



## **Self-monitoring during exploration of an interactive device**

CAROL-INA TRUDEL<sup>†</sup> & STEPHEN J. PAYNE

*University of Wales—College of Cardiff, School of Psychology, P.O. Box 901, Cardiff, CF1 3YG, Wales. email: carol-ina.trudel@inria.fr*

*(Received 19 July 1995 and accepted in revised form 20 August 1996)*

We report four experiments in which we manipulated the conditions under which subjects explored an unfamiliar interactive device without the benefit of assistance or instruction. Our aim was to increase the amount of knowledge subjects acquired about the device and to influence how efficient they were at later applying the operational procedures they discovered. The interventions were to interrupt subjects' exploration at regular intervals and ask them to verbally report on what they had learned so far (Experiment 1) or to state their future intention (Experiment 2). Both manipulations yielded significant benefits, when compared to subjects who explored the same device without such interruptions. In Experiment 3 there were four conditions in which interrupted subjects reported on topics of different levels of relevance to the task. This experiment showed that it is the relevance of the verbalised content as opposed to an "incubation" period which affects performance in a beneficial manner. Finally we investigated whether it was necessary to employ external prompts to obtain these effects (Experiment 4). It was found that subjects were in fact able to effectively interrupt their work in order to review their progress.

© 1996 Academic Press Limited

### **1. General introduction**

#### **1.1. EXPLORATORY LEARNING**

Trudel and Payne (1995) reported experiments which examined the nature of exploratory learning (of a simulated digital watch) and some factors that influence its success. Among the manipulations that we studied, the most striking effect was obtained by imposing a keystroke limit on subjects exploring the device. This resulted in substantially more successful learning as compared to subjects who explored without any constraints, despite the fact that the keystroke limit group spent less total time interacting with the device. We interpreted this result within a general framework of learning through reflection. Imposing a keystroke limit made each interaction with the device a more scarce and valuable resource, and thereby encouraged subjects to pay more attention to and think harder about the results of any interaction.

In the same study we also used verbal and behavioural protocols to classify the strategies—i.e. types of discoveries, hypothesis testing, goal abandonment, repetition, forgetfulness, reviewing progress, attention to feedback—that were employed

<sup>†</sup> Present address: INRIA, Projet Psychologie Ergonomique, Domaine de Voluceau, Rocquencourt, B.P. 105, 78153 Le Chesnay, Cedex, France.

by subjects as they explored the device. A *post hoc* division of subjects into Good Learners and Poor Learners highlighted the main differentiating factor between Good and Poor explorers: the latter were carrying out significantly more moves which impeded their progress toward a good understanding of the device. There was not much evidence that good learners adopted a greater number of "good" strategies; rather they seemed able to avoid poor strategies, such as repeating inappropriate actions. We reasoned that this may be due to the good learners being more inclined to review their past progress, both their accomplishments and their failures. In the current study this tendency is investigated in greater depth.

Whether or not reviewing yields benefits, we certainly did not find a natural inclination in most of our subjects to do so. Perhaps the "display based" characteristic of the device (e.g. Payne, 1991), which permits rapid and low-effort input, does not encourage reflection. It may be that the current generation of computer users (even if they are only occasional users) have learned that it is highly unlikely that their actions will bring about irrecoverable damage to the system. In any event, only 8.3% of all the strategies carried out by our subjects were explicit, verbalized reviews of what had been learned so far. This may be because subjects failed to notice that this strategy could be beneficial, were unmotivated to expend the cognitive effort to do so, or simply because they had no experience in reflecting on their actions in this manner.

A great body of educational research supports the view that the ability to monitor one's performance and to evaluate progress is a necessary component of proficient learning (e.g. Schunk, 1983; Ascher, 1984; Schunk & Rice, 1987; Lavoie & Good, 1988; Pintrich & deGroot, 1990; Pressley & Ghatala, 1990; Lan, Bradley & Parr, 1993). Similarly, research on verbalization reviewed by Ericsson and Simon (1993) has repeatedly demonstrated that what they call "level 3" verbalizations (i.e. when subjects are required to assess and generate explanations for their actions) lead to improved learning, problem solving and recall (e.g. Wilder & Harvey, 1971; Berry, 1983; Ballstaedt & Mandl, 1984; Berry & Broadbent, 1984; Stinessen, 1985; Ahlum-Heath & DiVesta, 1986). These beneficial effects of self-monitoring are almost certainly not due to the verbalizations *per se*. (Gagne & Smith, 1962; Chi, Bassok, Lewis, Reimann & Glaser, 1989; Ericsson & Simon, 1993; Veenman, Elshout & Groen, 1993). Wilder & Harvey (1971) for example produced the same beneficial changes in performance by asking subjects to generate explanations for their actions covertly.

