

# Ignoring Perfect Knowledge In-The-World for Imperfect Knowledge In-The-Head: Implications of Rational Analysis for Interface Design

Wayne D. Gray & Wai-Tat Fu  
 Human Factors & Applied Cognition  
 George Mason University  
 Fairfax, VA 22030  
 +1 703 993 1357  
 [gray/wfu]@gmu.edu

## ABSTRACT

Memory can be internal or external – knowledge in-the-world or knowledge in-the-head. Making needed information available in an interface may seem the perfect alternative to relying on imperfect memory. However, the rational analysis framework (Anderson, 1990) suggests that least-effort tradeoffs may lead to imperfect performance even when perfect knowledge in-the-world is readily available. The implications of rational analysis for interactive behavior are investigated in two experiments. In experiment 1 we varied the perceptual-motor effort of accessing knowledge in-the-world as well as the cognitive effort of retrieving items from memory. In experiment 2 we replicated one of the experiment 1 conditions to collect eye movement data. The results suggest that milliseconds matter. Least-effort tradeoffs are adopted even when the absolute difference in effort between a perceptual-motor versus a memory strategy is small, and even when adopting a memory strategy results in a higher error rate and lower performance.

## Keywords

interactive behavior, cognitive least-effort, errors, rational analysis, interface design, eye movements, eye tracking, direct-manipulation interfaces, satisficing

## INTRODUCTION

Knowledge can be in-the-world or in-the-head (see, e.g., Larkin & Simon, 1987; Norman, 1989). A well-designed interface can place knowledge in-the-world so that it is available in a known (or easily found) location when a user needs it. Placing knowledge in-the-world eliminates the need for the user to store knowledge in-the-head (estimated by Simon, 1974, to require 8s per chunk). The advantage of placing knowledge in-the-world versus requiring that it be stored in-the-head is widely touted as one of the main advantages of direct manipulation interfaces over command language ones (Frohlich, 1997; Hutchins, Hollan, & Norman, 1985; Shneiderman, 1982).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.  
 SIGCHI'01, March 31-April 4, 2001, Seattle, WA, USA.  
 Copyright 2001 ACM 1-58113-327-8/01/0003...\$5.00.

## The Research Issue from a Rational Analysis Perspective

Anderson (Anderson, 1990, 1991) casts human memory as an optimization process. In his rational analysis framework, the goal of human memory is to retrieve knowledge that would allow us to perform the task we are currently facing. The optimization process maximizes the expected utility of the memory system by balancing the cost of memory search against an assumed constant expected gain<sup>1</sup> of retrieving a desired memory item for the current task. A clear cost of memory search is time (and possibly a metabolic cost associated with time). Under Anderson's rational analysis framework, the human memory system would search a memory structure until the probability of getting the desired memory item (the expected gain) is lower than the cost of further search (i.e., when the expected utility becomes negative).

We take the position that the least effort perspective implied by rational analysis is sensitive to effort not to media (Fu & Gray, 1999, 2000). In this wider view, internal effort includes the effort of storing an item in memory as well as the effort of subsequently retrieving that item from memory. External effort includes the effort of searching the environment to locate an item as well as the effort of accessing an item at a known location. (Examples of “access” include an eye movement to an icon in a toolbar or moving the mouse and clicking on a pull-down menu.)

In most direct manipulation interfaces, information stays put; that is, icons, menu items, and tables have a fixed position. Hence, after a little experience, the effort of searching the environment is near nil (Ehret, 1999) and the only external effort of significance is the perceptual-motor (PM) effort involved in accessing the information. Hence, total effort can be considered as:

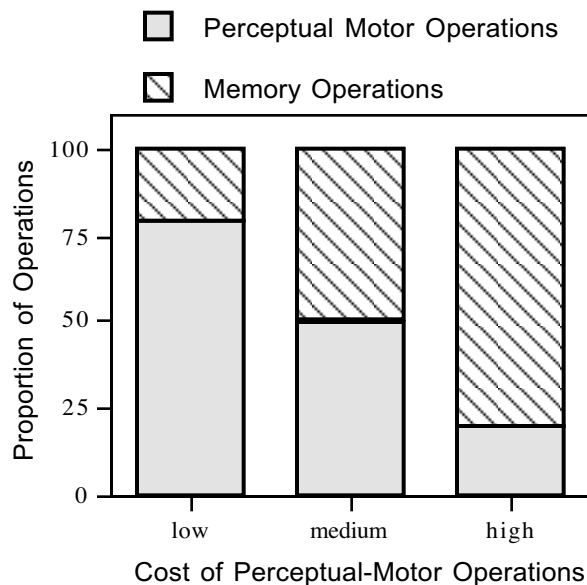
$$\text{TOTAL EFFORT} = \text{PM}_{(\text{access})} + \text{MEMORY}_{(\text{storage} + \text{retrieval})}$$

For users who have memorized information in advance (for example, “power users” who memorize keystroke

<sup>1</sup> The expected gain is defined as the product of P and G, where P is the estimated probability that the target memory item can be found, and G is the gain associated with retrieving the target memory item. If C is the memory search cost, then expected utility  $E = PG - C$ .

shortcuts), the *per interaction* effort for memory storage drops to zero.

