An eye fixation analysis of multialternative choice

J. EDWARD RUSSO and LARRY D. ROSEN

University of California, San Diego, P. O. Box 109, La Jolla, California 92037

Eye fixations were recorded while subjects chose their most preferred of six used cars. Fixation sequences of the form X-Y-X- . . . were used to identify pair comparisons. This assumption was validated by verbal protocols and by a comparison between X-Y-X and X-Y-X-Y- . . . patterns. The results of three experiments showed: (1) that the multialternative choice process was composed primarily of pair comparisons; (2) that evaluative processing took place *only* during these comparisons; (3) that strategies for selecting the pairs were based primarily on information processing convenience; (4) that the sequence of pair comparisons was not used by the subjects as a sequential elimination process; and (5) that subjects changed their strategies to adapt to different task environments. These results have implications for current choice models and for the use of eye fixations in other complex tasks.

Most of the research on human choice behavior has been confined to the two-alternative case, in spite of the fact that most natural choice situations offer several alternatives (e.g., consumer choices). The lack of attention to the multialternative choice problem is especially noticeable with regard to empirical work. The main purpose of this research is to conduct a detailed empirical investigation of the multialternative choice process. This study focuses on two aspects of that process, pair comparisons (binary processing) and the use of similarity relations. Both of these phenomena are directly related to research on binary choices.

BINARY PROCESSING

The first question was: To what extent do subjects partition the multialternative choice task into a sequence of pair comparisons? There are convincing a priori arguments for the adoption of a binary processing strategy. The choice task is inherently one of comparison, and some kind of direct comparison mechanism ought to be employed. It is assumed, of course, that the alternatives are at least moderately complex and are not trivially distinguishable with respect to personal perference or "utility." Information processing considerations favor a direct comparison among only a few alternatives. Certainly a comparison set of size three or four would stretch short-term memory (STM) capacity to its limits. A comparison subset of just two alternatives would minimize the information processing burden. One pairwise strategy is to compare a given alternative only to the best of the preceding ones. If the given alternative is judged superior, it can become a new standard against which each of the remaining ones will be compared. This strategy, termed standard revision, is designed to

minimize the STM load by retaining one member of a pair comparison as part of the next comparison.

Binary processing is by no means the only method for performing a multialternative choice. A process based on "elimination by aspects" has been proposed by Tversky (1972). According to this strategy, the subject selects some valued aspect (dimension) of the alternatives and eliminates all those alternatives that are found to be inadequate on that aspect. This process is repeated, i.e., another aspect is selected and another group of alternatives is eliminated, until only one alternative remains. Tversky has developed this strategy into a compelling mathematical model. The strategy of elimination by aspects may be contrasted with one of binary processing. The former strategy does not require direct comparisons among the alternatives. Rather, each alternative is judged against some elimination criterion. The two hypothesized processes are different enough so that a detailed explication of the choice process should reject at least one of them.

SIMILARITY RELATIONS

One way to investigate the tendency to use binary processing is to embed natural binary relations in the multialternative choice set. If pair comparisons occur at all, they should occur between those alternatives with a natural binary relation, one that can provide the inducement of easier processing. A highly similar pair of alternatives forms such a relation.

A basic phenomenon of binary choice behavior is that similarity facilitates choice. Holding utilities constant, the probability of a correct choice is an increasing function of the similarity between the two competing alternatives. This effect of similarity has been demonstrated for a wide variety of stimuli, including color patches (Krantz, 1964), geometric figures (Tversky & Russo, 1969), famous people (Rumelhart & Greeno, 1971), and gambles (Tversky, 1972).

If a similar pair is included in a multialternative choice, is the facilitative effect of the similarity relation

sufficient to induce a pair comparison? If binary processing will not take place on such pairs, it is unlikely that it will occur under any general circumstances.

THE USE OF EYE FIXATIONS

The proposed investigation requires process tracing data, such as an eye fixation sequence or a verbal protocol. Eye fixations were selected because they are unobtrusive, detailed, and difficult to misrepresent. The use of eye fixations in this research, however, differs from the typical use of eye fixations. First, the relevant data are sequences of fixations rather than summary measures like total number of fixations or average duration. In complex cognitive tasks, the meaningful units of behavior may subtend more than one fixation, so the unit of analysis must be extended to subsequences of fixations. A second characteristic of the present use of eye fixations is their interpretation without external aid. Typically, the interpretation of eye fixation sequences has been assisted by a concurrent verbal protocol (Winikoff, 1967), by the experimenter's knowledge of the task (DeGroot & Jongman, 1973; Tichomirov & Posnyanskaya, 1966) or by the experimenter's subjective abstraction of the observed fixation sequence (Noton & Stark, 1971). A goal of this research is to demonstrate that eye fixations can be used without recourse to either external information or subjective interpretation.

In order to accomplish this goal, two conditions must be satisfied. First, the process of interest, such as a pair comparison, must manifest a characteristic sequence of fixations. The second requirement is the minimization of processing differences caused by different retrieval times from long-term memory (LTM). To minimize these effects in a preferential choice task, the subject should be very familiar with the stimuli and with the values attached to them. When this is true, the processing sequence should reflect primarily the computational strategy that the subject chooses to apply to the retrieved preferences. It is these computational strategies that are of interest, not the retrievability of preference values for individual alternatives.

