the latter two categories contained six different brands, ketchup had only five. Although we believe that the scope of brand search is more appropriately measured by the proportion than by the number of brands acquired, we performed an additional two-way ANOVA on the latter. For the number of different brands fixed, the significant effect of the product category disappeared (F < 1.00).} \]
comparisons increases over time. These trends are exactly what would be expected if stage 2 contained the real work of evaluative inference, namely, comparison among alternatives and elimination until a single best alternative had been identified.

The Effect of Purchase Frequency

It is natural to ask whether there were differences in the choice processes of shoppers who purchased a product category more (vs. less) often and so were more familiar with it. Product familiarity can be a complex issue (Alba and Hutchinson 1987). However, we chose to assess only purchase frequency rather than other aspects of product familiarity, because it measures the repetition of the same choice process we are tracing. In addition, for our sample of shoppers and product categories, purchase frequency exhibited a large range, from less than once per year to as often as once per week ($\bar{x} = 13.3$ purchases per year; $SD = 14.0$).

To test for an impact of purchase frequency on the choice process, we first identified those characteristics of the choice process for which a directional prediction could be made. Greater purchase frequency was expected to reduce both work (total time and number of fixations) and breadth of search (number of different brands and sizes fixated) and to increase the focus of search (proportion of fixations on the chosen alternative and on single eliminations). Each of these six characteristics was regressed on the (natural) logarithm of purchase frequency separately for each stage for a total of 24 slope coefficients.\(^9\) No significant differences ($p < .05$; one-tailed $t$-tests) were observed, except in stage 2, and then only for the four work and breadth measures.\(^10\) Similar results are reported by Dickson and Sawyer (1990) and by Moore and Lehmann (1980). For each of the two focus measures, the slope was positive, as predicted, but failed to reach significance ($p > .10$ for both). Thus, the predicted direction of the purchase frequency's impact on stage 2 was confirmed for all six characteristics, with four reaching statistical significance. For completeness, we performed the same regression analyses for the four process characteristics not expected to be affected by purchase frequency. None of the 16 slope coefficients (four characteristics by four stages) significantly differed from zero ($p > .05$; two-tailed $t$-tests).

The absence of an effect of purchase frequency on the characteristics of stages 3a and 3b was not unexpected. These stages may be more sensitive to the immediate results emerging from stages 1 and 2 than to something as global as purchase frequency.

But, by the same reasoning, stage 1 might well be expected to exhibit the greatest impact of purchase frequency. Yet not one characteristic showed a significant effect. We considered one last possibility, that the impact of purchase frequency on this initial stage may not be as sensitive as a linear relation with the appropriate process characteristics but instead may be limited to whether the stage was needed at all. To investigate this less refined hypothesis, we asked whether the 11 cases without stage 1 differed in purchase frequency from the 127 with stage 1. The respective means were 22.1 and 12.6 purchases per year. A $t$-test of the log of purchase frequency showed this difference to be significant ($p < .05$; two-tailed). Thus, stage 1 is affected by purchase frequency, but this effect is limited to its occurrence and not to more refined characteristics like its length and search characteristics. For completeness, we performed the same test for stages 3a and 3b. There was no significant difference in purchase frequency when these stages were present alone or in combination (for all three, $p > .40$).

The Validity of Stages or Time Trends

We began our empirical report by describing four stages and our interpretation of their respective functions. Although various characteristics of these stages have been described, we have not formally tested whether the choice process is segmented into distinct stages separated by clear discontinuities or is a more or less continuous time trend. The latter hypothesis asserts that the same pattern of results, such as an increasing focus on the chosen alternative, can be accounted for by continuous trends just as well as by the discontinuous shifts that define the stage boundaries.

The required test was based on the following logic. If the stages are genuine, then the abrupt changes in process characteristics that occur at the stage boundaries ought to be greater than the corresponding changes at the same times in a continuous trend. To translate this logic into a statistical test, we constructed time segments to match the mean durations of the four stages. However, these segments had to be based on the number of eye fixations rather than on exact times (because individual fixations durations were not available). Thus, the resulting segments are an approximation to time on the basis of the number of fixations.

