Impact of numerical and graphical formats on dynamic decision making performance: an eye-tracking study

Cleotilde Gonzalez
Social and Decision Sciences
Carnegie Mellon University
Pittsburgh, PA 15213 USA
412 268 6242
conzalez@andrew.cmu.edu

Janice Golenbock
Social and Decision Sciences/ Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
412 268 9547
jsg@andrew.cmu.edu

ABSTRACT

This paper presents a study in which we manipulated the interface of a computer simulation into: graphical and numerical formats. We obtained both performance and eye-tracking learning curves from individuals assigned to one of these two conditions. Our findings indicate that although performance is not different between the two interfaces, the amount of attention as measured by the number of eye-tracking points was very different in the graphical and numerical conditions. Attention increased over time in the numerical condition, but was stable in the graphical condition. These results showed that the strategies used to make decisions in dynamic environments vary according to the form of information presentation.

Keywords

Eye-tracking, attention, graphical and numerical interface.

INTRODUCTION

It is clear that decisions made using computers are highly influenced by the interface design. Information may be presented in multiple formats, and these might influence the cognitive processes we follow in evaluating and comparing alternatives to make decisions.

Many characteristics of the interface have been investigated for their influence on decision making effectiveness: color, the pace of information presentation, and graphical and tabular interfaces. Color has been found to improve both time and accuracy performance in decision tasks [1]. Also, the pace of an animated interface influences decision effectiveness, where choppy animations degrade decision performance [2]. Finally, the form of information presentation, as numerical and graphical formats, is determinant for some decisions [1, 3]. It has been found that tabular charts, where numbers can be compared directly, result in superior performance compared to graphical charts [4]. Tabular numerical reports are found to be best for analytical thinking, since decision makers may compare specific numerical values of multiple options [5]. On the other hand, graphical interfaces may enhance recall in spatially-oriented tasks (tasks in which the positions of the graphical elements are important in making decisions) [6].

Research studies on the effect of interface characteristics on decision making performance have one thing in common: they deal with simple, static decision making tasks. A static decision making task does not change over time, as the decision maker is trying to make up his mind about the benefits and costs of alternatives. Frequently in these tasks options are explicitly given, such that the decision maker engages in the comparison process directly. Decision making frequently involves the comparison of at least two alternatives, and the computations of their expected value. Decision makers are not under time pressure to solve these tasks. Under these conditions, it is evident that numerical formats have an advantage over graphical interfaces. In this paper however, we do not deal with static decision making but rather with dynamic tasks.

Dynamic Decision Making (DDM) involves a series of multiple and interdependent decisions made in real-time in a continuously changing, autonomous environment [7, 8]. In these tasks, it might not be possible to act analytically, and to carefully evaluate the costs and benefits of different alternatives, due mainly to lack of time. For example, it has been shown that, although more accurate, tabular and numerical data may need significantly longer response time compared to graphs [9]. Therefore, in DDM, since the time of intervention is dictated by the environment rather than by the decision maker, graphical interfaces might be more beneficial than numerical ones.

In this paper we present the results of individuals performing a DDM task formatted in one of two ways: numerically and graphically. We test for differences in learning the task under these two conditions, and we present physical evidence of attention allocation under these two conditions. Eye-tracking is yet a relatively young field, and very little is known on how attention varies over time, specifically in dynamic tasks. But, for what is known in Psychology in terms of attention and learning theories, we would expect attention to drop as participants practice the task [10]. We would not be able to predict if the reduction of eye-tracking points would be higher or lower depending on the form of presentation.

METHODS

Dynamic Decision Making Task

A dynamic decision making simulation used in previous research was also used for this study [11]. The simulation was called the Water Purification Plant (WPP). A screen shot of the WPP simulation is provided in Figure 1. WPP simulates a water distribution system. The system is made up of chains of tanks and each chain is assigned a particular deadline by which a participant must distribute all the water out of the chain. The distribution of water occurs by the participant opening or closing pumps in the chain. The simulation is a dynamic environment, where quantities of water in any of tanks may increase (water input from outside of the system) without the participant's knowledge. WPP is a resource allocation task, where the maximum number of pumps opened is 5. The decisions that a participant makes are interconnected: opening a pump in one chain may prevent one from opening the water flow in another chain. The main performance measure is the number of gallons of water that were left in the system by the participant after each deadline. For the purposes of this study, it was possible to achieve the goal of zero gallons of water missed, indicating that the participant delivered all the water on time. Under a different scenario zero may not be possible but the goal is still to minimize the amount of water left behind, i.e. there is always a minimal solution to the task.

