Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V.
(North-Holland) © IFIP, 1985 Pages 737-741

ADAPTING A PSYCHOPHYSICAL METHOD TO MEASURE PERFORMANCE AND PREFERENCE TRADEOFFS IN HUMAN-COMPUTER INTERACTION

Jonathan Grudin, Wang Laboratories, Lowell, Mass., USA
Allan MacLean, MRC Applied Psychology Unit, Cambridge, England

ABSTRACT
An experimental methodology for contrasting certain design alternatives and quickly determining user preferences and performance tradeoffs is presented. It is shown how this experimental paradigm, used for psychophysical measurement, may be applied to the field of human-computer interaction. Where it can be applied, it promises a relatively quick determination of user preference and performance characteristics and tradeoffs on these measures with variation in parameters governing the user situation. Because the methodology is within-subject, it may also facilitate the study of individual differences.

INTRODUCTION

Often there are no optimal design features -- most features involve tradeoffs. The usefulness of a given feature may depend critically upon parameters that change within or across applications. (For a theoretical treatment of this issue, see Norman [1].)

However, the use of most standard experimental paradigms to determine the interacting influences of even a small number of variables may require a huge and often prohibitive investment of resources. A technique for collecting a large amount of data in as short a time as possible would obviously be highly desirable. This paper describes the adaptation to human-computer interaction of a technique designed to explore economically both performance and preference. The technique is in fact one of the oldest experimental paradigms in psychology -- psychophysical measurement.

Although the method has theoretical and pragmatic limitations (e.g., see Poulton [2]), it can potentially be used to investigate many situations where a tradeoff exists along a particular dimension. In a field quite distant from human-computer interaction, it is the method used by the optometrist when fitting lenses. The patient is asked to look through one lens, then through a second lens, and states his or her preference. The optometrist then varies one or both lenses along one or more dimensions and repeats the procedure, quickly homing in on the best fit.

This study is an illustration of how this method can be applied to the exploration of preference and performance tradeoffs in computer systems. We arbitrarily chose to examine two specific data entry methods for the purpose of illustrating the technique. We explored two procedural variations of the psychophysical method, our interest being more in that method than in analyzing data entry techniques. This is the first report of an experimental paradigm still under development, not a finished study of data entry alternatives.

THE DATA ENTRY TASK

On each trial a name appeared in the upper left of a VDU, and the subject was required to enter the code number associated with that name. The right side of the display consisted of an alphabetized

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V.
(North-Holland) © IFIP, 1985 Pages 737-741

directory of names and associated code numbers. Two different code entry methods were tried by each subject. (The procedures by which one method or the other was decided upon are described below.) One code entry method consisted of simply looking up the name in the directory, reading the associated number, and typing in that number using the keys above the top row of letters on the keyboard. Following the typing of the number the subject pressed the "ENTER" key. The second code entry method involved the use of an "option ring". When the subject pressed the space bar, the first name and associated number from the directory appeared in a small reverse-video window on the left side of the screen, below the target name. Each time the space bar was struck, the next name and code number from the directory appeared in the window. The subject thus used the space bar to step through the directory until the target name with its associated code number appeared in the window. Then the subject struck the ENTER key and the code number was entered, ending the trial.

In the code entry example in Figure 1, the code number is half entered. In the option ring example, the subject has begun stepping through the option ring.

A great many factors, ranging from the organization and layout of information on the screen (in the "directory") to performance incentives, could be varied within such a paradigm. In these studies, we varied two. One was the directory size, or the number of names through which the subject might have to search to find the target name and code number. The other was the length in digits of the code number that the subject had to enter.