Some further studies additionally demonstrate that adequate self-monitoring is not always spontaneous, and that prompts to monitor can improve learning outcomes. Brown, Campione and Day (1981), reporting on their approaches to train students to learn from texts, showed that students' ability to provide an adequate summary of what has been read is an important factor in overall understanding. They found however that poor learners found this to be a difficult task and that they needed training in order to be able to carry it out effectively. Gaskins (1988), describing a program for increasing underachievers' abilities, also stressed that poor learners need to be specifically taught, because they are unlikely to discover for themselves how to plan, self-monitor and evaluate their learning progress. Heins, Lloyd and Hallahan (1986) also demonstrated how the academic performance of

graduate students could be improved by self-monitoring but that the improvement was more evident in a condition where subjects were prompted, or cued, to do so. Chi, DeLeeuw, Chiu and LaVancher (1994) reported that subjects prompted to explain after reading each sentence of an expository text demonstrated enhanced understanding of the domain (i.e. human circulatory system) which they were learning about.

In a similar line, the studies by Veenman, Elshout and Bierman (1989, 1994) are perhaps the most closely related to those reported in this paper. These authors demonstrated that the metacognitive skills which are characteristic of a proficient working method include a systematic approach, review and elaboration of what is being learned. When subjects were regularly prompted to employ these skills before being allowed to carry out experiments with an "Electricity Lab" simulator their qualitative knowledge of the simulator was significantly better than that of subjects who engaged in unguided exploration of the device.

These studies adequately demonstrate that the more people reflect on their learning experiences, the more they learn. They also highlight the fact that in many situations people need to be prompted to engage in appropriate reflection. Nevertheless some questions remain.

In reviewing the literature on reflection, we have collapsed a wide variety of different senses of "reflection". On the basis of our earlier studies (Trudel & Payne, 1995) we conjecture that in the context of exploration of an interactive device, a minimal form of reflection, namely the bare reviewing and restatement of earlier learning experiences, even in the absence of any interpretation of these experiences, will benefit learners.

There is yet no direct evidence for this hypothesis in the literature. Although the empirical studies reviewed cover a wide range of learning situations, none directly related to exploratory learning of interactive devices, which does indeed have some quite unusual properties. For example, exploratory learners of interactive devices rely entirely on self-guided interactions—there is no instructional text whatsoever, whereas much of the literature on learning-by-reflection relates strongly to learning-from-text. In addition, because exploratory learners learn "by doing" in the absence of prior guidance, it is supposed that they will typically acquire declarative and procedural knowledge about the device simultaneously, in contradistinction to the classical declarative-then-procedural trajectory of most cognitive skills (e.g. Anderson, 1983). Most of the literature on learning-by-reflection relates either to the initial acquisition of declarative knowledge about a domain, or to the acquisition of procedural skills from some pre-existing declarative basis.

Furthermore, there is some empirical evidence in the problem solving literature which actually argues against our hypothesis. De Jong and Simons (1988) had subjects think aloud while reading texts describing how to operate a keyboard and edit online files. In an introductory session, subjects received practice in speaking aloud and were given a list of examples of thought expressions related to self-regulation. In the experimental session one group of subjects read the texts while freely thinking aloud while the other group were required to verbalize only thoughts related to self-monitoring activity. In addition, these same subjects read one text while verbalizing constantly and read another text while verbalizing only at

prompted intervals. De Jong and Simons found that the different verbalizations conditions yielded no differences in learning between the groups.

Schooler, Ohlsson and Brooks (1993) found that both retrospective and concurrent verbalization actually interfered with the successful solution of standard insight problems (e.g. Metcalfe & Wiebe, 1987). Specifically, subjects were less successful when interrupted to articulate their problem-solving strategies than subjects interrupted to perform an unrelated activity or subjects whose work was not interrupted. Their explanation is that verbalization disrupts the non-reportable processes that are critical to achieving insight solutions.