The occurrences of pair comparisons will be identified from fixation sequences of the form X-Y-X . . ., i.e., an alternation of three or more fixations. This assertion follows directly from the structure of the information processing sequence for a pair comparison. A similar, but more general assumption was made by Winikoff (1967). He defined "eye movement aggregates" as sequences of the form A-B-A and A-B-A-B, where A and B could be either single fixations or groups of fixations previously aggregated. However, Winikoff never interpreted the fixation aggregates as pair comparisons. Alternating subsequences have also been used to identify pair comparisons in pictorial matching experiments (Drake, 1970; Nodine & Steuerle, 1973; Vurpillot, 1968). This usage, however, differed critically from our own, because the locations of comparable pairs were predetermined. In the research to be reported here, no a priori information is used to determine a pair comparison. Any alternation of at least three fixations defines one.

METHOD

Stimuli

Each stimulus was a choice set containing six used cars. The subjects were instructed to choose their single most preferred automobile from the presented group of six. Each used car was described only by three pieces of information: make, year, and mileage. Pictures of the cars were not used. The six offerings in each set were selected from a total list of 240 possible used cars, every combination of eight makes, six years, and five mileage levels.

Two used cars were said to be similar if they were of the same make, e.g., two Datsuns; and two cars of different makes were said to form an independent pair. This definition of similarity took advantage of the predominant role of the make dimension and avoided the difficulties associated with constructing and verifying a general measure of similarity.

Each stimulus was a 2 by 3 rectangular array, with the six cars widely separated and with the three words defining each car closely packed. All stimuli were displayed on a storage CRT (Model 611, Tektronix) positioned about 40 cm in front of the subject.

Utility Collection

In order to construct choice sets with predetermined utilities for each subject, the utility of each car in the factorial set of 240 was measured. A subject first rated 28 cars on a 0 (worst) to 100 (best) scale. Choices were then made from among these rated cars, and the subject was required to reconcile any deviation between his ratings and choices. All 240 possible cars were then rated, and these ratings were subjected to an analysis of variance. The utility of each car was estimated directly from the three main effects (make, year, and mileage) and from any interaction terms that accounted for more than 1% of the variance. In order to assure that the used cars were effectively three-dimensional, it was required that each main effect account for at least 5%, but no more than 70%, of the variance in the ratings. Overall, about 40% of the subjects failed this criterion and were dropped. The only noticeable difference between the passing and failing subjects was that the former group seemed more knowledgeable about and comfortable with used cars as stimuli. Finally, 20 six-alternative choice sets were constructed from the estimated utilities in order to provide a test of the agreement between a subject's utilities and his choices. Any discrepancies were discussed with the subject, who then revised either ratings or choices. One cycle through this discussion-revision-test procedure was sufficient for all subjects.

In order to construct choice sets of equal difficulty across subjects, an estimate of each subject's choice discriminability was obtained. From subject's performance on the final test, that separation (in utility units) which produced a 75% rate of correct choices was estimated. The obtained values of this jnd unit varied between 3 and 6 utility units (on the 0 to 100 scale). Because this estimate of discriminability was not valid for cars of either very high or very low utility, only alternatives from the middle of the utility scale were used in the actual experiments.

Design

The design of all three experiments relied on pairs of choice sets matched in utility but differing in similarity. The six alternatives in each choice set contained zero, one, two, or three

similar pairs. Those with at least one similar pair were called similar choice sets, while those with only independent pairs were independent sets. All experiments consisted of 45 choice sets. In Experiments 1 and 2, 15 of these were independent and 30 were similar choice sets (10 each with one, two, or three similar pairs). In Experiment 3, 30 of the 45 sets were independent; the remaining 15 each contained one similar pair.

When only one similar pair was present, its alternatives were the two highest in utility. This was done to contrast the typical processing order with the deliberate organization of the processing around similarity relations. Typically, the two best alternatives survive until the end of a trial. If they form a similar pair, however, and if there is a strong effect of similarity, then these two alternatives should be compared first. Whenever a second similar pair was included, its alternatives were ranked third and fourth in utility.

The 45 choice sets were partitioned into 15 groups of 3 each, with the alternatives within each group matched in utility. That is, within each matched triple of choice sets, the three most preferred cars had (nearly) identical utilities, etc. This matching design permitted the effect of similarity to be separated from any effects due to utility differences.

Within all choice sets, the alternatives were equally spaced in utility, at intervals of approximately 1 jnd. No dominance relations were permitted. No used car was presented more than five times, including practice trials. Finally, both rank order of utilities and the location of similar pairs were counterbalanced over display position.

Apparatus

Eye position was detected by a photoelectric sensing apparatus (Russo & Mathews, 1975). The stimulus subtended visual angles of 20 deg horizontally and 8 deg vertically, with a minimum separation between alternatives of 5.5 deg; individual letters were .5 deg high. The large distances between alternatives both eliminated any opportunity for peripheral processing and also minimized errors in the identification of the object of regard. In order to restrict head movements, the subject's head was positioned on a bite bar.

Eye positions were sensed 100 times per sec and recorded by a PDP-12 computer (Digital Equipment Corporation). They were analyzed into a sequence of eye fixations by an algorithm based on a pretrial calibration and on a minimum fixation duration of 200 msec (Goode & Russo, 1970). During a trial the subject's eye position and calibration boundaries were displayed to the experimenter, who monitored the data. In this way, head movements or other sources of error could be detected and the trial discarded.

Subjects

Twelve subjects were used, four in each experiment. Subjects 1, 3, 6, 7, 8, and 11 were male; Subjects 2, 4, 5, 9, 10, and 12 were female. All subjects were naive student volunteers. The utility collection procedure did, however, provide every subject with extensive experience on both the stimuli and the task, prior to the recording of any eye fixation protocols. Subjects 1 and 2 also participated in a pilot study involving the collection of verbal protocols. All subjects were paid $1.88 per hour for their participation.