As Figure 1 illustrates, all interactive behavior involves the perceptual-motor and memory systems. What differs from one situation to another is the relative proportion of memory or perceptual-motor operations. If memory effort (storage + retrieval) is held constant, then as perceptual-motor effort increases, people should use fewer perceptual-motor operations and more memory operations. That is, by making implicit least-effort tradeoffs, users will come to rely more on memory storage and retrieval and less on perceptual-motor access. Conversely, as perceptual-motor effort decreases, people should use more perceptual-motor operations and fewer memory operations.



**Figure 1: A notional chart illustrating how the relative proportion of perceptual-motor to memory operations changes as a function of effort of perceptual-motor operations.**

The rational analysis framework makes an interesting and, perhaps, counterintuitive prediction. As the effort of accessing perfect knowledge in-the-world increases, the cognitive system may satisfice with imperfect knowledge in-the-head. The weaker the target memory, the more likely a competing memory is to be retrieved in error. Hence, increasing the effort of perceptual-motor access may cause the cognitive system to access imperfect knowledge in-the-head with a concomitant increase in errors. Of interest to the CHI community is whether a range of perceptual-motor effort that is representative of the effort required by modern interfaces is sufficient to induce a reliance on imperfect knowledge in-the-head with a concomitant increase in errors.

#### Manipulating Internal versus External Effort

In this paper, we present two studies that investigated the influence on performance of the effort required to use knowledge in-the-world versus knowledge in-the-head. In selecting a task, two criteria were important. First, we needed a clear separation between using the task interface

versus accessing information for the task. Second, we wanted a task that would not force users to keep or manipulate information in-the-head; that is, storage in memory for more than a few seconds should be an optional, not a necessary requirement of task performance.

These criteria led us to select the task of programming a VCR to record a television show. Meeting our first criterion, the VCR interface was constant across conditions. With the task interface held constant, we varied the ease with which information for the to-be-recorded television show (i.e., start time, end time, day-of-week, and channel) could be retrieved from memory or accessed by the perceptual-motor system. Meeting our second criterion, the VCR did not require users to keep or manipulate information in-the-head. Information from the world could be obtained, used immediately, and then forgotten.

Experiment 1 used three conditions: Memory-Test, Free-Access, and Gray-Box. We manipulated the effort of memory retrieval during task performance by requiring the Memory-Test group to memorize show information before they could begin programming the VCR. Well-learned material is typically retrieved faster (i.e., requires less effort) than less familiar material. Hence, material memorized to a high criterion in advance of use, should be better learned and easier retrieved than material accessed during task performance. Therefore, for each to-be-recorded show, the Memory-Test participants would have well-learned and easily retrievable information available in in-the-head. Neither the Free-Access or Gray-Box groups were required to memorize information. (As discussed below, during programming, show information was available to all groups.)

The effort of perceptual-motor access was manipulated in the Free-Access versus Gray-Box conditions. As they programmed the VCR, the Free-Access group had information freely available to them in a Show Information Window located immediately below the VCR. This same window was in the same location for both the Gray-Box and Memory-Test conditions. However, rather than being freely available, the information fields for these two conditions were covered by gray boxes. Accessing a field required moving the cursor to and clicking on the gray box that covered the field.

Table 1 shows the effect that we expected our manipulations to have on the effort of memory storage, memory retrieval, and perceptual-motor access. As the Memory-Test group was required to memorize show information, the storage effort for them was expected to be greater than for the other two groups (see top line of Table 1). We had no reason to expect that the Free-Access and Gray-Box groups would differ in effort of memory storage.

For memory retrieval, we expected the Memory-Test manipulation to lead to this group having show information readily available in memory. The time needed to retrieve a well-learned memory should be around 100 msec (Altmann & Gray, 1999, 2000a, 2000b). In contrast, we assumed that while the Free-Access and Gray-Box groups would store some show information during the trial, they would store

less show information to a lower strength than the Memory-Test group. Hence, the effort of memory retrieval would be greater for these two conditions than for Memory-Test.