The specific process of construction was as follows. We divided each trial into four segments that matched as closely as possible the mean proportion of fixations in each stage. For instance, since on average the last 8 percent of fixations were devoted to stage 3b, the last 8 percent of the fixations of every trial were assigned to

---

\(^9\) We thought that a logarithmic transformation better captured the increase in familiarity gained as purchase frequency increased from less than once per year ($<1$) to once per week ($\geq 1$) relative to the linear increase implied by no transformation at all. We had no independent measure of familiarity, however, so this transformation is based only on our judgment.

\(^10\) Of the 30 slope coefficients in stages 1, 3a, and 3b, one did reach statistical significance. The mean duration of stage 3a's fixations was positively related to purchase frequency. We could see no compelling reason for this one exception and, thus, dismiss it as likely to be a false alarm.
a segment corresponding to stage 3b. Furthermore, the 58 percent of each trial allocated to stage 2 was divided into four equal quartiles, which paralleled the analyses reported in Table 2. Thus, each choice was partitioned (by eye fixations) into seven segments that approximated, via the number of eye fixations, the durations of the four stages and of the four quartiles of stage 2. We now had two parallel partitions of the choice process, one by stages, the other by fixation-based segments approximating time.

To test for the validity of the stage structure, we examined the same seven process characteristics reported in Table 2 for differences across the three stage or segment boundaries. These were the boundaries between stage 1 and the first quartile of stage 2 (time segments 1 and 2), between the fourth quartile of stage 2 and stage 3a (time segments 5 and 6), and between stages 3a and 3b (time segments 6 and 7). To simplify the reporting of results, we exclude the 5 of the 21 possible tests for which we could not predict the direction of the shift across the boundary. Also excluded are the five remaining tests involving characteristics for which the definition of stages would create a natural advantage for an observed change across stage boundaries. For instance, the proportion of fixations on the chosen alternative is elevated for stage 3a, because it is always terminated by the required announcement of the chosen brand and size.

For each boundary, Table 3 reports the values for the remaining process characteristics on either side of the boundary and the shift or difference between these values. The key test of the power of the stage structure over a simple trend structure is whether the stage shifts are greater than the corresponding trend shifts. The two shifts and the critical difference between them are reported in Table 3 along with the results of the corresponding t-tests. These tests are one tailed except for the time trend shifts, which were two tailed to respect the weakness of the directional predictions across trend boundaries.

For the stage 1–stage 2 boundary, the stage shifts are greater than the trend shifts for all four tested characteristics. Three of these differences are statistically significant, while the one remaining, the proportion of different brands fixated, yielded \( p = .053 \). A similar pattern of results holds for the stage 2–stage 3a boundary. For all four tested characteristics, the observed stage shift significantly exceeds its corresponding trend shift. Note that the latter tend to mimic the direction of the stage shifts, although they are usually smaller in magnitude and, in a few cases, even violate the predicted direction. Finally, for the boundary between stages 3a and 3b the situation is reversed. None of the three predicted changes yielded a significant difference between the stage and trend shifts.

In sum, the results for stage 1 versus stage 2 and for stage 2 versus stage 3a support the validity of the stage distinctions. In contrast, for stages 3a and 3b the hypothesis of a continuous time trend cannot be disconfirmed in favor of a stage structure. Thus, for the three stages taken as a whole (i.e., considering 3a and 3b as a single stage), the tests confirm the value of the stage structure in differentiating subprocesses of the overall choice process.

Is Stage 1 Screening?

We began with two plausible interpretations of stage 1. Screening is the systematic acquisition of alternatives in order to eliminate most products as noncompetitive and to pass on the remaining alternatives to a second stage of comprehensive evaluation and final choice. This is the usual first stage of the commonly proposed two-stage framework. However, given the subjects' familiarity with the products and task used here, a rigorous, comprehensive screening of all 16 alternatives should not be expected. Orientation is an organizing overview sufficient to locate enough sizes and brands to prepare for subsequent evaluative processing. Such an overview might include some eliminations, as with screening, but it will do so more opportunistically than strategically.

These competing interpretations of stage 1 are not easily distinguished. Clear manifestations of the key difference, whether or not there is a systematic intent to eliminate most alternatives, are blurred by implicit eliminations based on familiarity and by opportunistic eliminations when only orienting is intended. Nonetheless, we analyze two process characteristics that provide some discrimination between the two interpretations, the mean fixation duration and the proportion of single eliminations (Table 1).

