"A Comparison of Two Process Tracing Methods for Choice Tasks"

95-10-08

Gerald Lohse and Eric Johnson

The Proceedings of the 29th Hawaii International Conference on Systems Sciences, January 1996

A COMPARISON OF TWO PROCESS TRACING METHODS FOR CHOICE TASKS

Gerald L. Lohse
Eric J. Johnson

95-09-03

Department of Operations and Information Management
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6366

Forthcoming: *Proceedings* 29th Hawaii International Conference on System Sciences, January 1996

A comparison of two process tracing methods for choice tasks

Gerald L. Lohse
Eric J. Johnson
The Wharton School of the University of Pennsylvania
Philadelphia, Pennsylvania USA

Abstract

The proliferation of computerized process tracing tools for monitoring information acquisition behavior demands an inquiry into their validity. This research compares information acquisition behavior for choice tasks using Mouselab, a computerized process tracing tool, and Eyegaze, an eye tracking system. In a paired difference experiment using apartment selection tasks and gambles, we found significant differences in information search patterns and subsequent choice behavior contingent upon the process tracing method. Overall, computerized process tracing tools increase the amount of effort needed to acquire information compared with eye tracking equipment. Our results show that subjects adopt information acquisition strategies to cope with the increased effort. These strategies tend to be more rigorous and systematic than those observed with eye-tracking equipment. Thus, computerized process tracing tools fundamentally alter the information processing behavior they are believed to track unobtrusively by limiting the ability of the decision maker to adapt their information processing behavior dynamically to the demands of the data. Additional research is needed to explore the magnitude and consequences of these differences.

Introduction

Information acquisition includes an analysis of the content of the information sought, how long the subject examines this information, the sequence of acquisition, and the amount of information acquired [12]. These data are important for research in behavioral decision making and decision support systems for at least three reasons. First, patterns of information processing suggest certain strategies for evaluating information [20]. Understanding these prototypical patterns for evaluating information helps identify behavior that could constrain or alter decision processes. Second, the information acquisition patterns directly influence cognition and memory. Subtle changes in presentation format can change the frequency of preference reversals [16], change decision making strategy [6, 15, 31] and alter decision performance [2]. Third, because the way information is displayed can change decisions, understanding these influences is important for electronic commerce [34]. For example, Russo [24] induced supermarket shoppers to purchase products with lower unit prices by providing unit price information on a single list. The advent of electronic shopping also allows marketers to capture data analogous supermarket scanner data but more complete, because it contains information about consumer's consideration of products not purchased. Consumer examination of products in on-line services or the World Wide Web can be an important source of market intelligence. More generally, knowledge about the effect of information acquisition patterns could help design future information products, retail kiosks, and multimedia information services producing new retail environments such as personalized electronic catalogs that reflect consumer preferences.

The relationship between choice processes and the decision environment raises an interesting and fundamental question about different methods of recording acquisition. Because they require different levels of acquisition effort on the part of the decision-maker, do various processes tracing methods change the decision process? In other words, are they reactive procedures?

Existing information acquisition research has used three major types of process tracing methods each with its own typical method for acquiring information: information display boards in which information is retrieved by reaching for an index card, computerized process tracing (CPT) systems in which acquisition is made using a mouse, light pen, keystrokes and the like, and eye tracking [26, 1] in which the natural discrete fixations of the eye are recorded. Information boards [7] present information in an envelope on a poster board containing an index card with one piece of information (e.g., rent for apartment C, see 20). The overt action of selecting and reading one card requires 3-4 seconds. CPT environments replace manual acquisition with a computer based pointer that requires less time (1-2 seconds, [5]) per acquisition typically. For comparison, an eye fixation typically requires 0.2 - 0.4 seconds to acquire the same piece of information [10].

Surprisingly, very little research has examined the influence of these different process tracing procedures upon the underlying process itself. Russo [26] did provide a review of the relationship between information display boards and eye movement which suggested significant changes in information acquisition behavior. However, the advent of personal computers spurred rapid development of CPT systems for capturing information acquisition behavior. Numerous CPT systems exist (e.g., *Mouselab*, [22]; *IS Lab*, [11]; *Search Monitor*, [8]; and other unnamed tools [31, 35]). Each system has many users and CPT systems have become the most common process tracing tool for studying information acquisition behavior. Despite a history of information acquisition research using CPT tools, researchers have not compared eye tracking and CPT tools.

We approach this comparison from a cost-benefit perspective. Consistent with prior research [5, 19], costs are measured as a function of the number of elementary information processes (EIPs). In the spirit of computational models of cognition [10], we assign timing parameters to each EIP and sum the time required for each EIP to predict total task completion time. Thus, costs include information acquisition effort and mental processing while benefits reflect the utility or accuracy of the decision.

The goal of this paper is two fold. The first is the development of an effort accuracy theoretical framework [22] for understanding any possible changes in decision processes that might occur for different process tracing methods. The second goal is to examine differences in two particular methods of observing information acquisition, eye movement recording and one particular mouse-based CPT. We offer this both because it illustrates how one might understand, in general, how different ways of presenting and monitoring information might change decision processes generally, and because it is interesting to know what differences exist in this specific case. Our general prediction is that differences in information processing strategies will be explained by differences in time required to acquire information using each method.

This study also has indirect implications for the development of systems for electronic commerce. The advent of cybershopping on Compuserve, American Online, and Prodigy as well as "Electronic Malls" available on the Internet, heightens the need for additional research on consumer information systems [4]. One major issue is how best to satisfy consumer needs through the design of aids to present information. By and large consumer products and services are advertised and displayed one product per screen. Additional information about product attributes is often made available using hypertext links. While these links provide additional product information, they also prevent consumer comparison of products by attribute because it is difficult to display multiple products on the same screen [3]. Thus, current electronic shopping interfaces restrict product comparability by attribute.

Of course we are not testing electronic shopping systems *per se*, but CPT systems are analogous to current electronic shopping interfaces that display information for a product on a single page. Any differences between choices made using CPT systems and those made using eye tracking equipment, would suggest that the interface design influences consumer choice. Such differences would have numerous public policy implications for electronic commerce systems regarding market efficiency and consumer choice.