Subjects were selected on the basis of two criteria, an adequate familiarity with used cars, as described in the discussion of utility measurement, and the nature of their eyes. The eye position sensor required that the sclera (white part of the eye) be exposed below the iris and above the eyelid while the subject was looking straight ahead. Only 50% of potential subjects passed this criterion. One-third of the remaining were excluded because they needed eye glasses.

Procedure

The subject initiated each trial by a buttonpress. The six alternatives were first displayed one at a time for 1.5 sec each (Experiment 1) or .75 sec each (Experiments 2 and 3). The initial reading times were reduced to .75 sec when it became clear from Experiment 1 that 1.5 sec was more than sufficient. This exposure duration allowed the subject enough time to encode each alternative (make, year, mileage, and position) without permitting any significant amount of choice processing. The initial viewing procedure also guaranteed that the subject saw all six alternatives before the choice process was begun. Following the initial viewing period, all six alternatives were available, until the subject signaled the end of the trial. Each subject participated in one to three sessions of no more than 2 h each, held on consecutive days and within 4 days of the utility collection.

In Experiment 1 only, an overt elimination procedure was used. As soon as the subject was sure that any alternative would not be his eventual first choice, he was required to signal that elimination by pressing a button while fixating the alternative to be eliminated. The designated alternative was removed from the CRT display. The purpose of the overt elimination procedure was to highlight the important components of the multialternative choice process while reducing the number of fixations due to searching or rechecking.

Data Analysis

The main use of the sequence of eye fixations was the inference of a pair comparison from alternations of Length 3 or more. The alternations were partitioned into two groups, those of Length 3 and those of Length 4 or longer. Such alternations were labeled, respectively, weak and strong binary processing. There was little doubt that alternating chains of four or more fixations indicated a pair comparison; but the inference of binary processing from the pattern X-Y-X was less certain. It was possible, for example, that the subject refixated X only to check one dimension whose value had been forgotten.

For a preliminary test of the validity of the assumption that an X-Y-X pattern represented a pair comparison, verbal protocols on 15 choices were collected from each of the first two subjects. These data were retrospective and prompted; i.e., the verbal protocols were collected after the choice task had been completed, and they were prompted by a replay of the eye fixation sequence just exhibited. Each fixation sequence of the form X-Y-X . . . was identified, and the verbal protocol was examined for an admission that a deliberate comparison between alternatives X and Y had been made at that time.

There were 110 alternations of Length 4 or more, and for 105 of them (95%) a comparison was reported in the verbal protocol. As expected, this confirmation rate was very high, leaving little doubt that an alternation of Length 4 or more fixations could be assumed to indicate a pair comparison. Of the 102 triple alternations, 70 (69%) were explicitly identified as comparisons in the prompted protocols. This figure was a lower bound, since subjects could not recall what they had been doing during every fixation. Overall, 83% of the pair comparisons that were identified from the fixation sequences were explicitly confirmed by the verbal protocols. This percentage was considered as sufficient preliminary support for the assumption that pair comparisons could be identified from the eye fixation data.

RESULTS

Experiment 1

Two questions had been asked: Did pair comparisons (binary processing) play a major role in the multialternative choice process, and were similarity

Table 1
Proportion of Refixations in Binary and Ternary Comparisons

| Subject | Type of Binary Comparison | | | Ternary Compari-son X-Y-Z-X-Y . . . | Total Number of Refix-ations |
	Weak X-Y-X	Strong X-Y-X-Y . . .	Total X-Y-X . . .		
			Experiment 1		
1	.15	.66	.81	.015	1769
2	.20	.58	.78	.004	1314
3	.18	.57	.75	.005	2478
4	.23	.37	.60	.036	1757
Mean	.19	.54	.73	.015	
			Experiment 2		
5	.28	.17	.45	.010	621
6	.18	.54	.72	.014	1800
7	.30	.34	.64	.015	1486
8	.20	.59	.79	.019	1139
Mean	.24	.41	.65	.015	
			Experiment 3	.	
9	.26	.29	.55	.016	697
10	.25	.38	.63	.010	978
11	.31	.23	.54	.039	1199
12	.23	.52	.75	.004	1229
Mean	.26	.36	.62	.017	

relations involved in the selection of the pairs? To answer the first question, the proportion of refixations that were devoted to binary processing was calculated separately for the weak and strong criteria. These data, presented in Table 1, showed that for all four subjects, binary processing accounted for most of the multialternative choice process. On the average, 73% of refixations were part of pair comparisons.

Although the majority of refixations involved pair comparisons, it was still possible that subjects were employing a multiple comparison strategy, especially a comparison among three alternatives. To assess the use of ternary comparisons, a count was made of all sequences that had at least five fixations on three alternatives and that contained no embedded binary comparisons. Such sequences were of the form X-Y-Z-X-Y . . . . The proportions of refixations covered by these sequences are given in Table 1. The observed incidence of ternary comparisons was very low, averaging 1.5% over the four subjects. Only Subject 4 seemed to show a tendency to employ ternary comparisons.