**Table 1: Storage, retrieval, and access effort for each condition.**

<b>Memory Storage</b>	Memory-Test	>	Free-Access	=	Gray-Box
<b>Memory Retrieval</b>	Memory-Test	<	Free-Access	=	Gray-Box
<b>Perceptual-Motor Access</b>	Memory-Test	NA	Free-Access	<	Gray-Box

For perceptual-motor access, we expected the Memory-Test group to do little perceptual-motor access (hence, comparing the effort of perceptual-motor access for Memory-Test with the other groups is not appropriate, or NA). In contrast, perceptual-motor access for the Free-Access group should require approximately 500 msec. This estimate is based on CPM-GOMS models (Gray, John, & Atwood, 1993) of the time required to move the eye to a known location and to perceive a simple word or number. Extrapolating from the models and data presented by Gray & Boehm-Davis (in press), the time required to move a cursor to a gray box, click on the box, and perceive a simple word should be between 1,000 to 1,500 msec. Therefore, we expected that the access effort for the Free-Access condition would be less than that for the Gray-Box condition.

Using time as an estimate of effort leads us to make the following predictions. First, the Memory-Test condition should perform best as for them the critical show information is well-learned and easily retrieved from-the-head. The Free-Access condition should perform well but perhaps, not as well as the Memory-Test condition. The 500 msec estimate of the time needed to access information with an eye movement is an optimal estimate; the actual time may be slower. Hence, the time required to access perfect knowledge in-the-world may be competitive with the time required to retrieve imperfect knowledge in-the-head. To the extent that retrieving imperfect knowledge in-the-head is the least-effort solution, then we might expect the Free-Access condition to perform poorer than the Memory-Test condition. Finally, our estimate of 1,000 to 1,500 msec for the Gray-Box condition to retrieve knowledge in-the-world is within the range of time estimated as needed to retrieve knowledge in-the-head under normal circumstances. (For example, ACT-R's default time, or latency factor, to retrieve an item from declarative memory is 1,000 msec, see Anderson & Lebière, 1998.) Hence, for the Gray-Box condition, there should be many occasions in which imperfect memories are retrieved faster than the time needed to access knowledge in-the-world. Of the three conditions, least-effort considerations lead us to predict that the Gray-Box condition will be the worst. Therefore, although the margin of superiority is not clear, least-effort considerations of

performance lead us to predict Memory-Test > Free-Access > Gray-Box.

Experiment 2 was run to provide eye-tracking data on the Free-Access condition. With these data, we could compare the frequency and patterns with which Free-Access versus Gray-Box participants accessed the fields in the Show Information Window.

## EXPERIMENT 1

Experiment 1 had three conditions; all used a simulation of a commercial VCR built in Macintosh Common Lisp. All clicks on any button object in the simulation were time stamped to the nearest tick (16.67 msec) and saved to a log file along with a complete record of the information in the VCR's displays (e.g., mode, time, day-of-week, channel, and so on).

## Method

### Procedure

With minor differences described below, the procedure for all conditions was the same. The study began with the participant watching as the experimenter programmed the first trial of show 0. After the first trial the experimenter watched as the participant programmed show 0 to criterion. At that point, the experimenter left the room while the participant programmed shows 1 through 4. (As show 0 was an instruction and practice show, it is excluded from the analyses reported below.)

Each participant programmed shows 1-4 to the criterion of two successive correct trials. Each trial began with the participant pressing a START TRIAL button and ended with the participant pressing STOP TRIAL. At the end of each trial, the participant was given feedback as to how long the trial took and as to whether the show had been programmed correctly. If the show was not programmed correctly, the participant was provided feedback on the first error that the software found. The order in which errors were checked was: clock time, start time, end time, day of week, channel, and program record.

For all conditions and both experiments, each trial began with the VCR covered by a black box with the Show Information Window clearly visible and immediately below the VCR. In addition to fields containing the show's name, start time, end time, day-of-week, and channel, the Show Information Window also contained the START TRIAL button. Clicking on this button began the trial, changed START TRIAL to STOP TRIAL, and removed the black box that had covered the VCR.

For the Free-Access condition, the labels and fields of the Show Information Window were clearly visible throughout each trial. In contrast, for the Gray-Box condition, the labels in the Show Information Window were visible but gray boxes covered the fields. For example, to see the channel field the participant had to move the cursor to and click on the gray box covering that field. The value remained visible as long as the cursor remained in the field.

For the Memory-Test condition, clicking on the START button removed the Show Information Window and opened a memory test window. The memory test window required

the participant to select the show's start-hour, start-10min, start-min, end-hour, end-10min, end-min, day-of-week, and channel from a series of pop-up menus. After setting the show information the participant clicked the OKAY button. If the information were not set correctly, the participant iterated between the Show Information Window and Memory Test Window until the memory test was passed. A memory test was required before each trial of each of the four shows.