The remainder of this paper provides further information about the tools used to implement the comparison, and then develops the general model and the set of hypotheses that follow. We then present the results of an experiment that compare the two process tracing methods and discuss the results. Finally, we discuss implications of information acquisition tools as well as the application of the accuracy-effort model to better understand the impact of information acquisition methods and computer based decision aids.

Process tracing tools

The Eyegaze System (LC Technologies Fairfax, VA) uses the pupil-center/corneal reflection method to determine eye gaze [36]. A video camera, sensitive in the infra-red range and positioned below the computer monitor, continually observes the subject's eye. A small, low power, infrared, light emitting diode (LED) located at the center of the video camera lens illuminates the user's eye. By means of video image processing, an algorithm determines the center of the pupil and the brightest reflection of the cornea (as illuminated by the LED). Specialized image-processing software generates x, y coordinates for the gaze point on the monitor screen. Other measures include: fixation duration, pupil diameter, and eye blinks. The observer's eye is about 20 inches from the screen of the computer monitor. The Eyegaze System collects data at 60 Hz or about every 16 milliseconds. The eye tracking equipment does not require attachments to the head. The calibration procedure for each new subject takes less than one minute.

Mouselab monitors the information acquisition stages of decision behavior for many tasks [22]. Subjects use a mouse to acquire information from a computer display. Boxes cover information on the display. The user selects a box using a mouse to reveal the information behind the box. When the user moves the mouse off the information, the box reappears and prevents the user from viewing the information behind the box (Figure 1). Mouselab time stamps when a user selects and releases a set of boxes that conceal information used in the choice task. Data include information about the time, sequence, and frequency of a user's information acquisition behavior. These detailed

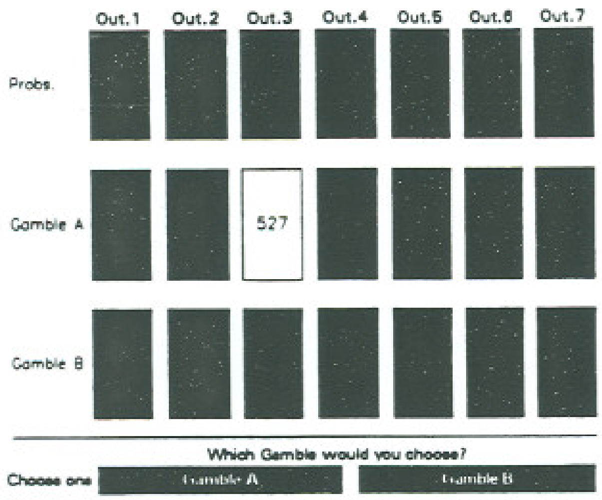

Figure 1: Gamble with two alternatives and seven attributes using Mouselab.

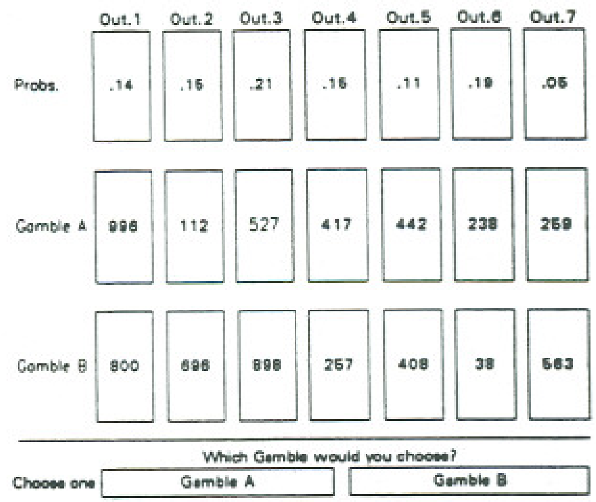

Figure 2: Gamble with two alternatives and seven attributes using Eyegaze.

data approach the level obtained with eye movement recording. However, Mouselab only captures about 60% of the process tracing data. Data about scanning to row and column labels on the display is not captured with Mouselab. Thus, eye tracking systems provide a more complete accounting of what information subjects attended to during a task.

Hypothesis / Cognitive Effort Framework

Bettman, Johnson, and Payne [5] proposed a set of EIPs to describe and measure cognitive effort for different decision strategies. This paradigm predicts cognitive effort as a function of the number of EIPs required to execute a particular decision strategy. Total task completion time is the sum of the subcomponent times for each EIP. We adopt this framework for calculating EIP counts and task completion times to predict differences between Mouselab and Eyegaze for collecting information acquisition data. Bettman, Johnson, and Payne show that 75% of the variance in decision times can be accounted for by a count of these mental operations, and that individual differences in the cost of these operations can help predict individual differences in the selection of decision strategies.

We illustrate the EIP counts using the gamble task from Figure 2 with two alternatives and seven attributes. Expected value calculations for gambles using a weighted adding rule have two steps. First, subjects evaluate each alternative by multiplying each probability times the attribute payoff and adding those products for all attributes. Second, subjects retain the value of the best alternative and its label in working memory while computing the value of the remaining alternatives. For each alternative, a subject compares the value of the current alternative to the best possible alternative. The EIPs are **Read** a piece of information into working memory, **Multiply** a probability times its payoff, and **Add** or **Compare** two values in working memory.

Humans process some information from a table using learned spatial patterns. Reading in English, for example, begins in the upper left-hand corner and proceeds from left to right in a top to bottom fashion. Of course, there are cross cultural differences among these learned spatial patterns (e.g., Hebraic and some Asian reading patterns). Assuming an English reading order of left to right and top to bottom, subjects read the first probability (.14) and the gamble A payoff for outcome number one (996). Next, subjects mulitply the two values to obtain the product (139.44) and retain this value in working memory. Then, subjects read the second probability (.15) and the gamble A

Table 1 Weighted additive rule applied to gambles

Gamble Alternatives	2	2	7	7	Eyegaze	Mouselab
Gamble Attributes	2	7	2	7	time in	time in
Gamble Probabilities	2	7	2	7	seconds	seconds
Reads (probabilities + attributes) * alternatives	8	28	28	98	0.23	1.19
Products (attributes x alternatives)	4	14	14	49	0.84	0.84
Additions (alternatives x (attributes-1))	2	12	7	42	0.84	0.84
Comparisons (alternatives -1)	1	1	6	6	0.84	0.84

payoff for outcome number two (112). Next, subjects multiply the two values to obtain the product (16.8) and retain this value in working memory. Subjects add the two products retained in working memory to obtain the sum (156.24). The process continues until determining the value of gamble A (436.25). After processing the first alternative, there would be fourteen Reads, seven Products, six Additions, and no Comparisons. A similar process determines the value of gamble B (523.78). After processing the second alternative, there would a Comparison of the values for gamble A and gamble B. Hence, there is a total of 28 Reads, 14 Products, 12 Additions, and one Comparison. Table 1 shows the EIP counts for other gamble tasks.