Next Code Needed: SAGGS

Code Entry

Type in Code Number
Then Press ENTER Key

9_

CODE DIRECTORY

ANTCLISS	43	LEVITT	20
BURGESS	18	NUTTING	80
CHILDERS	58	SAGGS	90
DAVIDSON	42	THORPE	38
FROY	17	UPSON	63
HUNT	21	WILSON	75

Next Code Needed: RINALDI

NATHAN 295109

Bar   Advances Option Ring
B   Backs Up Option Ring
ENTER Select Current Option

CODE DIRECTORY

ABBOTT	298734	LARGE	921022
BARNARD	510813	MARSHALL	781102
BREADMORE	549885	NATHAN	295109
CASEY	816420	PALMER	610083
COMINS	359071	RINALDI	891638
DOYLE	309812	SCOTT	644201
FINLAY	10084	STEVENS	523551
GREEN	891119	TRACEY	646209
HEDGE	287364	WHITEHEAD	784928
JOHNSON	654291	WRIGHT	198617

Figure 1. Code entry display (top) and option ring-display.

We expected the setting of these parameters to influence both the entry method producing better performance and the entry method subjects would prefer. With a directory of many names and one-digit code numbers, subjects would presumably prefer to scan the directory

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V. (North-Holland) © IFIP, 1985 Pages 737-741

visually for the name and enter the single digit. But, where the directory had few names and the code numbers were long, the subject would presumably prefer to step through the option ring to the target name and hit the ENTER key, avoiding the need to remember and type in the several digit code number. What we sought was a methodology to tell us what the performance curves for the two entry methods would be, and thus where the cross-over point from one entry method to the other would occur. We also were interested to know which entry method a subject preferred for a given parameter setting, and whether the performance and preference cross-over points were identical.

EXPERIMENT ONE

The basic procedure was as follows: A given directory size and code length were established. The subject was asked to look at the directory and choose one or the other code entry method. A target name then appeared, and the subject had to enter the associated code number using the entry method just selected. Following the subject's completion of that task, the directory size and/or code number length were changed and the process repeated.

Because the subject selects the code entry method prior to each trial, this method produces considerably more preference data. However, there may not be data for a direct comparison of performance with the two entry methods for a given setting of directory size and code length. In addition, the subject's choices are not immediately preceded by experience with both entry methods. These latter concerns are addressed in Experiment Two. The large number of preference decisions does allow a relatively efficient discovery of the preference tradeoff points.

<u>Method</u>. Four subjects were run for a single two-hour session. Each subject had 5 blocks of 6 practice trials to acquaint them with the code entry methods. (Code lengths were 4,1,1,7,7; directory sizes were 16,4,36,4,36 respectively.) Then the procedure to be used in the experiment was introduced with a set of practice trials with code length 4. In this procedure, a code length was established for a set of trials. On the first trial, the directory size was 16. Prior to the identification of the target name, the subject chose the entry method by pressing the space bar (for option ring) or escape key (located adjacent to the row of numbers on the keyboard, for code entry). Then the target name appeared, and they entered the associated code number. Thus, their choice of entry method could reflect code length and directory size, but not the position of the target in the directory. The directory size was changed for the next trial, being incremented if the option ring was chosen and decremented if code entry was chosen. The step size began at 8. After three changes in preferred entry method, the step size was reduced to 4, and after six reversals it was reduced to 2. (If the new directory size would have fallen outside the range of 4 to 40, it was made to be the closer of the two values.) The set of trials terminated when one entry method had been selected 3 more times than the other or after nine trials, whichever came first.

<u>Results</u>. For each subject and each code length, the method identified the tradeoff point as being the final directory size. The subject typically homed in on the point, switching entry methods and thus reversing directory size growth when crossing it, until the requisite number of reversals. Of course, the floor of 4 and ceiling of 40 for

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V. (North-Holland) © IFIP, 1985 Pages 737-741

directory size could prevent the tradeoff point from being reached, in which case the subject reached the boundary and stayed there. Figure 2 presents performance data from Experiment One. For each code length, the tradeoff point represents the directory size for which the subjects on the average switched from using option ring (for the smaller sizes) to code entry (for the larger). Thus, the region above the solid line represents the parameter values for which code entry would be preferred, and the region below the line represents the parameter values for which subjects tended to prefer the option ring. The psychophysical method appeared to produce a clean determination of this tradeoff curve.