As the VCR was being programmed, we encouraged the Memory-Test group to retrieve show information from memory by discouraging the use of the Show Information Window. As per the Gray-Box condition, gray boxes covered the fields of the Show Information Window. In addition, moving the cursor out of the VCR window, caused the VCR to be covered by a black box. The black box stayed until the participant moved the cursor back to and clicked on the VCR window. Hence, for the Memory-Test condition, when a participant moved to and clicked on a gray box, the corresponding setting of the VCR (indeed, all settings of the VCR) was covered by the black box.

#### Participants

Seventy-two George Mason University undergraduates, 24 per condition, participated for course credit. Participants were assigned to conditions randomly in blocks of threes. The experiment took approximately 30 min. Participants were individually run.

#### Results and Discussion

Three dependent measures were analyzed: two measures of performance and one of process. The performance measures are trials-to-criterion and a measure of goal suspension. The process measure is an analysis of accesses of knowledge in-the-world.

Given that, in each condition, show information was readily available, we might have expected all participants in all conditions to spend a maximum of two trials to program each show. Hence, trials-to-criterion greater than two may be interpreted as reflecting a reliance on imperfect memory in lieu of accessing knowledge in-the-world.

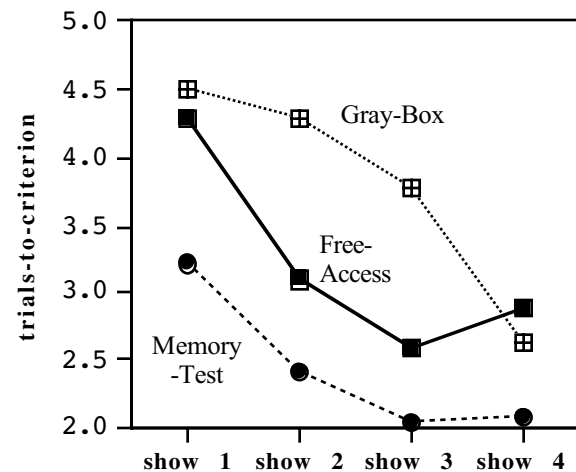
The measure of goal suspension is derived from Gray's (2000) goal-structure analysis of errors of performance. For the VCR simulation there are eight fields that must be set to correctly program the VCR; day-of-week, channel, start-hr, start-10min, start-min, end-hr, end-10min, end-min. Given the structure of the device, the measure of goal suspension is quite simple: Once a participant starts to change a setting, how often was it abandoned before being correctly completed? For example, if for show 2 the to-be-set channel was 21, and the current channel was 11, then if the participant began setting the channel but stopped before 21 (e.g., going off to set the day-of-week), then this is one goal suspension.

For goal suspensions, we examined only trials that were successfully programmed. In the context of a successfully programmed trial, goal suspensions are potential errors. They require that the participant detect that the setting is not complete and correct the setting before pressing the STOP TRIAL button. Note that accessing show

information during a setting was not considered goal suspension. For example, if a participant started programming the channel setting, interrupted his or herself to check the Show Information Window, and then resumed programming the channel – this would not be considered a goal suspension. We interpret goal suspensions as due to reliance, at least temporarily, on imperfect knowledge in-the-head rather than on perfect knowledge in-the-world. If participants compare the current setting of, e.g. channel, with the value of channel in the Show Information Window then they would not stop, but would continue programming until the current channel matched the goal channel.

Our process measure counts the number and the pattern of information accesses to the Show Information Window. For the Memory-Test and Gray-Box conditions, each click on a gray box was counted. The pattern of when information was accessed versus when the information was programmed was also recorded.

#### Performance Measures



**Figure 2: Trials-to-criterion for experiment 1. Participants were required to program each show to the criterion of two successive correct trials. Hence, for shows 3 and 4 the Memory-Test group is close to the minimum number of trials possible.**

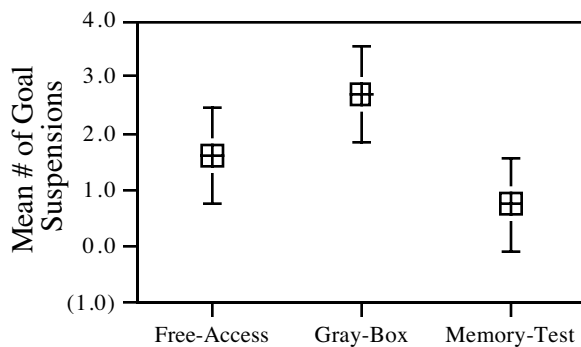
*Trials-to-Criterion.* A two-way analysis of variance (ANOVA) was conducted on the number of trials to reach the criterion of two successive correct shows. Condition (Free-Access, Gray-Box, Memory-Test) was a between-subjects factor and show (1-4) was within-subjects. The main effect of condition was significant,  $F(2, 69) = 4.48, p = .015$  ( $MSE = 10.04$ ), as was the main effect of show,  $F(3, 207) = 5.90, p = .0007$  ( $MSE = 5.05$ ). The interaction of condition by show was not significant ( $F < 1$ ) (see Figure 2).