The EIP counts are irrespective of the processing tracing method. Predicted differences between the two processes tracing tools depend on the amount of time required to execute each EIP. Mental computations for multiplication, addition and comparison are identical regardless of the process tracing method. Bettman, Johnson, and Payne [5] cite values of .84 seconds for additions, 2.23 for products, and .09 for comparisons. They note that their estimate for Product EIP is larger than that commonly reported in the literature and that the Comparison EIP time is very short. We use .84 seconds as an average value for all mental operations. The only difference between the two process tracing methods is the amount of time required to read a piece of information into working memory. For eye-tracking systems, Russo [25] lists 0.230 seconds as a typical saccadic eye-movement time to read and process one piece of information (travel + fixation time). The comparable time for Mouselab is 1.19 seconds for an average mouse move to read one piece of information [5]. This value agrees with the published value of 1.1 seconds in Card, Moran, & Newell [10]. Thus, our eyes are faster than the mouse for acquiring information from a display.

Total Time The total time required to make a choice measures cognitive effort. Table 2 shows the predicted total time to complete the choice tasks using Mouselab and the Eyegaze System. Times range from 7.7 seconds for the apartment selection task with two alternatives and 2 attributes to 198.1 seconds to complete the gamble task with seven alternatives and seven attributes. The ratio of Mouselab time to Eyegaze time predicts that Eyegaze will be nearly twice as fast as Mouselab. **Hypothesis 1: The time to complete a choice with Eyegaze will be twice as fast as Mouselab.**

Mean Fixation Duration and Total Number of Fixations The average time to point to an object using a mouse is 1.19 seconds [5] and the average saccadic eye-movement time is .23 seconds. Subjects could use up to five times as many Eyegaze fixations per Mouselab fixation. However, we expect subjects using Eyegaze will spend less total time on each choice task. Thus, we expect to find that average fixation time will be greater for Mouselab but that total number of fixations will be greater for the Eyegaze System. **Hypothesis 2: The mean time per fixation for Eyegaze is less than for Mouselab. Hypothesis 3: The total number of fixations will be greater for Eyegaze than for Mouselab.**

Accuracy Correct choices for the gamble tasks depend upon accurate mental addition and multiplication. Mental arithmetic involves well-learned procedures, problem-solving skills, and reliance on working memory. Most adults use associative memory to access the product of 3 x 4 directly. For more complex calculations (116 x 996), Hitch [13] demonstrated that mental arithmetic errors occur because subjects fail to retain an accurate record of carries and intermediate calculations in working memory.

Two factors suggest that the working memory load is higher for Mouselab than for Eyegaze. First, concurrent arm movement disrupts the retention of spatial information while concurrent visual input does not [17, 23, 30]. This suggests that the arm movements used to control the mouse for acquiring information from the display increase the information processing burden on working memory. Thus, the mental effort involved in using a mouse is slightly greater than that of using Eyegaze. Second, longer information acquisition times using Mouselab increase the total time information must be retained in working memory. With sufficient levels of activation, working memory retains information for about seven seconds. This duration limitation makes it much more difficult to maintain a large amount of information in working memory.

The expected value calculations for gambles place a very high load on working memory. We expect that Mouselab will cause more slips, forgetting and errors in the arithmetic calculations. As a result, performance will be less accurate with Mouselab than with Eyegaze. **Hypothesis 4: More subjects will make the optimal choice for gambles using Eyegaze than using Mouselab.**

Percent Information Searched The mean proportion of information searched examines the amount of information used in making a choice. Percent information searched is the number of cells examined divided by the total number of cells. A high proportion of information search indicates the use of compensatory processes whereas a low proportion of information search indicates the use of noncompensatory processes. The time and effort required to acquire a piece of information using the Eyegaze System is less than that for Mouselab. Thus, we expect the mean proportion of information search will be greater for the Eyegaze System. **Hypothesis 5: The percent information searched will be higher with Eyegaze than with Mouselab.**

Search Direction The search direction index [21, 28,

	Appear.	Distance	Rent	Safety	Kitchen	Landlord	Laundry
Apt. A	V.Clean	10 min.	$150	Average	Partial	Dffcult	Coin
Apt. B	Clean	24 min.	$200	Unsafe	None	Helpful	None
Apt. C	V.Dirty	24 min.	$550	Average	Partial	Dffcult	Free
Apt. D	Dirty	12 min.	$300	V. Safe	Full	Dffcult	None
Apt. E	Dirty	24 min.	$450	Safe	None	Helpful	Free
Apt. F	V.Clean	15 min.	$600	Safe	Partial	Dffcult	None
Apt. G	Dirty	30 min.	$300	Unsafe	None	Average	Coin

Which Gamble would you choose?

Choose	Apt A	Apt B	Apt C	Apt D	Apt E	Apt F	Apt G

Figure 3 Apartment selection task with seven alternatives and seven attributes

11], is a function of the number of transitions within the same attribute (Intra) and the number of transitions within the same alternative (Inter). Intra is the number of instances in which the *n*th+1 item searched is of the same attribute as the *n*th. Inter is the number of instances in which the *n*th+1 item searched is of the same alternative as the *n*th. A score of 1.0 represents a strict interdimensional search (e.g., weighted adding rule). A score of -1.0 represents a strict intradimensional search (e.g., elimination by aspects). Interdimensional search indicates each move to a new piece of information within the same alternative. Intradimensional search indicates shifts between alternatives. Choice tasks exhibit intradimensional search processes. Thus, Eyegaze should lead to more intradimensional search processes than Mouselab. **Hypothesis 6: Eyegaze should lead to more intradimensional search processes than Mouselab.**

Search Direction = (Inter - Intra) / (Inter + Intra) (1)