The observed proportions of fixations were compared against chance base rates. In addition to pair comparisons and ternary comparisons, "single" fixations were defined as all refixations not included in either of the other two categories, i.e., the left-over refixations. An estimate of fixation frequency was computed assuming that the location of each fixation was randomly determined. The chance frequencies were 70.4% for singles, 26.4% for pair comparisons, and 3.2% for triple comparisons. The observed frequencies of pair comparisons exhibited many more fixations than could

have been expected by chance, 73% vs. 26%. Although an exact calculation was not possible, the chance rate of triples was estimated to be about 1.2%, given the high observed proportion of fixations in pairs. Using this rough value, Subject 4 did show a small but reliable tendency to employ triple comparisons.

The second question was whether similar pairs played a prominent part in binary processing. In answering this question, it was important to distinguish between a deliberate strategy of using similarity relations and a more haphazard encountering and use of similar pairs throughout the decision process. Consequently, it was decided to test only whether the first comparisons, the ones immediately following the initial viewing period, involved similar pairs. Specifically, the proportion of the initial comparisons between similar pairs, out of the total number of possible similar pairs, was calculated. For example, if there were two similar pairs, only the first two pair comparisons were examined to determine whether they had taken place between similar cars. These proportions are presented in Table 2. Of the initial comparisons, an average of 83% were on similar pairs. From those independent choice sets that were matched in utility, it was possible to compute a baseline against which to compare this value. Specifically, the baseline was the proportion of initial comparisons on those pairs in the corresponding independent choice set that were equal in utility to the similar pairs. Table 2 also shows the baseline proportions for all subjects. The contrast between similar and independent pairs made it clear that there was a deliberate strategy of organizing the initial

Table 2
Proportion of Initial Comparisons Between Similar Pairs or Between Matched Independent Pairs

| Subject | Choice Sets | | | | z Score of Difference |
	Similar		Independent		
			Experiment 1		
1	.73	(60)†	.33	(30)	3.65*
2	.80	(60)	.20	(30)	5.48*
3	.92	(60)	.17	(30)	7.12*
4	.88	(60)	.10	(30)	7.22*
Mean	.83		.20		
			Experiment 2		
5	.56	(39)	.26	(27)	2.45*
6	.76	(59)	.17	(30)	5.36*
7	.57	(60)	.10	(30)	4.24*
8	.37	(59)	.20	(30)	1.66*
Mean	.57		.18		
			Experiment 3		
9	.13	(15)	.11	(27)	.21
10	.27	(15)	.28	(29)	−.06
11	.27	(15)	.08	(26)	1.66*
12	.13	(15)	.03	(30)	1.27
Mean	.20		.12		

†The number of observations on which the adjacent proportion is based.
*p < .05

binary processing around similar pairs. This use of similarity relations was statistically significant for all four subjects, based on a test for the equality of two proportions. The results of this test are reported in Table 2.

Experiment 1 showed that multialternative choice consisted mainly of binary processing and that there was a systematic tendency to use similarity relations. All of this had taken place, however, in a decision environment more structured than usual, because overt eliminations were required. This task might have encouraged both binary processing and the use of similarity relations. In order to examine the choice process in a more typical environment, Experiment 2 was conducted. It replicated Experiment 1, but without the overt elimination requirement.

Experiment 2

The same analyses, for binary processing and for the use of similarity relations, were performed. These results are shown in Tables 1 and 2. The average percentage of refixations involved in binary processing dropped from 73% to 65%. This reduction still left binary processing as the main component of the multialternative choice process. Again, the proportion of refixations involved in ternary comparisons was minimal, averaging 1.5% over the four subjects. No subject had a proportion clearly above the chance value (approximately 1.2%).

The use of similarity relations remained well above chance (Table 2), with 57% of the initial comparisons on similar pairs. The use of similarity relations was again significant for all four subjects individually. Similarity use did, however, decrease from Experiment 1 (from 83% to 57%). This decrease was presumably caused by the removal of the overt elimination requirement. It should be noted, however, that Experiment 2 not only dropped the overt elimination requirement of Experiment 1, but also lowered the initial reading time for each alternative from 1.5 sec to .75 sec. It was possible that this latter change may have contributed to the decreased use of similarity relations.

There was much more variation across subjects in Experiment 2 than in Experiment 1, especially noticeable in the low proportions of the use of similarity for Subject 5 and Subject 8. Subject 5 was characterized by a low rate of binary processing, the lowest among all subjects in any of the three experiments. This explains the low sample sizes in Table 2. Subject 8 used the standard revision strategy. That is, he systematically compared each used car with the best of the preceding ones. Given such a heuristic, it was not hard to believe that Subject 8 would exhibit the highest use of binary processing but the lowest use of similarity relations.

Experiment 2 showed that even without overt eliminations there was a strong tendency to base a multialternative choice on a sequence of binary comparisons and to organize the initial comparisons around binary similarity relations. This use of similarity relations occurred, however, in an environment rich in them. In 45 choice sets, there were 10 each with 1, 2, and 3 similar pairs, a total of 60 such pairs. It was natural to question whether similarity would have been an organizing strategy for binary processing if similar pairs had occurred only infrequently. Indeed, if the clear inducement of similar pairs had been removed, would a binary processing strategy still have been predominant? Experiment 3 was designed to answer these questions. Only 15 similar pairs were distributed over its 45 choice sets.

Experiment 3

Table 1 shows that the proportion of refixations in binary comparisons remained high in Experiment 3. Even without many similar pairs, the overall mean was 62%. Binary processing still accounted for a clear majority of all refixations for each of the four subjects. One subject in this experiment (Subject 12) was, like Subject 8, a user of the standard revision heuristic. Her proportion of refixations in binary processing was noticeably larger than that of any other subject in Experiment 3.