This finding leads us to question the effect that verbalization would have on the exploratory learning of an interactive device. This type of learning is characterized by action-oriented problem solving (i.e. pressing keys and observing the outcome) as well as by higher order cognitive activities, such as insight. In earlier work (Trudel & Payne, 1995) we transcribed and classified the verbal protocols of subjects engaged in the exploration of the same device used in the experiments reported here—though this was not done specifically for the purpose of identifying instances of insight. An excerpts of these protocols (presented below—italics are authors' explanation and protocols in bold highlight what we view as a moment of insight) nevertheless supports our view that learning by exploration can lead to discovery-by-insight.

**Subject A:** Um, I get this timer and it won't stay on. (*She is referring to the label CHRONO which only briefly appears before the watch face for that mode appears.*) It keeps going back to nought (*i.e. to the Chrono watch face*). Oh, I've started the stopwatch. I think. (*She has.*) **Oh, I see.** So back to mode again. (*She seems to have understood that the label appears before the watch face in that mode.*)

**Subject B:** I don't quite understand but obviously one's sort of going faster than the other, so it must be something to do with the laps because I don't understand that bit. So you start it and then you stop it and then you . . . **Oh right.** I think maybe the top one is showing you how many laps you're doing. (*Activating the stopwatch and then the lap/reset button records the different laps on the top screen while the bottom screen keeps a running total.*)

**Subject C:** I don't know what the time's doing now. Press lap again and it goes back . . . hold on . . . they're on different times now (*he's activated the lap function of the stopwatch so the top screen is recording those*). Um, that's got me confused. **Ah! I know what it is now.** Reset that. (*He starts over again to verify this discovery.*)

Schooler *et al.* (1993) also found that verbalisation did *not* disrupt the solution of non-insight problems (where a solution can be arrived at by working through a series of incremental steps) but nor did it produce beneficial effects on problem solving. This is not surprising given, as they point out, that concurrent verbalizations have been shown to not disrupt performance (Ericsson & Simon, 1980, 1993). It is nevertheless worth speculating what would happen in a situation where one must generate retrospective protocols in a situation which involves both insight and noninsight problem solving, as is typically the case in exploratory learning.

In summary then, the weight of the evidence from our earlier study and from the education literature summarized above leads us to hypothesize that verbalizations will enhance learning. Nevertheless, the conflicting evidence of de Jong and Simons (1988) and of Schooler *et al.*'s (1993) study reduces the certainty of this hypothesis.

The effects of self-monitoring on the exploratory learning of an interactive device remain to be empirically tested.

### 1.2. THE CURRENT EXPERIMENTS

This study looks at the effect of prompting subjects to verbally report on their learning progress while they explore an unfamiliar interactive device (described in Method section). Subjects were instructed at the start of the experiment that they would be interrupted, at regular intervals, to produce verbal reviews. We expect that these prompted reviews will result in greater learning when compared to the learning of subjects in a control condition where they will explore freely and not be asked to think aloud.

## 2. General method

In the interest of parsimony the shared general method for Experiments 1 to 4 are described here. Any element particular to one experiment will be described in that experiment's method section.

### 2.1. SUBJECTS

Subjects were undergraduates from the University of Wales, Cardiff. Only subjects who had never programmed a digital watch before were allowed to participate. There was a mix of subjects who participated in order to fulfil a course requirement and subjects who were paid £3 for participating.

### 2.3. MATERIALS

#### 2.3.1. Target Device

The experiment was run on an Apple Macintosh II vi. The watch was created in Hypercard version 2.0 (see Figure 1). This computer-simulated digital watch was based on the commercially available Timex Triathlon Series digital watch. The only button present on the actual watch that was not reproduced on the computer-simulated one was the light button, for obvious reasons. All other buttons and functions were in the same position and had the same name for both watches. The computer was programmed to keep a record of mouse button presses and their location on the computer watch display. As can be seen from the ideal answers to the questionnaire used in the experiments (see Appendix 1), the watch had four modes, which could be scrolled through by repeatedly pressing the button labelled "Mode". The other three buttons served a variety of functions depending on the Mode. The questionnaire answers (Appendix 1) list these functions.