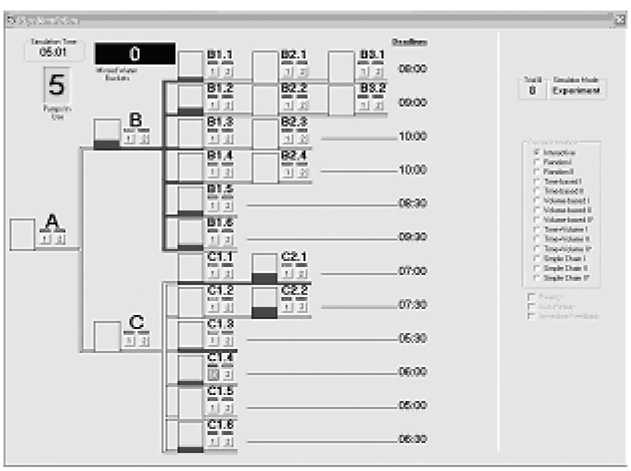

Figure 1. The layout of the WPP task

The eyetracker apparatus

The EyeLink eyetracker was produced by SensoMotoric Instruments, and allows the experimenter to know where the subject was looking at each millisecond of the simulation. The eyetracker consists of an apparatus worn on the subject's head with a small camera pointed at each eye. Sensors on the monitor help track where the head is moving and which direction it is pointing, to help calibrate the eye movement data.

Each day, each participant is calibrated in the eye tracker. This involves a set of steps where the subject follows a dot on the screen. After "Good Calibration" is reached, the data collection using WPP begins, with calibration in between each trial.

Experimental Design

This was a 2 x 16 mixed experimental design. The between factors variable is the form of information presentation (Graphical or Numerical) and the within factors variable is the number of practice trials allowed for each subject (1 to 16). The graphical version of the interface is shown in Figure 1, and the numerical version is shown in Figure 2.

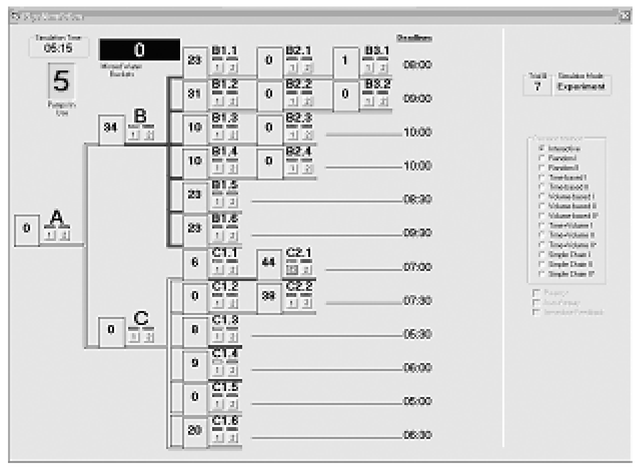

Figure 2. Numerical version of the WPP task.

Participants

Twenty-one College students from Carnegie Mellon University and the University of Pittsburgh participated in this study. Ten participants were randomly assigned to the graphical condition and eleven to the numerical condition. They averaged age of 23. Ads were sent out to recruit subjects, and they signed up online for a given time-slot.

Data collection and preparation

Behavioral data

The main performance variable is the total number of gallons missed. Participants had a running counter in the upper left corner indicating the number of missed gallons, updated after each deadline passed. There are many possible decision sequences for activating and de-activating pumps and achieving the optimal performance of zero (pumping all the water buckets in time). As a reasonable upper limited yardstick for performance, we ran the simulation making random assignments, maintaining no idle time (that is, never having idle pumps). We call this strategy the zero intelligence scheduler. The results for 30 replications of random assignments were a mean of 182.9

missed buckets with a standard deviation of 28.4. Therefore, reasonable performance is between zero and 200.

Eye-tracking data

The eyetracker software outputs data files with 250 samples every second. Each sample contains: Time, the *x*-axis location for the left and right eyes (LeftX, RightX), the *y*-axis location (LeftY, RightY), the pupil size of the left and the right eye. Each trial of the simulation lasted 8 minutes. Therefore we had about 120,000 points in each file. Each trial ran by each subject produced one of these files. So that, in total, we had 336 (21 times 16) files with eye-tracking data.