The incidence of ternary comparisons was again low (1.7%), although Subject 11's proportion (3.9%) was reliably higher than chance. Thus, over all three experiments, 2 of 12 subjects showed a small but deliberate use of triple comparisons. These two subjects used about one triple comparison every three trials.

In the use of similarity relations, the results were exactly the opposite of Experiment 2. When few similar pairs were included, three of the four subjects showed no significant tendency to begin their binary processing on similar pairs (Table 2). The one subject (Subject 11) who did significantly use similarity showed an effect smaller than the lowest similarity user in the previous experiments.

Experiment 3 showed that subjects adapted their strategy to the characteristics of the stimulus environment. When that environment was known to have many similar pairs, all subjects took advantage of these easy comparisons. When similar pairs occurred only once every three trials, however, it was not worth the effort to search for them. A second conclusion from this experiment was that even without similar pairs to encourage pair comparisons, binary processing was still the primary characteristic of the multialternative choice strategy. Whereas the organization of pair comparisons around similarity relations held only in the presence of many such pairs, binary processing characterized subjects' strategies in all three experiments.

Selection of Pair Comparisons

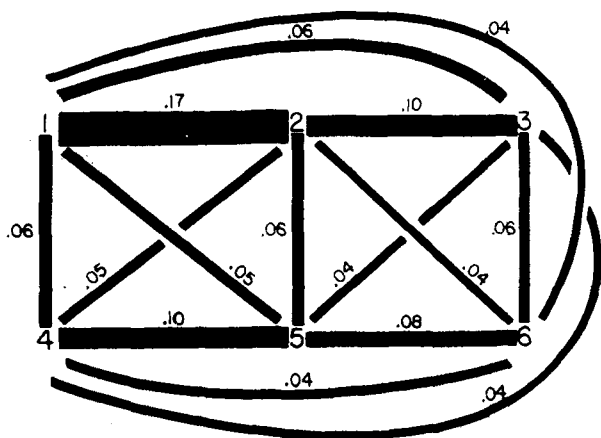Given that the essential work of the multialternative

Figure 1. The relative frequency of binary comparison by display position.

choice task was accomplished in a series of pair comparisons, it was natural to ask how those pairs were chosen. In particular, when similarity relations were not available, what strategies did subjects use to organize the sequence of pair comparisons?

One of the simplest strategies was to select pairs according to spatial proximity. Given such a strategy, the probability of two alternatives' being directly compared should have been a sharply increasing function of their proximity in the stimulus display. Figure 1 displays the relative frequencies of pair comparisons for each of the 15 possible spatial transitions. These data were based on the 748 pair comparisons occurring in independent choice sets of Experiments 2 and 3. The frequencies showed that physical proximity played a large role in determining which pairs of alternatives were directly compared. Although only 7 of 15 pairs (47%) were spatially adjacent, 63% of all pair comparisons were between adjacent pairs.

A second strategy reduced the information processing burden by conserving the contents of STM. One alternative from the previous comparison was retained as part of the next comparison. To test for this strategy, the probability that a contiguous pair of pair comparisons contained a common alternative was estimated. These data are presented in Table 3. For 75% of comparisons, at least one alternative was retained from the previous comparison. Since Subject 8 and Subject 12 used the standard revision strategy, it was expected that a large proportion of alternatives would have been retained from the previous comparison. Indeed, these two subjects retained exactly one alternative from over 80% of the pairs.

These analyses showed that the organization of pair comparisons, in the absence of similarity relations, was determined primarily by information processing

convenience. There was no evidence of selection strategies based on utility.

The Role of Pair Comparisons

To evaluate the role of the pair comparisons, the pairs were contrasted with single fixations. The critical question was whether all of the evaluative processing took place within the pairs, or whether the single fixations also played an evaluative role. Evaluative processing meant processing sensitive to the utilities of the alternatives, as distinct from what might be called bookkeeping functions such as searching, checking, etc. The higher the utility of some alternative, the more evaluative processing it should have received. The real issue was how much of this sensitivity to utility could be attributed to the single fixations relative to the fixations in pair comparisons.

Fixation frequency as a function of utility level was computed for the eight subjects in Experiments 2 and 3, for independent choice sets only. The average relative frequencies are displayed in Figure 2. The results were unequivocal. Only for the fixations in binary processing did frequency increase with utility. Single fixations exhibited no consistent change in fixation frequency as a function of utility. For completeness, Figure 2 includes the results for ternary comparisons, which also showed no sensitivity to utility. The slope of the best fitting straight line through these data served as a rough measure of the strength of the relation between fixation

Table 3
Proportion of Pair Comparisons Retaining Alternative(s) from Previous Comparison

Subject	Number of Alternatives Retained			Number of Pair Comparisons†
	One	Two	Total	
Experiment 1				
1	.65	.05	.70	99
2	.62	.06	.68	72
3	.62	.10	.72	124
4	.74	.04	.78	96
Mean	.66	.06	.72	
Experiment 2				
5	.67	.16	.83	12
6	.71	.07	.78	109
7	.56	.01	.57	73
8	.83	.08	.91	63
Mean	.69	.08	.77	
Experiment 3				
9	.55	.14	.69	42
10	.70	.05	.75	76
11	.61	.08	.69	107
12	.81	.04	.85	146
Mean	.67	.08	.75	