Planned comparisons by condition yielded a significant difference between Gray-Box and Memory-Test ( $p = .0002$ ) as well as between Free-Access and Memory-Test ( $p = .037$ ). The difference between the Free-Access and Gray-Box condition was not significant. Despite the ready availability of knowledge in-the-world, both the Gray-Box and Free-Access group made more errors than did the

group that had show knowledge strongly encoded in-the-head.

**Goal Suspension.** The trials-to-criterion measure focused our attention on the number of trials that ended in error; that is, the number of trials that ended with a show being incorrectly programmed. The more shows that were incorrectly programmed the greater the trials-to-criterion. In contrast, for goal suspension we examine errors that were made, but latter detected and corrected, on trials that ended successfully.

Goal suspensions are a rarity. Examining patterns of goal suspensions requires that a vast quantity of correct data be collected and parsed. Across all three conditions of experiment 1, 36,877 mouse clicks were collected and time stamped on correct trials. These mouse clicks were parsed into 12,560 goals using the action-protocol analyzer developed by Fu (in press). Of these goals, there were 122 goal suspensions. For each group the mean number of goal suspensions per participant is shown in Figure 3.



**Figure 3: Mean goal suspensions per participant across the three conditions. Statistical Significance Bars (SSB) show the pairwise statistical significance between means.**

An overall ANOVA produced a marginally significant effect,  $F(2, 69) = 2.64, p = .078$ . The statistical significance bars (SSB) in Figure 3 are based on planned comparisons. If two SSB's look different (i.e., they do not overlap), the corresponding pairwise comparison is different (at the .05 level of significance adopted for this study). (For more information on SSB's see Schunn, 2000.) As indicated by Figure 3, the Gray-Box condition made significantly more goal suspensions than did the Memory-Test condition, but there were no significant differences between the other comparisons.

A  $\chi^2$  comparison that looked at whether or not each participant had a goal suspension, was significant ( $p = .05$ ). Fifty percent of the Free-Access participants made goal suspensions, 75% of the Gray-Box participants, and 42% of the Memory-Test participants.

**Discussion of Performance Measures.** The first two dependent measures, trials-to-criterion and goal suspensions, yield a consistent pattern. The Memory-Test condition is best, and the Gray-Box is worst with the Free-Access condition somewhere in the middle. These data present us with an interesting quandary. All groups had

access to all show information at all times, yet they made errors that kept them in the study longer than they needed to be.

The participants had to program each show until they got it correct twice in succession. Hence, the penalty for ending the trial in error was having to stay in the experiment longer. Participants in the Free-Access or Gray-Box groups could have matched the performance of the Memory-Test group by simply comparing their settings against the Show Information Window before clicking the STOP TRIAL button. Similarly, the penalty for a goal suspension was having to go back and complete the suspended goal at a later time at the risk of ending the trial in error. Participants in the Gray-Box condition could have easily double-checked show information before suspending their current goal. Both of these measures, trials-to-criterion and goal suspension, suggest that participants were relying on imperfect memory for show information rather than more reliable perceptual-motor access.

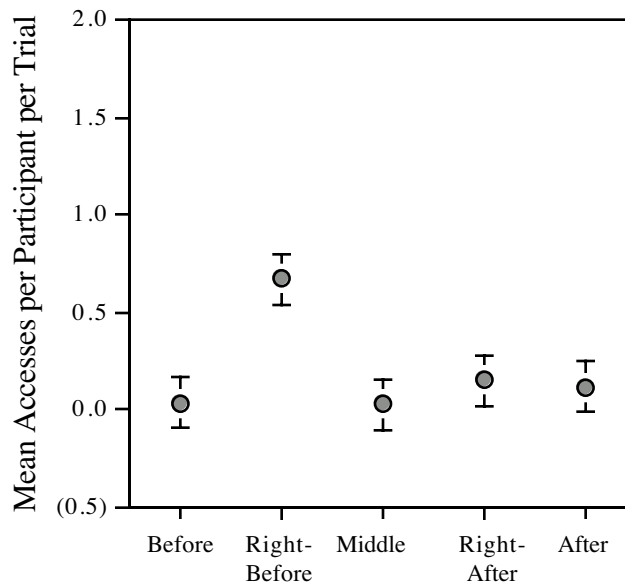
#### Process Measure: Accesses of Knowledge in-the-world

Our third measure can be used to address two questions. First is a construct validity issue (Gray & Salzman, 1998): Did the Memory-Test manipulation lead to the retrieval of show information from memory instead of accessing it from the display? The second examines what the patterns of information access reveal about the use of knowledge in-the-world.