Reacquisition Rate The time and effort required to acquire a piece of information using the Eyegaze System is less than that for Mouselab. Therefore, we also expect that there will be fewer reacquisitions of information using Mouselab as compared with the Eyegaze System. We compute reacquisition rate as the number of cells viewed at least twice divided by the total number of cells viewed. Thus, the denominator only reflects actual cells viewed not the total number of cells in the problem. **Hypothesis 7: The reacquisition rate will be higher for Eyegaze than Mouselab.**

Variability in Information Search In a typical recognize-act cognitive cycle, recognition is a fundamentally parallel task and the action phase is serial. Mouselab imposes strictly serial information processing. In contrast to a normal information processing environment, Mouselab eliminates the possibility of acquiring information from peripheral vision. Users can not scan ahead to direct attention. Mouselab supports goal-directed, top-down processing. Whereas Mouselab imposes serial processing, Eyegaze involves parallel recognition and a serial action phase. Eyegaze facilitates data-driven, bottom-up processing. Thus, variability in percent information searched will be greater for Eyegaze than for Mouselab. **Hypothesis 8: Eyegaze will lead to higher variation in information acquisition processes than Mouselab.**

We adopt measures for variability in percent information searched from Cook and Swain [11]. Variability by alternatives measures the standard deviation of the percent information searched per alternative across the set of available alternatives. Variability by attributes measures the standard deviation of the percent search per attribute across the set of available attributes.

$$\left[\left(\Sigma\ x_i^2 - \left((\Sigma\ x_i)^2\right)/n\right)/(n-1)\right]^{1/2} \qquad (2)$$

where n = number of total attributes and
x_i = percent of attribute i searched

Method

Stimuli Each gamble alternative was a row in a matrix (Figure 2). Each cell in a row was a payoff. The first row of the gamble choice task contained a vector of probabilities associated with each payoff dimension. The probability vector summed to one. The sum of the cross product of the payoff vector and the probability vector yielded the expected value of each gamble (Equation 3). The order of the rows and columns was the same for all subjects. For each choice task, one gamble maximized expected value.

$$EV_j = \sum_{i=1}^{n} (payoff)_{ij}\ (probability)_i \qquad (3)$$

The apartment choice task presented each apartment as a row in a matrix (Figure 3). There were two or seven attributes for each apartment. The seven attributes were Appearance (5 levels from very dirty to very clean), Distance (in minutes from campus), Rent, Safety (5 levels from very unsafe to very safe), Kitchen (3 levels none, partial, full), Landlord (3 levels difficult, average, helpful), and Laundry (3 levels none, coin, free). For choice problems with only two attributes, the attributes were Distance and Appearance. Information about each attribute was a column in a matrix. The order of the rows and columns was the same for all subjects. No alternative dominated another on all attributes.

Subjects Thirty-six undergraduate students from the University of Pennsylvania participated in the study. All subjects had prior computer experience using a mouse. All subjects had at least some familiarity with expected value from their course work (e.g., one lecture in an introductory Economics class). At the end of the second session, subjects received a payment of $20.00 for their participation. The experimenter told subjects that there were no right and wrong answers for the apartment selection task. In addition, subjects with the best performance for the gambles received an additional award of $100.

Procedure Subjects attended two sessions held one week apart. At one session, subjects used Mouselab; at the other they used the Eyegaze System. A practice session helped subjects learn to make eye movements with minimal head movement. Subjects using Mouselab had a similar practice session using the mouse.

After the practice session, subjects evaluated four apartment selection problems and four gambles representing the fully crossed combinations of alternatives and attributes. The choice tasks alternated (e.g., apartment selection, gamble, apartment selection, gamble, etc.). All factors were counter-balanced across subjects. The order of choice problems for each subject varied randomly.

Design Thirty-six subjects participated in the study. There were four within-subjects factors: process tracing tool (Mouselab versus Eyegaze), two alternatives (2 or 7 items), two attributes (2 or 7 items), and two tasks (apartment selection versus gamble). The resulting design had a 2x2x2x2 factorial arrangement of treatments. The within-subjects design allows a paired difference comparison between Mouselab and Eyegaze for all the dependent measures.

Results and Discussion

Every subject performed each task using both methods. Each observation subtracts Eyegaze performance from Mouselab. Differences significantly different from zero indicate performance was contingent on the process tracing method. The aggregate data includes total number of fixations, and total time in seconds to complete the task as well as whether the choices were identical using both methods. There are 288 observations for each dependent variable (36 subjects x 2 methods x 4 tasks per subject). A program computed the process tracing measures from 26,447 Eyegaze fixations and 16,992 Mouselab fixations over all 36 subjects. Process tracing measures include percent of the total information searched, search direction, reacquisition rate, variability by alternative, variability by attribute, and mean fixation duration. All statistical analyses use a within-subjects paired difference comparison. Table 3 summarizes the all hypotheses and their results. Detailed analyses of the aggregate and process tracing data follow.

Aggregate Data Results

A paired difference comparison of the aggregate data found significant differences for all three measures (Table 4). Almost one-third (93/288) of the choices changed as a function of the process tracing method [T(287)=11.70, p<.0001]. On average, subjects used 55.2 more fixations per task using Eyegaze than with Mouselab [T(287)= -11.58, p<.0001]. Subjects also spent 29.66 more seconds per task using Mouselab than with Eyegaze [T(287)= 8.30, p<.0001].

MANOVA examined these performance effects simultaneously while controlling for correlation among the dependent measures. Table 5 shows overall significant main effects for Subject, Task, Alternatives, Attributes and overall significant interaction effects for Task x Attribute as well as Alternatives x Attributes. Three univariate ANOVA models examine these effects in more detail for each dependent measure.

Accuracy If subjects are consistent in applying their preference criteria to the choice tasks, their choice should not change as a function of the method used to collect process tracing data. The paired difference data codes identical choices as zero and different choices as one. The mean is a ratio expressing the percentage of choices that are different. The univariate ANOVA explains 30% of the variance in choices. There were more differences in choices for gambles (.38) than for the apartment selection task (.26) [F(1,245)=5.51, p<.0191]. The difference also was greater for tasks with seven attributes (.44) than for tasks with two attributes (.20) [F(1,245)=23.60, p<.0001]. Given the subjective nature of the apartment selection task, it is difficult to determine which process tracing tool facilitated more accurate decisions. However, in half the gambles, Eyegaze had the optimum gamble more frequently than Mouselab (2x2 88.9% vs. 75% and 7x2 97.2% vs. 77.8%). There were no significant differences for the other gambles.