†*Each frequency is the total number of pair comparisons in the 15 independent choices, minus the 15 initial pair comparisons.*
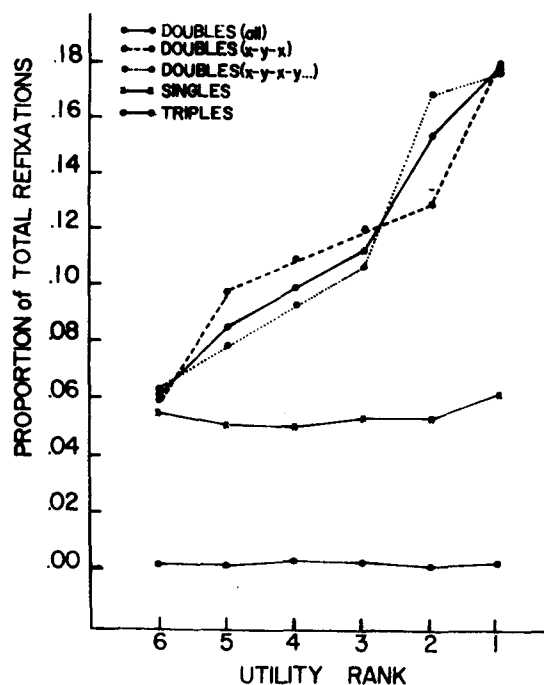
Figure 2. The relative frequency of refixations by utility rank, based on the 748 pair comparisons in the independent choice sets of Experiments 2 and 3.

frequency and utility order. The slopes for the data in Figure 2 were .023 for pairs, .001 for singles, and .0003 for triples.

The results for individual subjects were consistent with the grouped data. For each subject, the chi-square statistic was used to test the null hypothesis that fixations were distributed uniformly across all six utility levels. For the fixations in pair comparisons, all eight subjects showed significant departures from uniformity (p < .005 for every subject). For the single fixations, no subject showed a significant deviation from a uniform distribution (p > .20 for every subject).

There were indications that the singles were used mainly in a search capacity. A surprising characteristic of the single fixations was that they did not occur singly. The average length of a sequence of such singles was 3.1 fixations. A second result was that the pattern of transitions within a sequence of singles showed a very strong effect of physical proximity. Whereas only 14 out of the 30 possible pairwise transitions (47%) were between adjacent alternatives, 73% of all transitions to a "single" fixation were between adjacent alternatives. These two results suggested that the chains of singles often served as search units. A similar phenomenon, in a crypt-arithmetic task, was reported by Winikoff (1967); see also Newell and Simon (1972, p. 313).

Finally, the mean fixation duration in pair comparisons was compared with that of single fixations. These values were 1.20 sec and 1.13 sec, respectively. It had been expected that the single fixations would be

briefer, but only a negligible difference was found. Of the eight subjects, only five exhibited briefer durations on the singles. Thus, the duration data did not reliably distinguish the single fixations from those in pair comparisons.

The Individual Pair Comparison

The previous results indicated the importance of pair comparisons, but did not explain what processing was occurring during these comparisons. A natural question was whether the pair comparisons were similar to those in binary choice, i.e., whether the purpose was to eliminate the inferior alternative. To answer this question, the relation between the processing time for a pair comparison and the utility difference between the two alternatives (ΔU) was examined. If elimination was the goal of a comparison, then larger ΔUs should have led to briefer comparisons.

For the 15 pairs of utilities in the experimental design, Table 4 presents the average processing time for a pair comparison for the independent choice sets of Experiments 2 and 3. The value of the product moment correlation coefficient (r) between the mean processing time and ΔU equalled +.12. This result was not significantly different from zero and, indeed, was not in the predicted direction. Furthermore, none of the eight subjects exhibited a significant value of r. Apparently, the pair comparisons were completely insensitive to ΔU. It was concluded that whatever the purpose of the pair comparisons in Experiments 2 and 3, it had not been elimination of the inferior alternative.

As a check, the same analysis was performed on the data of Experiment 1. Since eliminations had been explicitly required, a clear relation between comparison length and ΔU was expected. The observed value of r was −.50 (p = .028), confirming the prediction.

Table 4
Processing Duration of Pair Comparisons as a Function of Utility

Utility Rank	Utility Rank				
	2	3	4	5	6
1 Mean	5.76	5.32	4.94	4.93	4.41
SD	3.87	3.98	3.35	4.79	3.16
N	120	78	74	58	46
2 Mean		6.26	4.69	6.05	6.00
SD		4.92	3.68	6.48	4.30
N		84	52	60	13
3 Mean			4.57	4.41	4.64
SD			3.34	2.76	2.69
N			52	32	29
4 Mean				4.41	4.20
SD				2.68	2.25
N				40	40
5 Mean					3.39
SD					1.77
N					36

Therefore, when placed under instructions to eliminate alternatives, subjects did use the pair comparisons to execute these eliminations. What, however, was the purpose of the pair comparisons when no explicit eliminations were required?

An alternative hypothesis was that the pair comparisons served as a convenient method for gaining knowledge about two alternatives. This hypothesis suggested that the importance of a pair comparison and, therefore, the processing time devoted to it should have depended upon the average utility ($\bar{U}$) of the two alternatives. To test this prediction, the correlation between $\bar{U}$ and the processing durations was computed. The observed value was $-.65$ ($p < .001$), indicating strong support for this alternative explanation. The eight subjects all exhibited negative correlations, although only four were significant at the liberal .10 level. More importantly, $\bar{U}$ yielded a stronger correlation with processing duration than did ΔU for all but one subject (Subject 8).