**Construct validity.** Did the Memory-Test group rely on memory retrieval or on perceptual-motor access? Throughout shows 1 through 4, the 24 participants in the Gray-Box condition clicked on information fields 293 times over 223 correct trials for an average of 1.31 checks per show. In contrast, the 24 participants in the Memory-Test condition clicked on an information field 10 times during 205 correct trials for an average of 0.05 checks per show. This contrast suggests that the memory manipulation was successful and that the Memory-Test group almost exclusively relied on retrievals from memory as their source of show information.

**Patterns of Information Access.** Given that participants in the Gray-Box condition could access information in-the-world whenever they wanted it, can their patterns of information access provide any clue regarding why this group did not do as well as the Memory-Test group?

Figure 4 shows the mean number of information accesses per correct trial per participant for the Gray-Box condition. Each information access was categorized by when it occurred in relation to when the information was used. For example, if a participant accessed channel information but set something else before setting channel, this access was classified as *before*. If after accessing the channel information the participant's next act was to program the channel setting, this access was classified as *right-before*. Any interruption of a setting to access the information for that setting was classified as *middle*. If immediately after setting the channel the participant's next act was to access the channel information, this access was classified as *right-*



**Figure 4:** For the Gray-Box condition from experiment 1, the graph shows the mean accesses per participant per trial, SSBs, based on 24 participants, show the pairwise statistical significance between means.

*after*. Any later access of an information field was classified as *after*.

A within-subject ANOVA yielded significant between-category differences in when the Gray-Box group accessed show information,  $F(4, 92) = 15.36, p < .0001$  ( $MSE = 0.11$ ). The SSBs in Figure 4 are based on the Tukey Honestly Significant Difference (HSD) test. As shown by the SSBs, more accesses were performed right-before the information was needed than at any other time. There were no significant pairwise comparisons between any of the other access categories.

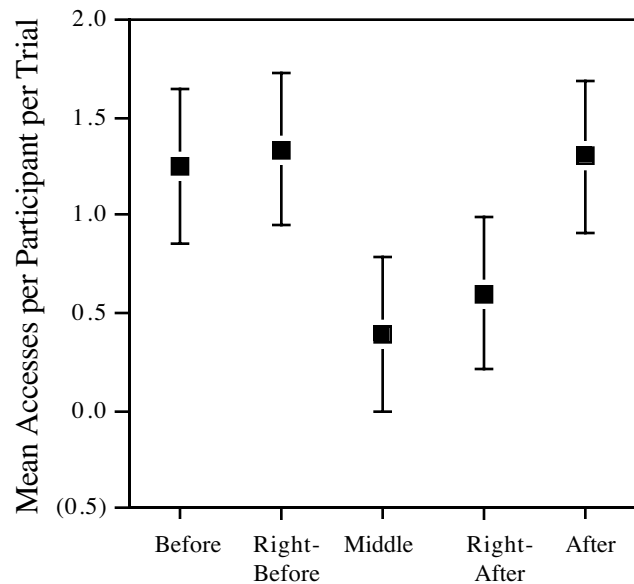
Apparently, the Gray-Box group will access knowledge in-the-world when the information they need to program the next setting is not available in-the-head. However, they are unlikely to access this information while they are programming a setting (middle), nor are they likely to compare a setting with information in the Show Information Window anytime after they have programmed it. If the Gray-Box group is making such comparisons, the comparisons must be based on access of imperfect knowledge in-the-head.

## EXPERIMENT 2

Experiment 1 was interesting but incomplete as it provided no information on how often or when the Free-Access condition accessed show information. To remedy this deficit we conducted experiment 2.

### Method

Experiment 2 had one condition that replicated the Free-Access condition with one main difference: participants were eye-tracked as they programmed the VCR. To facilitate eye-tracking, the size of the Show Information



**Figure 5:** For the Free-Access condition from experiment 2, the graph shows the mean accesses per participant per trial, SSBs, based on 8 participants, show the pairwise statistical significance between means.