We programmed the device so that when the experiment started the normal display mode (with current time and date) appeared on the screen. It is important to note that whenever the mode button was pressed a "label" announcing the *next* mode flashed briefly on the screen (as it does on the real watch) and automatically disappeared again so that the watch face for that mode could appear. For example, pressing the mode button from the normal display would make the label "Chrono" appear—since the chronograph (i.e. stopwatch) was the next mode one

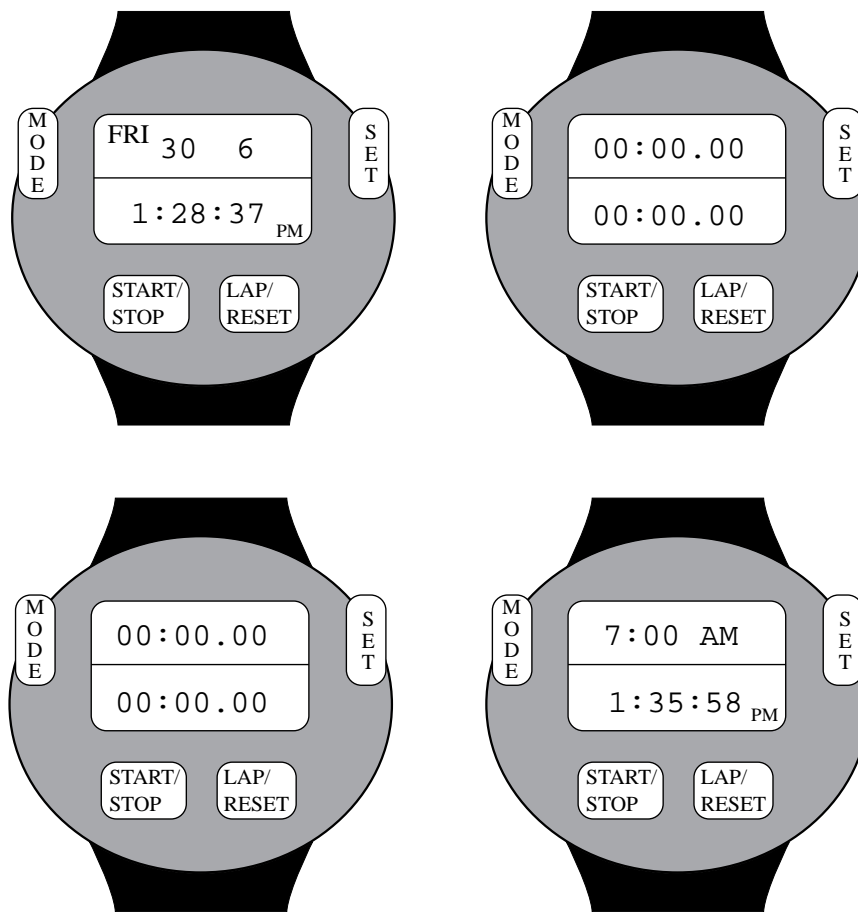


FIGURE 1. Device used in all experiments. The computer-simulated digital watch consisted of four separate modes: Normal Display (with date & time—top left), Chronograph (i.e. stopwatch—top right), Timer (bottom left) and Alarm (bottom right).

could call forth. Pressing the mode button a second time would bring up the label “timer”, and a third time would bring up the label “alarm”. From the alarm mode however, one went directly back to the normal display, without a label preceding this change. We mention this in order to show that although at a first glance the Chrono and Timer modes look similar, the subject could rely on the label to be certain of which mode she/he was in. Another distinction is that the Chrono allowed one to time min/s/ms while the Timer mode allowed one to set h/min/s. Before any of the function buttons were pressed the watch faces for each mode appear as they are shown in Figure 1, except for the normal display which automatically displayed the current date and time, and for the bottom screen of the alarm mode which also displayed the current time.

### 2.3.2. *Tape Recorder*

A Panasonic tape recorder was used to record the verbal protocols of subjects in all the Reflection conditions.

### 2.3.3. *Questionnaire*

A simple pencil and paper questionnaire asked subjects to write down all they could recall about the functions of each of the four buttons of the watch (see Appendix 1).