Unfortunately, without a chinrest and with the long time per trial, the eyetracking data resulted in poor final calibration. To recalibrate the data after collection, we used a LISP program developed at CMU for this purpose [12]. This program uses the Interpretation Tree Algorithm which uses pattern matching to match the data we collected to a model we define previously. This algorithm can therefore detect and correct systematic bias caused by miscalibration. The data analyses show the number of samples as they change over time and per condition.

Procedures

The data for this experiment was collected over a 4-week period. Each week for 4 weeks, data from 5-7 subjects were gathered, since only one subject could run the simulation at a time and each session took one hour.

On the first day, participants were given instructions on the objective of the task and how to use the simulation to perform the task. Instructions were provided by following a standard script, and by allowing the participants to run the simulation in a training mode (at a very slow speed of 30 minutes for a single trial). Participants were randomly assigned to either the Graphical or Numerical condition for the last four days (days 1-4 of experimental sessions). During the experimental hours, participants ran four 8-minute trials each day while wearing the eyetracker.

RESULTS

Our first hypothesis was that we would see a difference in the way participants learn the dynamic task while using a Graphical versus Numerical interface. Figure 3 shows the average performance per trial for all individuals in the two interface conditions. The statistical analyses showed significant learning overall ($F(15,285)=17.91$, $p<.000$), but not significant difference between numerical and graphical conditions ($F(15,285) = 1.41$, n.s.). Although not a significant result, Figure 3 shows better performance by individuals assigned to the Graphical condition in most of the trials.

Our second hypothesis was that there would be a reduction of attention as participants practiced the task. We expected an overall attention reduction, but did not know of enough previous research to expect a difference in attention between Graphical and Numerical interfaces. Our

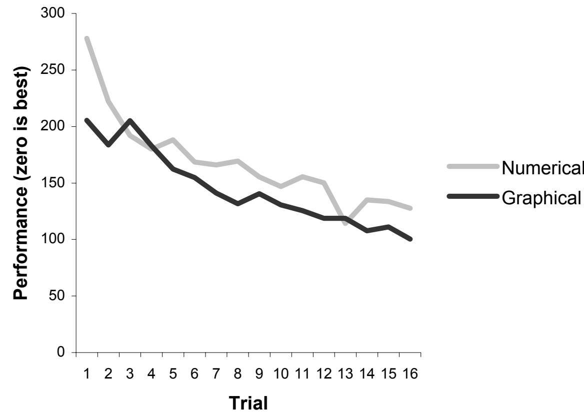

Figure 3. Performance in Graphical and Numerical interfaces.

measure of attention is the number of fixation points over time. The statistical analyses showed no difference in the number of eye tracking samples over time ($F(15,285)=1.18$, n.s.).

Figure 4 shows the average sample points per trial for both numerical and graphical conditions.

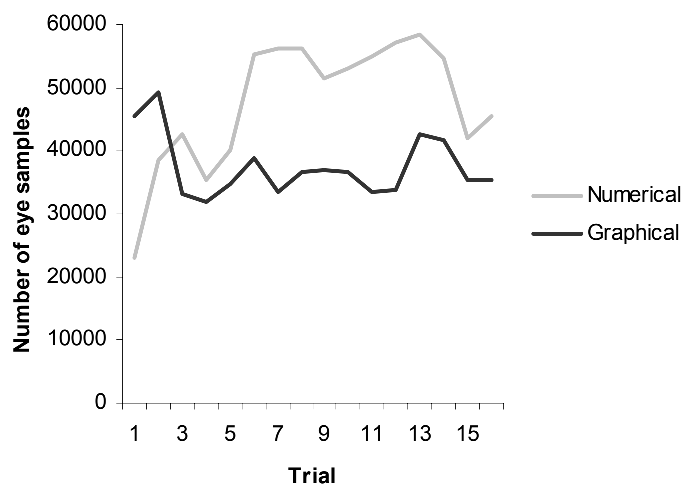

Figure 4. Fixation points in Graphical and Numerical Interfaces

According to this data we could expect a significant difference in the fixation points due to the numerical and graphical interfaces. The statistical analysis showed that the attention learning curves are different for these two interfaces ($F(15,285=1.82$, $p<.05$). Analyses of learning that occurred within each of the interfaces indicated that in the numerical condition the amount of attention as measured by the number of eye-tracking points increased significantly over time (F(15,150)=4.33, p<.000). On the other hand the amount of attention in the graphical interface did not change significantly over time (F(15, 135) = .38, n.s.).