Eye fixation data could not specify in fine detail the processing during a pair comparison. Within the constraints imposed by the finding that processing duration was sensitive only to $\bar{U}$ there remained at least two possible interpretations. First, the alternatives were being evaluated relative to each other *and* relative to the choice set. The essence of this hypothesis was that an ongoing evaluation of the used cars was taking place during the pair comparisons. Alternatively, there might have been no evaluation occurring at all, but rather the organizing of the stimulus information into a structure for later evaluation. This latter hypothesis could explain the sensitivity to $\bar{U}$ by having the subject take greater care in organizing the more highly valued alternatives or dimensions. The prompted protocols from Subject 1 and Subject 2 and informal verbal reports from the other subjects lent support to the former hypothesis. The subjects clearly claimed to be evaluating and at least tentatively eliminating alternatives throughout the choice process. The outcome of a typical pair comparison was to discard the inferior alternative and to retain the better one for further evaluation. The subjects also frequently reported a decision to discard both alternatives because they were both relatively unattractive, or to save both members of a pair because they were judged to be two of the best alternatives.

Validation of X-Y-X Patterns

The use of eye fixations in this research relied on the assumption that a pair comparison could be validly identified by an alternation of three or more fixations. Preliminary verbal protocols provided support for this assumption, especially for the case of alternations of Length 4 or longer. For the pattern X-Y-X, however, the verbal protocols confirmed a comparison only 69% of the time. For the result depicted in Figure 2 it was

possible to consider separately the weak (X-Y-X) and strong (X-Y-X-Y . . .) pair comparisons. Figure 2 displays the relative frequencies of fixations in pairs, separately for weak and strong pairs, after each was normalized to the same total frequency as all fixations in pairs. The resulting slopes were .019 for weak pairs and .024 for strong pairs.

From these slopes and the fact that 95% of strong pairs were valid pair comparisons (as shown by the verbal protocols), it was possible to calculate that 75% of the weak pairs indicated valid pair comparisons. This estimate agreed well with the value based on the verbal protocols (69%). The difference could well have been caused by the failure of the verbal data to be complete. This result indicated that the X-Y-X patterns represented something other than pair comparisons about 25% of the time.

DISCUSSION

The Use of Eye Fixations

The remarkable uniformity of processing patterns across subjects deserves some comment. Eye fixation protocols have typically shown that there are many possible ways to perform such tasks as pictorial recognition (Noton & Stark, 1971), chess playing (Tichomirov & Posnyanskaya, 1966), and visual search (Thomas & Lansdowne, 1963). Different subjects exhibit different sequences of fixations for the same task. In contrast, the present experiments have demonstrated a remarkable uniformity of eye fixation patterns across subjects in the multialternative choice task.

The primary reason for the observed consistency across subjects was the successful reduction in that part of the processing variability caused by retrieval of information from LTM. The measurement of utilities provided subjects with extensive experience with the stimuli and their values, prior to the choices. Thus, when a subject was confronted with a choice from six used cars, he did not require widely different times to evaluate different cars. His processing could reflect primarily his computational strategies for combining the evaluations. Imagine how different the processing might have been if a subject were unexpectedly presented with an unfamiliar set of six used cars. There is little doubt that the minimization of LTM retrieval variability played an essential role in the consistency of processing across subjects.

Subject selection may also have contributed to the observed uniformity of processing. Subjects were dropped from an experiment when their utilities indicated that the used cars were not effectively three-dimensional. More importantly, subjects probably did not volunteer for an experiment unless they were moderately familiar with used cars. It is difficult to

evaluate the effect of subject selection on the experimental results. This uncertainty is the price to be paid for using quasinaturalistic stimuli like used cars, which necessarily exclude from the research those subjects unfamiliar with them.

One of the surprises (and disappointments) of the analyses of the eye fixations was the lack of any relation between the fixation durations and cognitive processes. For example, the mean duration of singles was no different than that of pair comparisons. Similarly, in spite of extensive searching for a pattern of durations for the three or more fixations within a pair comparison, no pattern could be found. These failures should be contrasted with the usefulness of the total durations of a pair comparison (Table 4) which were consistently related to the utilities of the two alternatives being compared. These findings support the claim that processing times should be measured only over complete cognitive units.

Implications for Existing Choice Models

The existing models of multialternative choice behavior are stated in terms of choice probabilities and seldom propose a process heuristic. Exceptions are recent papers by Tversky (1972) and by Corbin and Marley (1974), both of which propose sequential elimination processes based on Tversky's elimination by aspects procedure. The results of the present experiments suggest that these elimination processes cannot be accepted as descriptive heuristics. First, the eye fixation data have shown that the choice process is a sequence of pair comparisons. This process bears no resemblance to the sequence of repeated individual evaluations predicted by an elimination by aspects strategy. Second, it was shown that the multialternative choice process does not appear to be based upon eliminations at all. It is always possible, however, to view an alternative that was not chosen as having been implicitly eliminated. When examined in this light, the sequential elimination processes proposed by Tversky and by Corbin and Marley do not conflict with the present findings.

Current models of multialternative choice rely upon very general stimulus representations. A set theoretic representation is used by Tversky (1972), Herniter (1973), and Bass (1974); and Thurstone (1945) proposed a representation in terms of a multivariate Gaussian distribution. In each of these cases, there is a new parameter for each possible subset of alternatives. For choices among N alternatives, this yields on the order of 2^N parameters, a number no one seriously accepts, even for moderate values of N. Indeed, Bass and Thurstone both specify a reduced set of parameters.

The result that all of the evaluative processing took place within pair comparisons suggests that binary interaction parameters *may* be sufficient. Furthermore, not all possible pair comparisons occur, indicating that the majority of binary parameters might not be used at all. One must be careful in using the process tracing results to infer the structure of a different class of models. It is only claimed that the pair comparison structure suggests the sufficiency of binary parameters. There is no evidence, however, of a correspondence between a pair comparison and the need for a corresponding interaction parameter.