### 2.3.4. *Test*

A separate “test device” (i.e. version) of the watch was programmed for each of six questions (Appendix 3). Each test device was programmed so that only the correct sequence of keystrokes had the usual effect. Any erroneous or out of sequence keystroke had no effect at all, except to be recorded in the data file. Number of erroneous keystroke is thus a good index of the degree to which a particular method has been learned. As an example (which was not given to subjects) question number 1 required subjects to change the time on the watch. The correct sequence was: Set button, Start/Stop button to move the select box from seconds to minutes area, Lap/Reset button to advance the minutes by 10 units, etc. Any deviation from this sequence was recorded as an error. There was a time limit for each of the six questions, so as to minimize the opportunity for new exploratory learning during the test. To anticipate the results, no subject reached that time limit on any question.

## 2.4. DESIGN

For Experiments 1, 2 and 4 subjects were randomly assigned to one of two conditions; Prompted Verbalizations or Unstructured Exploration.

In the Unstructured Exploration conditions, subjects explored the watch for 20 min without any assistance, supporting documentation or constraint. Subjects in the Prompted Verbalization conditions also explored the device for 20 min, without aid or constraint. However, they were required to stop after every 2 min of exploration to verbalize on their future actions or on their progress (the precise nature of requested verbalization is described in the design sections of Experiments 1, 2 and 4).

Experiment 3 was slightly different in that there was no Unstructured Exploration group. Rather, there were four groups of subjects prompted to verbalize, after every 2 min of exploration, on topics of varying task relevance (described in Design section of that experiment).

We chose 2 min intervals because it was felt that this would give subjects enough time in which to have discovered something new about the device and was often enough to allow us to gain a representative view of a subject’s exploration approach.

All of the verbal protocols generated by subjects in these four experiments were transcribed, though no in-depth analyses of the content of the verbal protocols were carried out. The instructions given to subjects in the four experiments are presented in Appendix 2.

## 2.5. PROCEDURE

Each subject was tested individually in a laboratory. Before starting the experiment subjects were asked whether they had used a mouse before. Those who had not were shown how to do so, and allowed to practise moving the mouse and observe the input cursor moving around the screen. Each subject was also shown which four buttons on the watchface were interactive (i.e. mode, start/stop, lap/reset, set) and told that these were the only buttons they should use, i.e. they were told that clicking directly on the watchface or on any other part of the screen would have no effect. Following this, the subject read the instructions for the experimental condition to which they had been assigned. Once they stated that they understood the instructions and that they had no questions the experiment was started.

There were three ordered phases to the experiment: Exploration, Questionnaire, and a Test. Each subject was told, before the Exploration phase, that they would subsequently be given an on-line exercise to test the knowledge they had acquired about the watch functions. The experimenter was present in the laboratory during the experiment in order to hand out materials, monitor the functioning of the equipment and, for the Reflection groups, to time the interruptions.

Subjects in all conditions were told that their task was to discover how to operate every function of the watch. They were also informed that afterward, they would be given a test which would draw on their knowledge of the different device functions.

The additional instructions given to subjects in the reflection conditions are described in the Design sections of each of the three experiments.

Following Exploration subjects were given the questionnaire, which they had not been told about at the beginning of the experiment. They were not allowed to refer to the watch while completing the questionnaire. There was no time limit; subjects simply handed their answer sheet to the experimenter after they felt they had written all they could recall. Following this, subjects had a 3 min break.

After the break, subjects received instructions for the test. They were told that there was a time limit for each question and that they would be told what it was before they saw the question. They were also told to value accuracy rather than speed (taking into account the time limit of course—but, in fact, no subject ever exceeded the time limit for any of the test questions. The time limit was determined by doubling the amount of time it took an “expert” to carry out that test question in a pilot study.) After each question was completed the Experimenter re-initialized the watch and presented the next question. The six test questions were presented one at a time, in a fixed order.

## 3. Experiment 1: prompted reviews

### 3.1. METHOD

#### 3.1.1. Subjects

Subjects were 16 undergraduates (11 females, five males). With regard to previous computer experience, three subjects had none, five subjects had 6 months or less, six subjects had 2 years or less, and two subjects had 5 years or less. All subjects in the latter three categories listed word processing, and of these subjects, five also listed



games, six listed statistics and two listed graphics as packages they had used in the past. No subject had any programming experience.

### 3.1.2. Design

We compared the performance of Unstructured Exploration (UE) subjects to that of Prompted Review subjects (PR). In the latter condition the experimenter asked subjects to stop working, every 2 min and state, out loud, what they had learned so far (i.e. in the preceding 2 min). Note that in all prompted conditions the first prompt was signalled after the first 2 min of subjects' exploration.