DISCUSSION

Although our results showed no significant difference in performance of people assigned to the numerical and the graphical interfaces, we can see that individuals working in the numerical interface do slightly worse than those assigned to the graphical interface. This result suggests there might be an advantage in working in a graphical interface in a DDM task. The main advantage, we think, is that graphical interfaces do not promote analytical thinking. Analytical thinking may help performance in situations in which *timing* of the decisions is not important for their accuracy. When individuals have unlimited time they should be more accurate using numerical interfaces rather than graphical interfaces. However in DDM the time allowed to consider different alternatives is limited. The pace of decisions is determined by the environment rather than by the decision maker. This hypothesis seems to be confirmed by the eye-tracking data. Individuals in the numerical interface increased their number of attention points in the interface, while individuals in the graphical interface did to change attention over time. Perhaps the numerical interface demands one to pay more attention to the numbers, which are used to calculate the value of the alternatives, slowing down the decision process. This is not a general result and we have not tested it on other types of tasks. This result might change according to the characteristics of each specific task.

Contrary to our expectations, the total number of eye-tracking points did not decrease over time. In this analysis we used the raw number of points capture by the sampling rate (which was constant) across trials. We believe that attention must be determined not only by the number of points, but rather by the actual fixations (the length of time attending to an area) and saccades (movements from between areas in the screen) across the display. In future analyses of this data we plan to use EyeTracer, a tool designed to address the identification of fixations and saccades from raw eyetracking data [12]. Another future experiment is to have participants use one mode and then the other. We could do a post-interview to discover users' preferences and their perception of their performance in each mode. This would help determine which interface users would choose given the choice.

ACKNOWLEDGEMENTS

This research was supported by the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory (DAAD19-01-2-0009), and the Office of Naval research (N00014-96-1-0322).

REFERENCES

1. Hoadley, E., *Investigating the Effects of Color.* Communications of the ACM, 1990. **33**(2): p. 120-125.
2. Gonzalez, C. *Does animation in user interfaces improve decision making?* in *Proceedings of the International Conference on Human Factors in Computing Systems. CHI'96*. 1996. Vancouver, Canada.
3. Kennedy, M., D. Te'eni, and J.B. Treleaven, *Impacts of decision task, data, and display on strategies for extracting information.* International Journal of Human-Computer Studies, 1998. **48**(2): p. 159-180.
4. Davis, L.R., *Report Format and the Decision Maker's Task: An Experimental Investigation.* Accounting, Organizations and Society, 1989. **14**(5/6): p. 495-508.
5. Benbasat, I., A.S. Dexter, and P. Todd, *The Influence of Color and Graphical Information Presentation in Managerial Decision Simulation.* Human-Computer Interaction, 1986. **2**: p. 65-92.
6. Umanath, N.S. and R.W. Schmell, *An Experimental Evaluation of the Impact of Data Display Format on Recall Performance.* Communications of the ACM, 1988. **31**(5): p. 562-570.
7. Brehmer, B., *Strategies in real-time, dynamic decision making.*, in *Insights in Decision Making*, R.M. Hogarth, Editor. 1990, University of Chicago Press. p. 262-279.
8. Edwards, W., *Dynamic Decision Theory and Probabilistic Information Processing.* Human Factors, 1962. **4**: p. 59-73.
9. Engberg, J.M. and F.L. Wallace, *Information Extraction from Different Data Representation Forms on a CRT: Charts and Tables*. 1994, University of North Florida: Jacksonville, FL.
10. Logan, G., *Toward an Instance Theory of Automatization.* Psychological Review, 1988. **95**(4): p. 492-527.
11. Gonzalez, C., F.J. Lerch, and C. Lebiere, *Instance-Based Learning in Dynamic Decision Making.* Cognitive Science, In Press.
12. Salvucci, D.D. *An interactive model-based environment for eye-movement protocol analysis and visualization*. in *Proceedings of the symposium on Eye tracking reserach & applications 2000*. 2000. Palm Beach Gardens, Florida, United States: ACM Press.