Generalizability of the Results

The main finding of the present study is the pair comparison structure of the multialternative choice process. In spite of the uniformity of this finding across subjects and experiments, however, there is reason to doubt that pair comparisons would always be characteristic of multialternative choices. Consider a choice from many alternatives, say, 20, as is typical of a restaurant menu or a supermarket shelf. Most such alternatives would probably be eliminated by a strategy based on individual evaluations, including possibly the elimination by aspects strategy. After the choice set was reduced to a more manageable size, a binary processing strategy might be systematically employed. Svenson (1974) reports some verbal protocol data that support this notion. Similarly, obvious interrelations among three alternatives would almost certainly increase the rate of triple comparisons.

Probably the most distinguishing characteristic of these experiments is the used car stimuli, especially their dimensional structure. Their multidimensionality may have played a large role in determining the subjects' choice strategy. For example, a strategy of comparing alternatives one dimension at a time would generate an alternating pattern of eye fixations. Such fixation sequences might not occur with unidimensional alternatives like crimes (Thurstone, 1927) or famous people (Rumelhart & Greeno, 1971). On the other hand, recent studies of semantic processing have shown that unidimensional stimuli like animals are often processed as a set of features (Smith, Shoben, & Rips, 1974; Rumelhart & Abrahamson, 1973).

The present experiments confirm that processing strategies adapt to specific choice environments. The failure to use similarity relations for three or four subjects in Experiment 3 shows sensitivity to the stimulus environment. Similarly, the use of pairwise eliminations in Experiment 1 demonstrates changes in strategy caused by task constraints. It is expected that other changes, such as alterations of the stimulus display or of the cost/payoff structure, would evoke different processing strategies. This finding suggests that the generalizability of the results to other choice tasks will depend closely on the similarity of a new task environment to the ones used in these experiments. Until an adequate theory of the relationship between

task environments and possible strategies is available, however, the generalizability of the present results will remain an empirical question.

REFERENCES

Bass, F. M. The theory of stochastic preference and brand switching. Journal of Marketing Research, 1974, 11, 120.

Corbin, R., & Marley, A. A. J. Random utility models with equality: An apparent, but not actual, generalization of random utility models. Journal of Mathematical Psychology, 1974, 11, 274-293.

DeGroot, A. D., & Jongman, R. W. Perception and memory in chess: An experimental analysis of the master's professional eye. RITP Memorandum, No. 024, University of Amsterdam, 1973.

Drake, D. M. Perceptual correlates of impulsive and reflective behavior. Developmental Psychology, 1970, 2, 202-214.

Goode, F.M., & Russo, J. E. A computerized system for collecting eye movement data. Proceedings, 70th Annual Convention, American Psychological Association, 1970, 845-846.

Herniter, J. An entropy model of brand purchase behavior. Journal of Marketing Research, 1973, 10, 361-375.

Krantz, D. H. The scaling of small and large color differences. Unpublished doctoral dissertation, University of Pennsylvania, 1964.

Newell, A., & Simon, H. A. Human problem solving. Englewood Cliffs, N. J.: Prentice-Hall, 1972. Pp. 310-382, 378-401, 755-780.

Nodine, C. F., & Steuerle, N. L. Development of perceptual and cognitive strategies for differential graphemes. Journal of Experimental Psychology, 1973 97(2), 158-166.

Noton, D., & Stark, L. Eye movements and visual perception. Scientific American, 1971, 224 (b), 34-43.

Rumelhart, D. E., & Abrahamson, A. A. A model for analogical reasoning. Cognitive Psychology, 1973, 5, 1-28.

Rumelhart, D. L., & Greeno, J. G. Similarity between stimuli: An experimental test of the Luce and Restle choice models. Journal of Mathematical Psychology, 1971, 8, 370-381.

Russo, J. E., & Mathews, P. A photoelectric eye position sensor. Unpublished manuscript, University of California, San Diego, 1975.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model for semantic decisions. Psychological Review, 1974, 81, 214-241.

Svenson, O. Coded think aloud protocols obtained when making a choice to purchase one of seven hypothetically offered houses: Some examples. Progress report on project of cognitive processes and decision making, University of Stockholm, 1974.

Thomas, E. L., & Lansdowne, E. L. Visual search patterns of radiologists in training. Radiology, 1963, 81, 288-292.

Thurstone, L. L. Method of paired comparisons for social values. Journal of Abnormal Social Psychology, 1927, 21, 384-400.

Thurstone, L. L. The prediction of choice. Psychometrika, 1945, 10, 236-253.

Tichomirov, O. K., & Posnyanskaya, E. D. An investigation of visual search as a means of analyzing heuristics. Soviet Psychology, Winter, 1966-67, 5, 2-15. Transl. from Voprosy Psikhologii, 1966, 2(4), 39-53.

Tversky, A. Elimination by aspects: A theory of choice. Psychological Review, 1972, 79, 281-299.

Tversky, A., & Russo, J. E. Substitutability and similarity in binary choices. Journal of Mathematical Psychology, 1969, 6, 1-12.

Vurpillot, E. The development of scanning strategies and their relation to visual differentiation. Journal of Experimental Comparative Psychology, 1968, 6, 632-650.

Winikoff, A. W. Eye movements as an aid to protocol analysis of problem solving behavior. Unpublished doctoral dissertation, Carnegie-Mellon University, 1967.