First, each screening fixation involves acquiring an alternative and evaluating it thoroughly enough to decide between the alternative's final elimination and its acceptance as a serious contender (for stage 2's evaluation). In contrast, an orientation fixation mainly acquires information about sizes and brands. The greater processing burden of screening implies a longer mean...
fixation duration. The durations of stage 1, .64 seconds on average, are quite brief, which would seem to favor orientation over screening. However, subjects' familiarity with the product categories may have greatly reduced the additional time per fixation needed for evaluative processing. If so, a brief duration like .64 seconds might only negligibly favor orientation over screening. To get some sense of this reduction in processing time due to familiarity, we returned to the test of the relation between (log) purchase frequency, our index of familiarity, and the mean fixation duration in stage 1. If that stage is devoted to screening, we expect a negative relation between purchase frequency and fixation duration, especially over the wide range of purchase frequency for our subjects. Recall, however, that there was no significant effect of purchase frequency. Thus, the fixation data seem to favor the orientation interpretation of stage 1 over the screening interpretation.

Second, screening implies elimination (i.e., not re-fixating most of the alternatives seen). However, the proportion of elimination fixations (.27) seems too low. Instead, the majority of the alternatives seen in stage 1 (73 percent) are fixated at least once more in a later stage. Again, note that product category familiarity may make the low elimination rate compatible with screening. Because subjects are familiar with various brands and sizes, they can avoid the least attractive ones with very few fixations and find the likely choice candidates
instead. As above, we tested this possibility by regressing the proportion of fixations on eliminations in stage 1 on (log) purchase frequency. As before, there was no significant relation, which disabled the claim that familiarity facilitates a low rate of eliminations.

Even though both measures point to an orientation interpretation over screening, they are far from conclusive. Thus, our interpretation of the goal and activities of stage 1 as being devoted more to orientation than to screening is tentative on the basis of weak empirical evidence that, although favoring orientation, is not inconsistent with either interpretation.

Two Kinds of Verification or One?

Are the verifications in stages 3a and 3b qualitatively different? With one exception, the evidence reported above, especially in Table 3, suggests not. That exception is the longer durations of stage 3a (1.01 seconds) relative to 3b (.61 seconds). It is unfortunate that, because individual durations were not assessed, we cannot know whether the shortening of fixation durations may have occurred in a roughly continuous manner throughout stages 3a and 3b rather than in an abrupt shift at their boundary. Thus, the existing evidence provides no support for a qualitative difference in the natures of stages 3a and 3b.

DISCUSSION

Our intent was to trace the choice process for nondurables (in one particular task environment) with the goal of identifying process characteristics that might guide a theory of the choice process within the constructivist framework.

Stages

The presence of distinct stages was identified from eye fixation data and was verified against the competing structure of continuous trends over time (see especially Table 3). This is our main finding. Two aspects of this result merit further consideration: (1) Why did we find three stages while the generally accepted view contains only two? (2) What implications about real supermarkets can be drawn from our laboratory environment?

How Many Stages: Two, Three, or More? We found three stages, uniform across three product categories and distinguished reliably from simple time trends. Yet almost all other studies report support for a two-stage structure. (An exception is Olshavsky [1973], but in a somewhat different situation.) This discrepancy can be explained by two characteristics of our study, the task environment and the process-tracing method.

First, consider briefly the typical task environment in which two stages have been found, and contrast it to our laboratory simulation of supermarket shelving. The typical task environment is a brand-attribute matrix, often with no brand names and with unspecified ranges of attribute values. In this situation, subjects cannot rely on recalled knowledge to efficiently trim the processing burden—for instance, by eliminating some brands or all of one size. Instead, it might seem irresponsible not to look at all alternatives, at least if there are no more than 16. Also, the separate presentation of individual attributes may convey an implicit demand for thorough processing. More generally, if the choice process is adaptive, as we believe, the different task environments in which the two-stage and three-stage structures have been found at least partially explain the different findings themselves.