Total time Overall, subjects using Mouselab required 67% more time to complete the tasks. The magnitude of the effect is contingent upon the task (apartment selection or gamble), the number of alternatives and the number of attributes. The univariate ANOVA explains 45% of the variance in total time. Time differences were greater for gambles (45.4 seconds) than for apartment selection (13.9 seconds) [F(1,245)=30.50, p<.0001]. The difference was greater for tasks with seven alternatives (43.1 seconds) than for tasks with two alternatives (16.2 seconds) [F(1,245)=22.10, p<.0001]. The difference was also greater for tasks with seven attributes (44.6 seconds) than for tasks with two attributes (14.7 seconds) [F(1,245)=22.10, p<.0001]. Interpretation of a significant two-way interaction for Alternative x Attribute [F(1,245)=4.11, p<.0438] shows that task completion time increases as task complexity increases. Tasks with two alternatives and two attributes differed the least in the amount of time required to complete the task (13.5 seconds

Table 2 Comparison of predicted times and actual times for gambles

Gamble (alternatives x attributes)	2x2	2x7	7x2	7x7
Predicted total time for Eyegaze	7.7	29.1	29.1	104.0
Predicted total time for Mouselab	15.4	56.0	56.0	198.1
Predicted Ratio Mouselab:Eyegaze	1.99	1.92	1.92	1.90
Actual average time for Eyegaze	32.9	57.4	36.6	90.7
Actual average time for Mouselab	46.5	103.3	71.5	178.2
Actual Ratio Mouselab:Eyegaze	1.41	1.80	1.95	1.96

Table 3 Summary of hypotheses and results.

	Hypothesis	Results	Statistical Test
1	Eyegaze is twice as fast as Mouselab.	7 alternatives x 7 attributes tasks were 96% longer with Mouselab	F(1,245)=4.11, p<.0438
2	Eyegaze time per fixation is less than Mouselab.	An average Eyegaze fixation is 38% of the duration of a Mouselab fixation	F(1,43248)=3057.4,P<.0001
3	Total number of fixations for Eyegaze is greater than Mouselab.	Eyegaze averaged twice as many fixations as Mouselab to complete each task.	T(287)= -11.58, p<.0001
4	Eyegaze is more accurate than Mouselab.	Eyegaze had the optimum gamble more frequently than Mouselab	F(3,32)=3.13, p<.0391
5	Percent information searched is greater with Eyegaze than Mouselab.	Mouselab examined 3.3% more information than Eyegaze users	F(147,139)=2.06, p<.0001
6	Eyegaze should lead to more intradimensional search processes than Mouselab.	Subjects had interdimensional search (rowwise) using Mouselab but had intradimensional search (columnwise) using Eyegaze	F(147,139)=2.73, p<.0001
7	Reacquisition rate is higher for Eyegaze than Mouselab.	Subjects reacquired more information using Eyegaze than with Mouselab	F(42,244)=4.32, p<.0001
8	Eyegaze should lead to higher variation in search processes than Mouselab.	Eyegaze greater variation in the proportion of information searched than Mouselab	F(147,138)=1.28, p<.0693

Table 4 Paired difference comparisons for aggregate data (Mouselab - Eyegaze)

Source	N	Mean	Std Error	T value	Significance
choices different	288	.323	.0276	11.70	.0001
number fixations	288	-55.2	4.7700	-11.58	.0001
time in msec	288	29,661	3,574.2	8.30	.0001

Table 5 MANOVA for aggregate paired difference data (choice differences, total time, number of fixations)

Source	DF	Wilk's Lamba	F value	Significance
Subject	105, 728.6	.3376	3.0334	.0001
Task	3, 243	.7968	20.6504	.0001
Alternatives	3, 243	.8729	11.7962	.0001
Task x Alternatives	3, 243	.9733	2.2257	.0858
Attributes	3, 243	.6686	40.1442	.0001
Task x Attributes	3, 243	.9543	3.8780	.0098
Alternatives x Attributes	3, 243	.9516	4.1237	.0071
Task x Alternatives x Attributes	3, 243	.9845	1.2714	.2848

or 41% longer). Tasks with seven alternatives and seven attributes differed the most in the amount of time required to complete the task (87.5 seconds or 96% longer). Thus, the more complex the task, the greater the difference in total time required to complete the task using Eyegaze and Mouselab.

Number of fixations Subjects using the Eyegaze System averaged twice as many fixations to complete each task. The univariate ANOVA explains 41% of the variance in total number of fixations. There was not a significant main effect for task or number of alternatives nor were any interaction terms significant. The difference was greater for tasks with seven attributes (73.3 fixations) than for tasks with two attributes (37.2 fixations) [$F(1,245)=20.88$, $p<.0001$].

Aggregate Data Discussion

The within-subjects, paired-difference comparison provides a very powerful test for accuracy differences. Choice behavior is contingent upon the process tracing method. For gambles, Mouselab was never more accurate than Eyegaze. In half of the gambles, Eyegaze was significantly more accurate than Mouselab. Because of a greater burden on a capacity constrained working memory, we postulated that Mouselab would cause more slips, forgetting and errors in the arithmetic calculations. As a result, performance would be less accurate with Mouselab than with Eyegaze. If this were true, it is more likely that there would be greater accuracy differences for gambles with seven attributes. We found the opposite. The gamble tasks with two attributes and either two or seven alternatives had significant accuracy differences. The accuracy analysis does not rule out the possibility that the choice differences reflect a change in preferences during the one week interval between trials. The experiment may be capturing a shift in preferences rather than true differences attributable to the process tracing method. While this might be true for apartments, gambles only have one optimum value.

Table 2 shows predicted and actual total times for gambles using Mouselab and Eyegaze. The ratio of predicted times compares favorably to the actual average time ratio. Except for the 2x2 gamble, Mouselab required nearly twice as much time to complete each task. The large and significant task completion time differences establish that Mouselab was more effortful than Eyegaze. While internal cognitive processes are unobservable, we attribute the total time differences to greater working memory loads imposed by Mouselab. These results also provide strong empirical support for the predictive validity of the componential cognitive effort technique described by Bettman, Johnson, and Payne [5].