Window was increased to increase the visual separation of each of the information fields.

### Participants

We report results from the first eight undergraduates who we could successfully eye track. All participants, whether or not they could be eye-tracked, received course credit for their participation. Because of the necessity to calibrate the eye-tracker on each participant, experiment 2 took approximately 45 min.

### Eye Tracking

Eye tracking was performed using an ASL 504 remote optics eye tracker. Head movements were tracked using a Flock-of-Birds™ magnetic head tracker. Eye data was sampled and saved to a log file 60 times per second (once every 16.67 msec).

Fixations were determined using the algorithm developed by Karsh and Breitenbach (1983). Basically, we say that a fixation occurs when at least six consecutive data points fall within a 3 x 4 pixel rectangle (where the definition of "consecutive" points is that they have to be less than 32 msec apart). Areas of interest were created around each information box. Consecutive fixations in the same area of interest were counted as a single access.

### Results

Figure 5 shows the mean number of information accesses per correct trial per participant for the Free-Access condition in experiment 2. A within-subject ANOVA showed the between category differences to be significant,  $F(4, 28) = 5.38, p = .002, MSE = 0.29$ . The SSBs in Figure 5 are based on the Tukey HSD test. The SSBs show no difference in number of accesses between the before, right-before, and after categories. However, each of these three

categories significantly differs from the middle category and is marginally different from the right-after category. There are no differences between the middle and right-after categories.

### Discussion

The pattern of information access across the Gray-Box condition of experiment 1 and Free-Access condition of experiment 2 has some interesting similarities. Both groups are more likely to access information right-before they need it instead of when they are using it (i.e., middle) or right-after. Apparently, both groups are so complacent in their ability to retrieve the correct information from memory that they are unwilling to pay the perceptual-motor effort needed to verify that the current setting is, indeed, the target setting.

The differences in patterns of information access are as revealing as the similarities. First, the lower the perceptual-motor effort required to access information, the more frequent the accesses. Over all categories the Experiment 2 Free-Access group is 4.3 times more likely to access information than is the Gray-Box group. However, although the number of accesses decreases in all categories between the Free-Access and Gray-Box conditions, the one category that is partially protected is the right-before category. The Gray-Box group appears to devote a higher proportion of its information accesses to the right-before category than does the Free-Access group. The higher number of accesses before and after suggests that the Free-Access group does more advance storage than the Gray-Box group and more comparing of the VCR settings to the show information.

### SUMMARY & DISCUSSION

The perceptual-motor and memory effort manipulated in these studies are of the same order of magnitude as the effort paid by the typical user of direct-manipulation interfaces. The effort associated with the Memory-Test condition is similar to that paid by the power user who has memorized critical or frequent shortcut keys. The Free-Access condition is similar in perceptual-motor effort to many situations in which information that is available in one open window is required by a program running in another open window. Finally, participants in the Gray-Box condition spent an effort equivalent to that required by users who must move to and click on a partially covered window to bring the information it contains to the foreground. Indeed, given the pedestrian nature of the manipulations, it is interesting and important that the three conditions produced the pattern of results that they did.

The most striking aspect of the between group differences in performance is that all were avoidable. All performance differences can be traced to differences in willingness to either memorize or access show information. For each trial the Memory-Test group had quick and reliable access to show information in memory. The other groups made more undiscovered errors that resulted in more trials-to-criterion. Apparently, verification is lower effort – and hence more likely – if based on knowledge in-the-head rather than accessing knowledge in-the-world.

Although the level of perceptual-motor and memory effort manipulated in these studies was representative of that encountered in many human-computer interactions, this effort is much lower than that involved in many other human-computer interactions. For example, for the typical user, searching the web for information may involve recalling or locating a slightly different login and password for each website visited. Similarly, unlike the situations we studied, visually busy web pages impose a substantial search effort on accessing knowledge in-the-world.

Extrapolating from the findings presented here, we would expect the prototypical visually busy website to present a situation in which least-effort considerations would lead users to tradeoff perfect knowledge in-the-world for imperfect knowledge in-the-head. Since for many sites the user's knowledge in-the-head would be based only on recent interactions (i.e., the last 10-30 sec), this may produce a tunnel-vision like effect in which users concentrate on the little that they know and fail to mine the riches awaiting at unaccessed locations on the screen.

### CONCLUSIONS

The results support the predictions of the rational analysis framework. As the effort of obtaining knowledge in-the-world increases, performance deteriorates. This prediction holds even though the absolute effort of perceptual-motor access seems trivial, whereas the future effort required by an undiscovered error is more time (and trials) in the experiment.