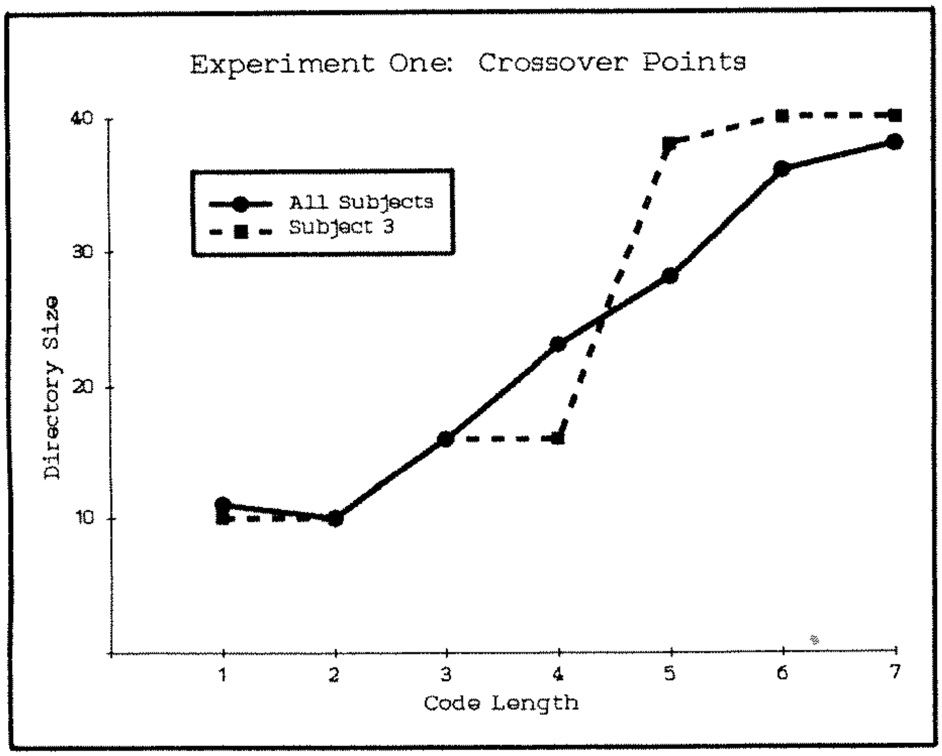

Figure 2. Tradeoff curve. Parameter space above solid line is code entry preference, below the line option ring is preferred.

Of course, not all subjects adhere closely to the mean preference curve. Whatever their drawbacks, within-subject measures afford a look at individual differences, and the psychophysical method is designed for efficiency in this regard. The dashed line in Figure 2 represents the preference data of one subject. This subject preferred the code entry method for all but the smallest directory sizes through code length 4, but relied almost exclusively on the option ring for longer code numbers. Other subjects showed other patterns. The degree to which such preference differences might be based on differences in memory span, or motivation, or other factors requires further exploration. Some people may tend to prefer the more repetitive but less tiring route of leaning on the space bar in the option ring method, while others may respond to the challenge of trying to commit the longer code numbers to memory. In any case, exploration of such individual differences may require a method such as this to uncover the differences in the first place.

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V. (North-Holland) © IFIP, 1985 Pages 737-741

EXPERIMENT TWO

Experiment Two was essentially a procedural variant of Experiment One. The basic procedure was as follows: A given directory size and code length were established. Then, one of the two code entry methods was randomly selected. The subject was given four target names, one at a time, and asked to enter a code number for each using the specified code entry method. Then the subject used the other code entry method on four new targets. Finally, the subject chose the preferred code entry method to use on a final eight trials. Following those eight trials, the directory size and/or code number length were changed and the process repeated.