Subjects used twice as many fixations with Eyegaze than with Mouselab. The rate of fixations per minute is actually much higher since subjects using Mouselab required nearly twice as much time to complete the task. The number of fixations per minute was three times greater for Eyegaze than for Mouselab. The result is similar to that found by Russo [26] for information boards. The number of fixations per minute was at least ten times greater for eye movements than for information boards. The differences highlight the importance of the granularity of the time unit used to collect process tracing data. The coarser the unit of time per fixation, the more difficult it will be to interpret information acquisition patterns.

Process Tracing Data Results

MANOVA examined these performance effects simultaneously while controlling for correlation among the dependent measures. There were significant main effects for Subject, Task, Alternatives, Attributes and overall significant interaction effects for Alternatives x Attributes, Subject x Task, and Subject x Alternatives. Six univariate ANOVA models examine these effects in more detail for each dependent measure.

Percent Information Searched Overall, subjects using Mouselab examined 3.3% more information than Eyegaze users. The magnitude of the effect is contingent upon the number of alternatives. The univariate ANOVA explains 69% of the variance in percent information searched. The difference was greater for tasks with seven alternatives (5.9%) than for tasks with two alternatives (0.7%) [$F(1,138)=7.74$, $p<.0062$]. Thus, the more complex the task, the greater the difference in percent information searched between Eyegaze and Mouselab.

Search Direction Overall, subjects acquired information from Mouselab with a slightly interdimensional search (rowwise) whereas subjects acquired information from Eyegaze with a slightly intradimensional search (columnwise). The magnitude of the effect is contingent upon the number of alternatives and number of attributes. The univariate ANOVA explains 74% of the variance in search direction. The paired difference was greater for tasks with seven alternatives (-.265) than for tasks with two alternatives (.021) [$F(1,138)=36.64$, $p<.0001$]. The paired difference was also greater for tasks with two attributes (-.317) than for tasks with seven attributes (.076) [$F(1,138)=74.25$, $p<.0001$]. The difference is striking considering the same subject used a completely different search strategy contingent upon the process tracing tool.

Reacquisition Rate Overall, the reacquisition rate was 47% for Mouselab and 69% for Eyegaze. Subjects re-examined more information using Eyegaze than with Mouselab. The magnitude of the effect is contingent upon the number of alternatives, the number of attributes, and the task. The univariate ANOVA explains 65% of the variance in reacquisition rate. The paired difference was greater for the apartment selection task (-.320) than for gambles (-.125) [F(1,138)=30.05, p<.0001]. Tasks with two alternatives (-.270) had a higher difference in the reacquisition rate than tasks with seven alternatives (-.175) [F(1,138)=6.62, p<.0111]. The paired difference was also greater for tasks with two attributes (-.262) than for tasks with seven attributes (-.183) [F(1,138)=5.25, p<.0235]. Thus, the less complex the task, the greater the difference in reacquisition rate between Eyegaze and Mouselab.

Variability by Alternative Overall, Eyegaze had a marginally greater variation in the proportion of information searched by alternative than Mouselab [F(147,138)=1.28, p<.0693]. The magnitude of the effect is contingent upon an interaction between the number of alternatives and the number of attributes. The univariate ANOVA explains 58% of the variance in variability by alternative. The paired difference was greatest for tasks with seven alternatives and two attributes (-5.4%) [F(1,138)=4.48, p<.0361]. Tasks with two alternatives and two attributes or seven alternatives and seven attributes had the least difference in variation in the proportion of information searched by alternative. Thus, tasks with more attributes than alternatives or visa versa exhibited greater differences in the variability by alternative.

Variability by Attribute Overall, Eyegaze (11.6%) had a significantly greater variation in the proportion of information searched by attribute than Mouselab (7.7%). The magnitude of the effect is contingent upon an interaction between the number of alternatives and the number of attributes. The univariate ANOVA explains 60% of the variance in variability by attribute. The paired difference was greatest for tasks with seven alternatives and two attributes (-13.5%) [F(1,138)=6.02, p<.0154]. Tasks with two alternatives and two attributes or seven alternatives and seven attributes had the least difference in variation in the proportion of information searched by attributes. Thus, tasks with more attributes than alternatives or visa versa exhibited greater differences in the variability by attribute.

Fixation Duration Overall, Mouselab averaged 1.267 seconds per fixation whereas the average for Eyegaze was .377 seconds. The magnitude of the effect is contingent upon the number of alternatives, the number of attributes, and the task. The univariate ANOVA explains 76% of the variance in mean fixation time. The paired difference was greater for gambles (.74) than for the apartment selection task (.54) [F(1,138)=23.68, p<.0001]. Tasks with two alternatives (.78) had a higher difference in fixation time than tasks with seven alternatives (.50) [F(1,138)=42.73, p<.0001]. The paired difference was also greater for tasks with two attributes (.75) than for tasks with seven attributes (.53) [F(1,138)=28.71, p<.0001]. The interaction between alternatives and attributes shows that the difference in mean fixation duration is greatest for tasks with two alternatives and two attributes and least for tasks with seven alternatives and seven attributes. Thus, mean fixation duration is three times longer with Mouselab than with Eyegaze.

Process Tracing Data Discussion

Results for percent information searched are counter to our predictions. Considering the cognitive effort associated with information search, we expected that subjects using Eyegaze would search more information. These data suggest that Eyegaze allows subject to be more adaptive to the data whereas Mouselab favors a more systematic search. Subjects using Mouselab made a more systematic search and considered more information. It appears that subjects using Eyegaze considered less information by being more adaptive to the data and ignoring data that were not relevant to solving the problem. Another possible explanation is that subjects were able to gather information from the periphery of their eye without making a saccadic eye movements. This type of information acquisition is not captured with the Eyegaze System. It is also important to note that the analyses only compares data that Mouselab was capable of capturing - about 60% of total task time. Data about scanning to row and column labels on the display is not captured with Mouselab. Thus, we can not determine what additional information subjects viewed using Mouselab.

Payne [20] observed a general tendency for search direction to become more intradimensional (column wise) as information load increased. Payne and Braunstein [21] reported that search direction became more intradimensional (column wise) as the number of alternatives increased. Cook and Swain [11] report similar results. Our findings also found that search direction became more intradimensional (column wise) as the number of alternatives increased. We also found that search direction is contingent upon the process tracing method. Search direction for the same subject tended to be in one direction for Eyegaze and in the opposite direction for Mouselab. Mouselab tended to exhibit interdimensional search whereas Eyegaze tended to exhibit intradimensional search. The difference in search direction can be attributed to the effort required to extract information. Mouselab users may have adopted a less effortful search pattern. A natural reading order (left to right and top to bottom) dominants the information acquisition behavior for subjects using Mouselab. This may facilitate mental accounting of what information has been and has not been seen.