Some readers may object, as did one reviewer, that if the VCR had been designed differently then the observed failures to access knowledge in-the-world would not have occurred. However, this observation is not an objection to the current research but, rather, is precisely the point of the current research. The goal of the research is to understand how interactive behavior emerges from the constraints and opportunities provided by the interaction of embodied cognition (Kieras & Meyer, 1997) with the task being performed and the interface designed to perform the task. (Embodied cognition includes the perceptual-motor system as well as core cognitive areas such as memory and decision-making.) The difficulty lies in understanding how small changes in interface design interact with embodied cognition to produce interactive behavior. Hence, the proper focus of our study is not the interface per se, but the human. What is important is not the trivial observation that different interface designs produce different patterns of interactive behavior, but understanding the interaction of design with embodied cognition that leads to these different patterns.

Paying the effort to store knowledge in-the-head may be expensive, but it produced the best performance. During performance the Memory-Test participants could quickly and reliably retrieve show information with no perceptual-motor effort. In contrast, participants in the Gray-Box condition preferred to rely on imperfect memory rather than paying the higher effort associated with accessing perfect knowledge in-the-world. Although the Free-Access condition came close to matching the Memory-Test

condition in performance apparently, in some cases, even the effort of an eye movement was too great of a price to pay to access perfect knowledge in-the-world.

That least-effort decisions guide interactive behavior introduces a new level to interface design where milliseconds matter (see also, Gray & Boehm-Davis, in press). Although these effects are small and subtle, they are at work whenever an interface offers the user more than one way of performing an action or more than one source of information. The challenge for interface designers is not to react by limiting the user's alternatives, but to design interfaces where least-effort decisions lead to effective and error free performance.

#### ACKNOWLEDGEMENTS

We thank Jeni Paluska, Michael J. Schoelles, Wolfgang Schoeppek, and Christian Schunn for commenting on earlier versions of this paper. We also thank John Sellers and Melanie Diez for help in collecting data. The work reported was supported by a grant from the National Science Foundation (IRI-9618833) as well as by the Air Force Office of Scientific Research AFOSR#F49620-97-1-0353.

#### REFERENCES

1. Altmann, E. M., & Gray, W. D. (1999). Serial attention as strategic memory. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 25-30). Hillsdale, NJ: Erlbaum.
2. Altmann, E. M., & Gray, W. D. (2000a). An integrated model of set shifting and maintenance. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modeling* (pp. 17-24). Veenendal, NL: Universal Press.
3. Altmann, E. M., & Gray, W. D. (2000b). Managing attention by preparing to forget, *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
4. Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
5. Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(3), 471-517.
6. Anderson, J. R., & Lebière, C. (Eds.). (1998). *Atomic components of thought*. Hillsdale, NJ: Erlbaum.
7. Ehret, B. D. (1999). *Learning where to look: The acquisition of knowledge in display-based interaction*. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.
8. Frohlich, D. M. (1997). Direct manipulation and other lessons. In M. Helander & T. K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (Second ed., pp. 463-488). New York: Elsevier.
9. Fu, W.-t. (in press). ACT-PRO: Action protocol tracer -- a tool for analyzing simple, rule-based tasks. *Behavior Research Methods, Instruments, & Computers*.
10. Fu, W.-t., & Gray, W. D. (1999). Redirecting direct-manipulation, or, what happens when the goal is in front of you but the interface says to turn left?, *Proceedings of the CHI 99 Extended Abstracts* (pp. 226-227). New York: ACM Press.
11. Fu, W.-t., & Gray, W. D. (2000). Memory versus Perceptual-Motor Tradeoffs in a Blocks World Task, *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society* (pp. 154-159). Hillsdale, NJ: Erlbaum.
12. Gray, W. D. (2000). The nature and processing of errors in interactive behavior. *Cognitive Science*, 24(2), 205-248.
13. Gray, W. D., & Boehm-Davis, D. A. (in press). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*.
14. Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, 8(3), 237-309.
15. Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13(3), 203-261.
16. Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction*, 1(4), 311-338.
17. Karsh, R., & Breitenbach, F. W. (1983). Looking at looking: The amorphous fixation measure. In R. Groner & C. Menz & D. F. Fisher & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 53-64). Hillsdale, NJ: Erlbaum.
18. Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391-438.
19. Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.
20. Norman, D. A. (1989). *The design of everyday things*. New York: DoubleDay.
21. Schunn, C. D. (2000). Statistical significance bars (SSB): A way to make graphs more interpretable. *Manuscript submitted for publication*.
22. Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behaviour and Information Technology*, 1(3), 237-256.
23. Simon, H. A. (1974). How big is a chunk? *Science*, 183(4124), 482-488.