The immediate implication of such adaptivity to the specific task environment is the absence of any incompatibility between the common two-stage structure and our observation of three stages. In terms of these stages interpreted goals, it is fully possible that screening, as well as evaluation, occurred in our stage 2. Similarly, both the orientation and verification goals can, in principle, be embedded in a two-stage structure. Thus, we see no fundamental conflict between our results and the two-stage view. Instead, the underlying theme of adaptivity to the choice environment embraces and unifies both.

The second reason for our observing a three-stage structure in lieu of the usual two stages lies with the methodology used to trace the choice process. Recall how the first and third of our stages were not always present. Eye fixations accurately reveal this, whereas verbal protocols and other process-tracing methods, which are always less detailed and often less complete, might not (for a comparative analysis of such methods, see Russo [1978]). Verbal protocols in particular risk incompleteness in reporting brief operations, such as those in stages 1 and 3b where short fixation durations suggest that the operations underlying them may be quite brief. Finally, given that the stages can start at widely different times for different choices and with some delay before speaking begins, imprecision in timing further handicaps the reliable identification of stages by verbal protocols.

In sum, the methodology may matter as much as adaptivity to the task environment. Eye fixations offer detail, completeness and alignment in time with the underlying cognitive process that may have been essential in revealing the three-stage choice process that is our fundamental finding.

Laboratory versus Supermarket. Despite our attempt to simulate the supermarket shopping environment (e.g., real products and prices and similar shelf arrangements), the laboratory setting was different in a number of important respects. These include, hypothetical purchases, the presence of an experimenter, and the generation of a verbal protocol. Given the adaptivity of the choice process to the particular task environment
we must expect substantial differences in the resulting process. 

There is one empirical measure from both our laboratory and typical supermarkets that enables an assessment of how closely the former resembles the latter. This is the total decision time. Over several studies, the mean decision times observed (unobtrusively) in actual supermarkets ranged from 12 to 18 seconds, which is well below our mean of 29.6 seconds (Cobb and Hoyer 1985; Dickson and Sawyer 1990; Hoyer 1984; Kendall and Fenwick 1979). Some of the prolongation of our decision time is due to inconsequential factors like the need to weigh our times by purchase frequency to match the times observed in supermarkets. This will lower our mean time, because the more frequent purchasers decide faster. Nonetheless, much of the longer laboratory time reflects genuine differences in the two choice processes. Overall, we believe that subjects’ choice processes were more deliberate than when shopping normally in their own supermarkets. When our distribution of response times is compared to that of Kendall and Fenwick (1979); the supermarket study that reports the most complete distribution, the sole difference is fewer brief choices. For instance, 56 percent of their actual shoppers took no more than 8 seconds to complete the task, a speed achieved by only 7 percent of our subjects. (In a noteworthy agreement with Kendall and Fenwick [1979], Hoyer [1984] reports a median total time of 8.49 seconds.) However, beginning with choice times of 13 seconds, the two distributions are remarkably similar in shape. We believe that brief decision processes were underrepresented in our study, because they may have made some subjects feel like they were not cooperating fully or were not giving a rational enough picture of their own decision making. Biehal and Chakravarti (1986) reach a similar conclusion, reporting that “subjects gave concurrent verbal protocols and may have tried to appear rational in their choices” (p. 399).

Given such a clear empirical difference between the laboratory and the supermarket, we consider the specific ways in which the three-stage picture that emerged from our eye-fixation records might fail to reflect the actual choice process in supermarkets. Stage 1 is clearly problematic. Shoppers should feel a greater need to orient themselves to our novel product display than do they in their own supermarkets (for a related result, see Park, Iyer, and Smith [1989]). The relative unfamiliarity with our laboratory display of products may have prompted or at least prolonged an orienting stage. Stage 2 might not have been dominated by direct pair or triple comparisons in the natural environment, where consumers are not accompanied by researchers (which might possibly prompt greater thoroughness of their choice process). Similarly, the use of verification, in either stage 3a or 3b, may have been magnified by some subjects’ desire to appear to be careful shoppers. For none of the three stages can these doubts about their generalizability be verified until we trace the choice process in real supermarkets. What we hope that we have accomplished, however, is to add to the many studies using a matrix of product attributes and brandless alternatives a choice environment that moves a step closer to real supermarkets.