Russo [26] reported a reacquisition rate of 7% for information boards and 56% for eye tracking equipment. Reacquisition rate for Eyegaze, 70%, exceeds the value of 56% reported by Russo. These differences probably reflect differences in stimuli and eye tracking equipment. The reacquisition rate of 47% for Mouselab is much closer to the rate for eye tracking equipment than the rate for information boards. Mouselab has a much higher reacquisition rate than an information board which reflects the smaller effort required to acquire one piece of information using Mouselab than an information board. The pattern of differences shows that subjects using Mouselab reacquired more information for the large tasks than for small tasks. One possible explanation is that subjects memorized information for small tasks and performed mental calculations internally. For large tasks, it would be too difficult to memorize information.

Subjects adopted information acquisition strategies to cope with the increased effort of using Mouselab. These strategies tend to be top-down, goal-directed, more rigorous, and systematic. Increased variability in search patterns suggests that Eyegaze caused more selective, data-driven (bottom-up) information processing. Our findings suggest that consumers are even more adaptable to data than previously thought and shift their information processing strategy accordingly.

The overall Mouselab mean fixation duration ranged from 1.034 to 1.630 with a mean of mean of 1.267. The overall Eyegaze mean fixation duration ranged from .341 to .439 with a mean of .377. Both the Mouselab and Eyegaze fixation times are higher than the timing parameters reported by Card, Moran, and Newell [10], 1.10 and 0.230 respectively. The Mouselab time also is slightly longer than value of 1.19 reported by Bettman, Johnson, and Payne [5]. The values reported by Card, Moran, and Newell are for skilled, error-free, ideal behavior. It is not uncommon to find longer times for empirical data from unskilled users prone to errors. More importantly, the gamble data reflect mental arithmetic calculations in addition to acquiring a piece of information. Thus, it is not surprising that these times are slightly longer than those reported previously.

Implications for Research

Comparison studies As long as data collected with CPT tools compare theoretical differences between two treatments, then the CPT tool should not cause any systematic bias of the data. Of course, this assumes there is not an interaction effect between information acquisition effort and the dependent variable of interest. In addition, studies that compare groups using a CPT tool to a control group without a CPT tool must consider the possibility that the results are an artifact of the tool. CPT tools impose a greater information processing burden especially on a capacity constrained working memory. This, in turn, could affect process measures such as total time to complete the task.

These results suggest that consumers are even more adaptable to the data than previously thought and shift their information processing strategy accordingly. We caution researchers about making strong implications about using their CPT findings to aid the design of consumer information systems, decision support systems and user interfaces. More importantly, it is unlikely that unaided subjects use the same information acquisition patterns as aided subjects. For example, Johnson, Payne, and Bettman [16] report an average search pattern from Mouselab data. It will be difficult to compare this value across studies using a different CPT tool because of the different amount of effort involved. Search patterns found using a specific CPT tool would probably not generalize beyond the scope of the study. More importantly, it is unlikely that these patterns would indicate the search patterns of unaided subjects.

Directions for Future Research

Process tracing tools have largely been used for small choice tasks with 2-10 alternatives and 2-8 attributes [16, 28, 31]. Often the CPT tool limits the size of the choice task because of pragmatic concerns such as font size in an 80 column by 40 row character-based display. These concerns suggest that some stimuli are not appropriate for CPT tools. The stimuli include: non-matrix displays, multi-page sequential displays, displays with missing information, and tasks with a large number of alternatives and attributes. Each of these are important areas for future research exploring the effects of information acquisition patterns on consumer choice.

In contrast, eye-tracking equipment is useful for large choice tasks with an irregular arrangement of information (e.g., Yellow Pages directories [18] and retail catalogs [14]. Eye-tracking equipment has limited precision and accuracy for detecting small regions on a display. Current eye-tracking systems are able to detect regions with an area of 1.5 square centimeters. Unfortunately, the cost of eye-tracking equipment ($20,000-$100,000+) is prohibitive to many research institutions. However, significant differences in information acquisition and choice behavior contingent upon the process tracing tool suggests that an investment in better process tracing tools might be warranted, especially for the following stimuli.

Non-Matrix Displays Contemporary information displays rarely present a matrix of alternatives and attributes in the row-column format used by Mouselab and other CPT tools. Eye movement equipment allows researchers to use test stimuli with very irregular arrangements of information. Examples include: retail catalogs, newspaper advertisements, Yellow Pages directories, retail kiosks, multimedia information services, and digital information products for electronic commerce. Display format has a strong influence on information acquisition and subsequent consumer behavior. For example, Russo [24] induced supermarket shoppers to purchase products with lower unit prices by providing unit price information on a single list. Further, consumer information systems could be developed that influence consumer purchasing behavior or provide personalized electronic catalogs that reflect consumer preferences.

Scanning sequence The American Airlines airline reservation system is a classic example of the effects of scanning order on choice. The Sabre reservation system listed American Airline flights first. Being first resulted in more bookings. The government tried to block this unfair use of reservation systems by forcing the systems to begin giving competing flights equal display on computer screens. After a 12 year legal battle, Sabre ordered flights randomly.

Like the inability to make an optimal choice in the apartment selection task, it is also difficult to identify an optimal choice using an airline reservation system. A particular choice reflects individual preferences on a number of attributes as well as the number of alternatives. As the choice set becomes larger, people tend to consider less information in making their choice. Given large information search costs, people satisfice and select an alternative that is "good enough". Thus, scanning patterns have implications for interface design of consumer information systems, especially for systems which force consumers to process information serially. The matrix format used by CPT tools is not useful for studying scanning sequence effects, especially for long multiple page listings.

Incomplete Information Very little research has explored the effects of incomplete or missing information on choice [9, 29]. The matrix information display used by CPT tools heightens the awareness of missing information (an empty cell). However, it is generally much more difficult to discern what information is missing when the consideration set is large and the number of attributes is also large. If information were organized in a non-linear manner, missing information would be much less noticeable.