## 3.2. RESULTS

It may be of interest to note that subjects in the Prompted Review condition spoke for an average of 3 min and 36s (S.D. 1 min and 17 s). The time they spent speaking took up part of the 20 min allotted for exploration of the device. The mean number of speaking episodes (i.e. how often they were interrupted to verbalize, which can vary depending on the length of time they speak) was 6.

Data analysis: *t*-tests were computed to compare the PR and the UE groups in each of the analyses described below.

### 3.2.1. Number of key presses: exploration

There was a significant difference with regard to how many key presses subjects employed while they explored the device. On average, the PR subjects made 366.9 (S.D. 59.8) key presses while the UE subjects made 483.9 (S.D. 135.2), ( $t = 2.24$ ,  $df = 14$ ,  $p < 0.02$ , one tailed).

### 3.2.2. Post-exploration questionnaire scores

The questionnaire responses were scored by assigning 1 point to each of the 16 possible functions that could be listed under the four buttons (see Appendix 2; any understandable paraphrases of the functions were accepted). There was no significant difference between the PR or UE groups with regard to the amount of device functions recalled. On average, the UE group listed 9.4 (S.D. 2.6) items while the PR group listed 10.6 (S.D. 1.5), ( $t = 1.19$ ,  $df = 14$ ,  $p = 0.13$ , one tailed).

### 3.2.3. Number of errors at test

Figure 2 illustrates the average number of errors made by PR subjects (8.7, S.D. 4.8) and UE subjects (45, S.D. 13.8). Indeed, subjects who were interrupted during exploration and asked to state what they had learned in the preceding 2 min made significantly fewer errors at Test than the group who explored the device without interruptions ( $t = 7.00$ ,  $df = 14$ ,  $p < 0.0001$ , one tailed).

## 3.3. DISCUSSION

When tested for recall subjects in both conditions listed slightly more than half of the 16 device functions asked for in the questionnaire. There was no significant difference between the two groups, perhaps simply because the questionnaire test was insufficiently sensitive. However, where the procedural test is concerned, the

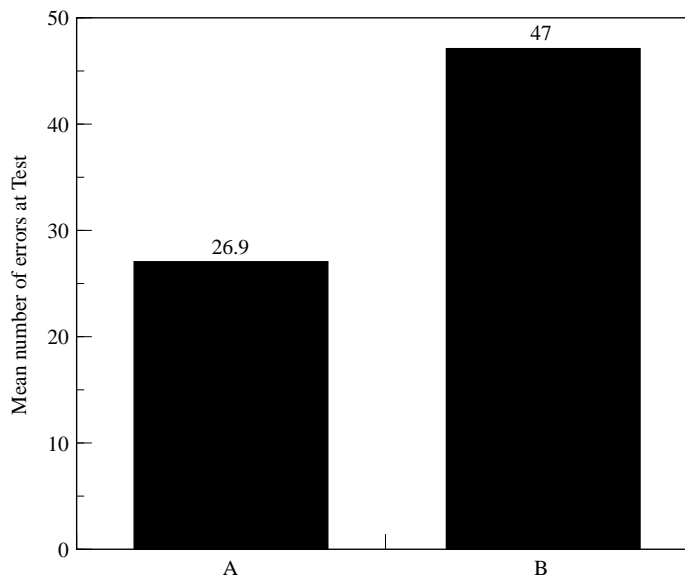


FIGURE 2. Mean number of errors at test. A: Prompted Reviews, B: Unstructured Exploration.

subjects who were asked to review their learning outperformed the group who explored the device without any constraints.

So while both groups performed well when their declarative knowledge was tested, it would appear that the unstructured exploration group were not able to transform or extend this knowledge into procedural knowledge that could be applied during the online test. A similar distinction between subjects' acquisition of these two types of knowledge has previously been reported by Chi *et al.* (1989). They found no differences in the ability of poor and good students to write down the declarative definitions and principles introduced in a physics test. However, the poor students were not successful in translating these into specific inference rules which could be used to solve problems.