Toward a Process-based Theory of Choice

The overarching conclusion from our results is that the planned, comprehensive analysis of choice alternatives implied by the utility maximization models is, once again, disconfirmed in favor of the simplifying, adaptive and constructive process described by Bettman (1988) and others. The adaptivity based on product familiarity or knowledge is noteworthy; but the claim that the choice process is constructive is neither new nor helpful. Instead, what have we learned that might guide theory development within the constructive framework?

Progress within Stages. Within a stage, the process is one of continual movement toward the completion of that stage’s goal. Thus, verification is not necessarily completed by the instructed announcement that ends stage 3a but sometimes continues through stage 3b (31 of 66 choices). Indeed, the announcement seems like an exogenous intrusion into a progressive process of verification. Similarly, stage 2 exhibited more or less continuous trends in, for example, direct comparisons and focus on the chosen alternative, without being able to be partitioned into distinct subphases. This view of roughly continuous progress within stages raises two questions for future research: (1) what drives the within-stage activity and (2) how progress is deemed sufficient to move on to the next stage.

The Role of Operations. Recall that stages were identified by overt behaviors like eye fixations but were interpreted by their goals: orientation, evaluation, and verification. These, in turn, correspond to characteristic operations. We see these operations, rather than stages, as the building blocks of the choice process. They reflect the subtasks that must be accomplished.

One of the less expected findings of our experiment is the overlap of some of these operations across stage boundaries. For instance, information acquisition, as indicated by a first eye fixation, occurs during verification (see also Biggs and Mock 1983), and elimination, as signaled by a final fixation, occurs during orientation. Apart from complicating the identification and even the existence of distinct stages, the overlap of operations across stages signals an important characteristic of the choice process. The construction of that choice process is optimized globally (i.e., across all stages) rather than locally within each stage. Thus, if stage 1 is truly devoted to orientation but an easy elimination can be accomplished, this opportunity will be seized. Why postpone all eliminations to stage 2 when time and effort can be saved by finishing a few in stage 1?
One implication of a complex relation between stages and specific cognitive operations is that identifying macrocomponents of the choice process, such as stages, from microcomponents, such as acquisitions, comparisons, and so on, will not be a simple task. Instead of being able to link a single characteristic operation to each stage or other macrocomponent, a deeper analysis, possibly based on unique patterns of operations, may have to be performed. In sum, although stages are an important part of the choice process, operations may be more fundamental building blocks of a theory of consumer choice.

**Ending a Stage.** The view of continual progress toward a stage's characterizing goal (on the basis of fundamental operations) raises the need to decide when to terminate a stage. For instance, when do subjects determine that enough overview (stage 1) has been achieved to begin a comparison-based evaluation (stage 2)? Is it when they have located the size they usually buy, have found two alternatives worthy of direct comparison, have overloaded short-term memory and need to winnow (via direct comparisons), or have developed a mental representation of the stimulus display that is sufficient to efficiently locate the alternatives worthy of considered evaluation? Similarly, we can ask how subjects (1) decide to terminate the direct comparisons of stage 2, (2) reach a sufficient confidence to announce their preferred brand-size, and (3) end verification (by turning away from the display) and with it any chance to reverse the announced choice. In sum, the rather homogenous nature of the activities within a stage suggests that a key element of any process theory is a set of stopping rules.

**Decision as Discrimination.** What characterizes the overall structure of the choice process besides the three stages with their particular goals and operations? Our data suggest a gradual process of discrimination among the alternatives (Hagerty and Aaker 1984; Svenson 1992). During the evaluation process, one alternative emerges as at least tentatively superior. Continued evaluations confirm the superiority of an alternative or possibly reject it in favor of some other alternative. Eventually, enough confidence in this discrimination is accumulated that the identified alternative is announced as the choice, all verification ceases, and the announced alternative is accepted as final. This discrimination process extends across stages, although each stage may contribute to it in different ways.

It also suggests an alternative to the conventional interpretation of an elimination as a considered and final conclusion. Instead, evaluating a brand-size may result in a decision not to eliminate it; yet later, when one or more clearly superior alternatives are encountered, a decision is implicitly made never to return to the first alternative. Of course, this amounts to an elimination and is manifested as never fixating the alternative again; but it has neither the directness nor the timing of a classic elimination. In sum, both across and within stages, choice is an imperfect discrimination process based on partially clarified preferences and on tentative inferences.