Large Problem Spaces In many instances, consumers make choices from dozens of alternatives each with numerous attributes. Large problem spaces (greater than 56 cells) are not well suited for study with CPT tools. The common use of multiple page layouts for airline flight reservations, Yellow Pages ads, and newspaper classifieds exacerbates the difficulty of collecting information acquisition data for real world choice tasks. As the choice set becomes larger, the importance of human processing limitations such as working memory looms ever larger. Since CPT tools like Mouselab increase the information processing load on working memory, it is likely that researchers will need to use eye-tracking equipment or verbal protocols to study information acquisition strategies for large problem spaces.

Conclusion

While CPT tools provide important process tracing data, they increase the amount of effort needed to acquire information. Our results show that subjects adopt information acquisition strategies to cope with the increased effort. These strategies tend to be more rigorous and systematic than those observed with eye-tracking equipment. Our findings suggest that consumers are even more adaptable to data than previously thought and shift their information processing strategy accordingly. Thus, CPT tools may fundamentally alter the information processing behavior they are believed to track unobtrusively by limiting the ability of the decision maker to adapt their information processing behavior dynamically to the demands of the data. Payne, Bettman, and Johnson [22] argue that choice rules are rarely "pure" and represent subcomponents of choice behavior. Most of the research on choice rules has been based on data collected with information boards [20], CPT tools [8, 11, 22, 31, 35], or analyses that did not use process tracing data [32, 33]. Only the study by Russo and Dosher [27] used eye movement process-tracing data. Since information boards and CPT tools fundamentally alter information processing behavior, some of these subcomponents of choice rules (e.g., lexicographic and elimination by aspects) might be artifacts of the information acquisition effort required when using CPT tools. Thus, descriptive choice rules may not reflect the information search patterns consumers use for large choice sets with incomplete information such as retail catalogs, newspaper advertisements, Yellow Pages directories, retail kiosks, multimedia information services, and digital information products for electronic commerce. Additional research is needed to explore the magnitude and consequences of these differences.

Bibliography

[1] Abelson, R. P. & Levi, A. (1985). Decision making and decision theory. In G. Lindsay & E. Aronson (Eds) *The Handbook of Social Psychology*, 3rd edition, 231-309.

[2] Ashton, A. H. & Ashton, R. H. (1988). Sequential belief revision in auditing. *The Accounting Review*, 63(4), 623-641.

[3] Baty, J. B. II, & Lee, R. M. (1995). Intershop: Enhancing the vendor/customer dialectic in electronic shopping. *Journal of Management Information Systems*, 11(4), 9-31.

[4] Beales, H., Mazis, M. B., Salop, S. C., & Staelin, R. (1981). Consumer search and public policy. *Journal of Consumer Research*, 8, 11-22.

[5] Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organization Behavior and Human Performance*, (45)1, 111-139.

[6] Bettman, J. R., & Kakkar, P. (1977). Effects of information presentation format on consumer information acquisition

strategies. *Journal of Consumer Research*, 3, 233-240.

[7] Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, 41, 1-19.

[8] Brucks, M. (1988). Search monitor: An approach for computer-controlled experiments involving consumer information search. *Journal of Consumer Research*, 15, 117-121.

[9] Burke, S. J. (1990). The effects of missing information on decision strategy selection. In M. E. Goldberg, G. Gorn, & R.W. Pollay (Eds.), *Advances in consumer research* (Vol. 17, pp. 250-256). Provo, UT: Association of Consumer Research.

[10] Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

[11] Cook, G. J. & Swain M. R. (1993). A computerized approach to decision process tracing for decision support system design. *Decision Sciences*, 24(5), 931-952.

[12] Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 32, 52-88.

[13] Hitch, G J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, 17, 417-444.

[14] Janiszewski, C. (1992). The influence of page layout on directing and sustaining attention. Working paper, College of Business Administration, University of Florida.

[15] Jarvenpaa, S. L. (1989). The effect of task demands and graphic format on information processing strategies. *Management Science*, 35(3) 285-303.

[16] Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, 42, 1-21.

[17] Logie, R. H. (1986). Visio-spatial processing in working memory. *Quarterly Journal of Experimental Psychology*, 38A, 229-247.

[18] Lohse, G. L. (1995). An eye fixation analysis of consumer choice from phone directory advertisements. In (Ed.), *Winter Marketing Educator's Conference*, La Jolla, CA, February 11-14, Chicago, Il, AMA Press.

[19] Newell, A., & Simon H. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

[20] Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.

[21] Payne, J. W., & Braunstein, M. L. (1978). Risky choice: An examination of information acquisition behavior. *Memory and Cognition*, 5, 554-561.

[22] Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge, England: Cambridge University Press.

[23] Quinn, J. G., & Ralston, G. E. (1986). Movement and attention in visual working memory. *Quarterly Journal of Experimental Psychology*, 38A, 689-703.

[24] Russo, J. E. (1977). The value of unit price information. *Journal of Marketing Research*, 14(2), 193-201.

[25] Russo, J. E. (1978). Adaptation of cognitive processes to eye movement systems. In J. W. Senders, D. F. Fisher, and R. A. Monty, eds. *Eye Movements and Higher Psychological Functions*, 89-109, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[26] Russo, J. E. (1978). Eye fixations can save the world: a critical evaluation and a comparison between eye fixations and other information processing methodologies. In H. K. Hunt (Ed). *Advances in consumer research (Volume 5)*, 561-570, Ann Arbor, MI: Association for Consumer Research.

[27] Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 676-696.

[28] Schkade, D. A. & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, 44, 203-231.

[29] Slovic, P. & MacPhillamy, D. Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance*, 6, 649-744.

[30] Smyth, M. M. & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology*, 41A, 235-250.

[31] Todd, P., & Benbasat, I. (1991). An experimental investigation of the impact of computer based decision aids on decision making strategies. *Information Systems Research*, 2(2), 87-115.

[32] Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1), 31-48.

[33] Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281-299.

[34] Widing, R. E. & Talarzyk, W. W. (1993). Electronic information systems for consumers: an evaluation of computer-assisted formats in multiple decision environments. *Journal of Marketing Research*, 30, 125-141.

[35] Williams, J. D. (1990). Risk assessment: An investigation of auditor's decision processes. *Proceedings of the Audit Judgment Symposium*, University of Southern California.

[36] Young, L. R. & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods and Instrumentation*, 7(5), 397-429.