For each directory size - code number length pair examined, we recorded the following information: the time to complete four trials with each entry method, the subject's preference, the subject's accuracy rate, and the subject's time to complete the final eight trials. This procedure provided a very controlled comparison of performance data. It also insured that the subject had experience with both code entry methods immediately prior to choosing between them. Neither' of these was obtained with the procedure of Experiment One. However, with 16 trials for each preference decision, it yielded less preference data than did the procedure of Experiment One, despite requiring longer participation by each subject.

Method. Four subjects were run. One was run in three sessions of one hour apiece and three were run in two sessions of two hours apiece. In each case, the first session was considered practice and not examined, leaving two hours of data for analysis. Three subjects were completely naive to the purpose of the experiment, and one (Subject 4) had some familiarity with it. (The design of this exploratory study was somewhat looser than that of Experiment One.)

In both practice and recorded sessions, subjects were given a block of trials with each of the code lengths from one to seven, pseudo-randomly ordered. An initial directory size was selected. Unlike in Experiment One, the initial directory size varied with the code length. The intention was to begin with a directory size distant from the hypothesized tradeoff point, in order to collect performance data across a wider range of parameter values. Step direction was tied to entry method choice as in Experiment One. Step functions were initially larger (as high as 16), decreasing by 4 with each second reversal to a minimum of 4. As in Experiment One, the directory size was kept between 4 and 40.

The code length for the block having been set and the initial directory size determined, one of the two entry methods was randomly chosen and the subject informed to use it for the subsequent four trials. A directory of names and numbers appeared. The four targets were randomly selected and presented with the constraint that their average distance from the beginning of the directory was half-way through. Then a new set of names and numbers (of the same code length and in a directory of the same size) appeared and the subject used the other entry method for four trials. Upon their completion, the subject was asked to choose the preferred method by typing "C" or "O". Then there were eight more trials on still a new directory of the same size and code length, using the chosen method (giving the subject a reason to choose carefully).

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V.
(North-Holland) © IFIP, 1985 Pages 737-741