**Conclusion and Prospect**

The goal of our work was to contribute to the empirical foundation of a veridical model of the choice process for common nondurables. Our study suggests at least two elements of an adequate process model. First, stages require stopping criteria. In contrast to a planned process, like an individual evaluation of every alternative in an initial screening stage, an adaptive process must have stopping rules contingent on the progress that has been made. Second, we must identify the basic processing operations or smallest cognitive units that, for instance, yield an inference. (They are more macro and are more likely to be available for conscious reporting than elementary information processes, although these too may eventually be needed by the model; see Payne, Bettman, and Johnson [1993]). Although we are a long way from an adequate theory of the choice process for common nondurables, these two topics may help guide further work. Equally important, the use of methods like eye fixations offer the prospect of extending the empirical base on which to construct an acceptable theory.

**APPENDIX**

Analysis of Eye Fixations

The eye fixation sequences were determined from a videotape recording of subjects' faces through a one-way mirror as they made their purchase decisions. The videotapes of the 138 trials (recall that three of the original 141 were lost because of equipment failure) were coded independently by two judges working side by side. Two coders were not needed to ensure reliable identification of the fixated alternative. This identification should have been easy (and perfectly accurate) because of the large separation among the 16 containers and the prohibition against subjects' moving from a preset position in front of the display. (The experimental assistant who video recorded subjects' faces from behind the one-way mirror had only to make an occasional adjustment for a very tall or short subject, something that could be anticipated by observing the subject's height as she arrived at the laboratory.) Two coders were used because the identification of fixated alternatives was tedious, and working jointly succeeded in reducing fatigue and boredom.

The coded fixations were grouped into sequences that corresponded to an initial estimate of possible stages. The duration of each stage was calculated by both coders working simultaneously. They used stopwatches and frame counting to estimate durations. The primary difficulty was determining on which frame the eye began.
to move or ceased moving. This procedure was time-consuming and is the main reason why we made no attempt to measure the duration of each individual eye fixation.

Comparisons between Alternatives

Pair and triple comparisons were identified from the eye-fixture sequences following the scheme of Russo and Rosen (1975). An alternating sequence of three or more fixations between two products (X-Y-X . . .) is considered a pair comparison. The comparison is said to be weak if it contains exactly three fixations and strong if it contains four or more. An alternating sequence of the form X-Y-Z-X-Y . . . is considered a triple comparison. It is weak if it contains exactly four fixations and strong if it contains five or more. Note that a triple comparison cannot be of the form X-Y-X-Z . . ., which is defined as a weak pair comparison between X and Y. In other words, pair comparisons take precedence over triples. Sometimes pair and triple comparisons overlap (e.g., X-Y-X-Y-Z-X-Y). According to our definitions, the third and fourth fixations belong to both a pair and a triple comparison. Because pairs take precedence, these two fixations are assigned to the pair (unless otherwise indicated). The triple is still said to occur, but only the last three fixations are attributed to it.

Validity of Comparisons

The assumption that specific fixation patterns indicate the mental process of direct comparison among alternatives requires validation. Using a verbal protocol to test this assumption, Russo and Rosen (1975) observed a minimum confirmation rate of 69 percent for weak pairs. That is, for 69 percent of pairwise alternations of exactly three fixations, a comparison process was reported in the corresponding verbal protocol. Thus, some of the X-Y-X sequences may have been falsely identified as representing pair comparisons.

Our subjects’ verbal protocols were not sufficiently detailed to enable a similar validation of the fixation alternations. (Russo and Rosen [1975] used a retrospective protocol prompted by a replay of the eye-fixture sequence to generate a protocol that was much denser and more apt to describe processes like comparisons; see Russo et al. [1989]!) To further compound the problem of validation, we were suspicious of one class of weak pairs. The sequence X-Y-X frequently occurred when the transition from X to Y reached an edge of the 4 × 4 product display. The refection of X may have represented only the reversal of a failed search process rather than a deliberate comparison between X and Y. Such “perimeter pairs” posed a special risk of false classification.