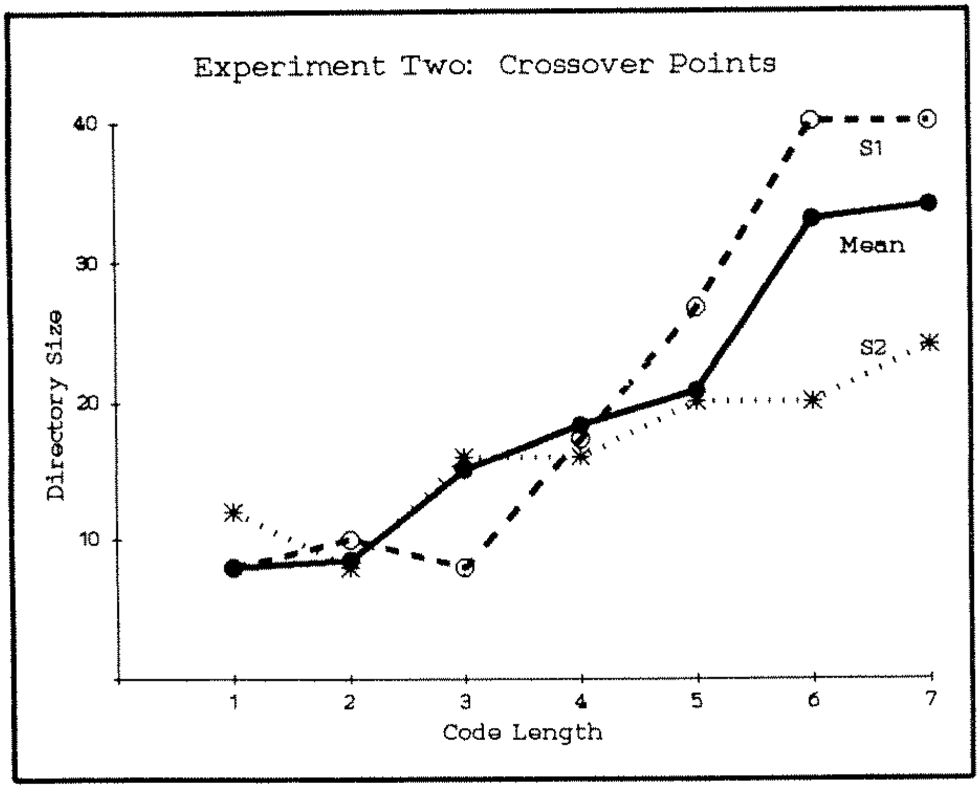

Figure 3. Tradeoff curve. Parameter space above solid line is code entry preference, below the line option ring is preferred.

Following the completion of the eight trials, the directory size would be adjusted as described above and another set of 16 trials begun. Following a number of such 16-trial sets, the block would terminate and a new code length would be chosen.

We tried minor variations on the number of 16-trial sets prior to changing code lengths. The 3-session subject was given shorter sets and went through each code length twice, while the other subjects had longer sets and saw each code length for just one block (following the practice session).

Results. The tradeoff curve in Figure 3 was calculated in the same way as that of Figure 2 of Experiment One. The average of the two tradeoff determinations was used for the one subject (Subject 1) who had two blocks of real trials with each code length. The mean curve is very similar to (hat of the first experiment, each based on 4 subjects. Individual data for two subjects are shown, once again for the purpose of indicating the underlying individual differences captured by this method. Subject 1 shows a pattern similar to Subject 3 of the previous experiment, except that Subject 1 does not completely abandon code entry for larger directories with code length 5. Subject 2 however is unusually unperturbed by code length changes.

This experiment allows us to determine whether performance (measured by completion time) matches preference. For the subjects studied, there may be a relatively close match when averaged, but there are

Human-Computer Interaction - INTERACT '84 / B. Shackel (ed.) Elsevier Science Publishers B.V. (North-Holland) © IFIP, 1985 Pages 737-741

clear individual exceptions. Subject 2, for example, is entering target numbers over twice as quickly with the option ring at code length 7, directory size 24 and code length 6, directory size 20, yet he consistently chose the code entry method for those parameter combinations, despite having just practiced with each. For code lengths 1-3, however, Subject 2 matches preference to performance. At the other end, Subject 4 (not shown) uses code entry for directory sizes of 4 with code lengths 1 and 2, despite performing more quickly with the option ring. These isolated examples indicate that the psychophysical method employed in this experiment offers one way to contrast performance and preference for certain tasks.

CONCLUSIONS

These exploratory studies indicate that the psychophysical method of threshold determination may be adapted in several ways to the exploration of certain human interface issues. First, though, we need more work on the adaptation of the method itself, and we must understand the limitations imposed by within-subject design and the repetitiveness of the task, which is artificial to most situations. However, the potential uses of the method are broad. By quickly mapping out the relationships among factors that may individually or collectively influence user preference and performance, it can produce results of both practical and theoretical significance. The method could he used to contrast specific alternative design options, prior to subsequent between-subject testing in the laboratory or in the field. It could balance our general reliance on performance measures with simultaneous measures of preference. The individual differences it uncovers could lead to investigations of their underlying cognitive and motivational structures. And the technique is also likely to be useful in evaluating and extending a model, such as that of Card, Moran, and Newell [3], when applied to parameters falling within the scope of the model.

ACKNOWLEDGEMENTS

The authors wish to thank Phil Barnard, Nick Hammond, and Richard Young for their contributions of ideas and criticisms.

REFERENCES

(1) Norman, D. A. Design principles for human-computer interfaces. Proc. CHI'83 Human Factors in Computing Systems (ACM, 1983) 1-10
(2) Poulton, E. C. Range effects in experiments on people. American Journal of Psychology 88 (1975) 3-32
(3) Card, S., Moran, T., & Newell, A. Applied information-processing psychology: The human-computer interface. (Erlbaum, 1983)