To help resolve the validity issue, we conducted two tests. First, we reasoned that noncomparison alternations would represent search and other functions in which the fixated brand-sizes are likely to be adjacent. In contrast, a genuine comparison is driven less by physical contiguity and is more likely to take place between separated alternatives. For each transition within pairwise and triple alternations, we measured the physical distance between the fixated items. We expected that genuine comparisons would exhibit transition distances greater than the benchmark of transitions between “single” fixations (i.e., those not assigned to any comparison). The mean transition distances are reported in Table A1 along with the frequencies of the various comparisons.

These data exhibit the expected pattern with the transitions between single fixations covering the shortest distance. A one-way ANOVA tested for significant effects. Perimeter pairs were excluded from this analysis, because they have zero variance. According to the results of a Duncan test, strong pairs and both triples had significantly more distant transitions than did singles. Replicating Russo and Rosen (1975), the weak pair comparisons achieved the lowest validity (as indexed by transition distance).

---

**TABLE A1**

<table>
<thead>
<tr>
<th>Fixation pattern</th>
<th>Frequency</th>
<th>Total number of eye fixations</th>
<th>Mean transition distance $^a$</th>
<th>Proportion of fixations on chosen alternative $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single fixation</td>
<td>1,998</td>
<td>1,998</td>
<td>1.29$^{a,b}$(1,620)</td>
<td>.091$^*$</td>
</tr>
<tr>
<td>Perimeter pair</td>
<td>109</td>
<td>326</td>
<td>1.00(198)</td>
<td>.166</td>
</tr>
<tr>
<td>Weak pair</td>
<td>184</td>
<td>506</td>
<td>1.37$^{a,b}$(387)</td>
<td>.320$^{**}$</td>
</tr>
<tr>
<td>Strong pair</td>
<td>106</td>
<td>511</td>
<td>1.43$^{a,b}$ (440)</td>
<td>.311$^{**}$</td>
</tr>
<tr>
<td>Weak triple</td>
<td>121</td>
<td>318</td>
<td>1.50$^{a,b}$ (292)</td>
<td>.217</td>
</tr>
<tr>
<td>Strong triple</td>
<td>22</td>
<td>88</td>
<td>1.43$^{a,b}$ (85)</td>
<td>.204</td>
</tr>
</tbody>
</table>

Note.—Values that share a superscripted letter are not significantly different.

$^a$Parentheses show the number of transitions between the adjacent eye fixations on which the corresponding mean is based.

$^b$The proportions of fixations on the chosen alternative are based on the total number of eye fixations, which are reported in the second column of data.

$^*p < .05$

$^{**}p < .001$
A second validity test was based on the proportion of fixations on the alternative eventually chosen. If the single fixations are devoted mainly to search and to similar bookkeeping functions (Russo and Rosen 1975), they should exhibit a proportion of fixations on the chosen alternative not much above chance. Similarly, if the alternating pair and triple sequences represent genuine comparison, they should involve the chosen alternative more often than chance. The relevant proportions, reported in Table A1, were tested against the chance base rate. We did not use .062 (1 in 16) as the chance rate, since rarely were all 16 alternatives fixated. (On average, only 11.6 were, which implies an overall chance base rate of .08.) For testing purposes, the chance fixation rate was computed for each trial based on the actual number of fixated alternatives out of the 16 possible. An arcsine transformation was used to stabilize the variance of the proportions, and separate t-tests were performed on the transformed proportions for each comparison type. Because not every trial contained fixation patterns of each type, the degrees of freedom varied from 138 for singles to 22 for strong triples.

The results of these tests, which are reported in Table A1, show significantly greater fixation rates on the chosen alternative for three of the four comparisons (for strong triples, $p = .08$). Note that the highest rate belongs to weak pairs, which strongly suggests that they can reliably be interpreted as comparisons. On the basis of the combination of transition distances and fixations on the chosen alternative, we exclude perimeter pairs because they are likely to contain too many false positives but treat all other comparisons as acceptably valid.

[Received January 1993. Revised November 1993. Kent B. Monroe served as editor for this article.]

REFERENCES


Leong, Siew Meng (1993), "Consumer Decision Making for Common, Repeat-Purchase Products: A Dual Replica-
tion" Journal of Consumer Psychology, 2 (September), 193-208.