#### 4. Experiment 2: prompted intentions

The results of Experiment 1 led us to speculate whether more deliberate planning ahead, as opposed to reviewing what has already been learned, would also yield beneficial effects with regard to subjects' learning. We predicted that subjects who are prompted to articulate plans would be more successful than the group who will explore the device in an unconstrained manner because deliberate planning will force subjects to construct and revise goals throughout the exploration session, thus acting as a form of self-monitoring. Subjects in the unstructured condition will be more *ad hoc*, we suppose, choosing and abandoning goals opportunistically.

#### 4.1. METHOD

##### 4.1.1. Subjects

Subjects were 24 undergraduates (19 females, five males). With regard to previous computer experience, nine subjects had none, two subjects had 6 months or less, four subjects had 1 year or less, six had 3 years or less, and three subjects had 5 years or less. All subjects in the latter four categories listed word processing, and of these subjects, three also listed games, four listed statistics and one listed graphics as packages they had used in the past. Only one subject also had programming experience.

##### 4.1.2. Design

We compared the performance of Unstructured Exploration (UE) subjects to that of Prompted Intentions (PI) subjects. In the latter condition the experimenter asked subjects to stop working, every 2 min and state, aloud, what they planned to work on in the *following* 2 min.

#### 4.2. RESULTS

Subjects in the Prompted Intention condition spoke for an average of 2 min and 5 s (S.D. 42 s). The time they spent speaking was deducted from the 20 min allotted for exploration of the device. The mean number of speaking episodes was 7.

Data analysis: *t*-tests were computed to compare the PI and the UE groups in each of the analyses described below.

##### 4.2.1. Number of key presses: exploration

There was a significant difference with regard to how many key presses subjects employed while they explored the device. On average, the PI subjects made 488.4 (S.D. 100.9) key presses while the UE subjects made 669.7 (S.D. 202.3), ( $t = 2.77$ ,  $df = 22$ ,  $p < 0.005$ , one tailed).

##### 4.2.2. Post-exploration questionnaire

There was no significant difference with regard to the average number of items subjects remembered. On average the PI groups recalled 9.2 (S.D. 1.6) items while the UE group recalled 8.5 (S.D. 2.2), ( $t = 0.956$ ,  $df = 22$ ,  $p = 0.17$ , one tailed).

##### 4.2.3. Number of errors at test

Figure 3 illustrates the average number of errors made by PI subjects (26.9, S.D. 16.1) vs. those of the UE subjects (47, S.D. 26.6). Subjects who were interrupted during exploration and asked to state what they intended to work on in the following 2 min made significantly fewer errors at Test than the group who explored the device without interruptions ( $t = 2.23$ ,  $df = 22$ ,  $p < 0.05$ , one tailed).

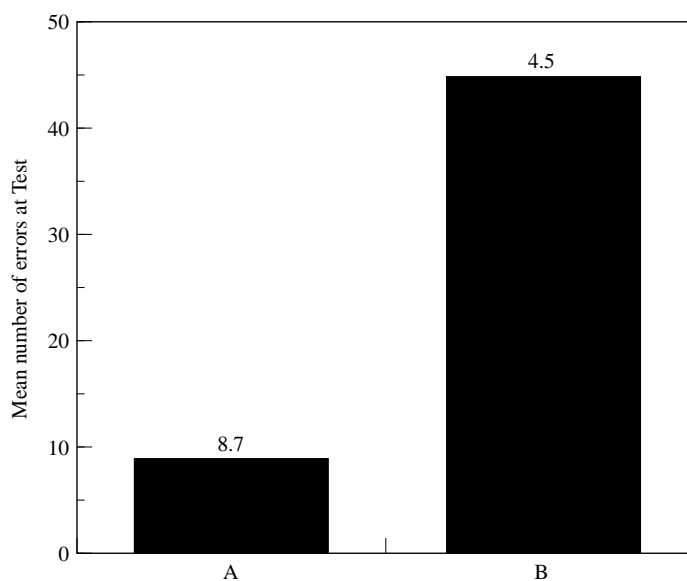


FIGURE 3. Mean number of errors at test. A: Prompted Intentions, B: Unstructured Exploration.

#### 4.3 DISCUSSION

There was no difference in the number of items recalled by subjects in both conditions, as tested by the questionnaire. However, once again we did find a difference in how efficiently subjects were able to perform the online test; subjects who had stated their future intentions throughout the experiment made fewer errors than the group who explored without interruptions.

It is interesting to compare the performance of the two Prompted groups from Experiments 1 and 2. Table 1 shows that the type of protocols being generated were very much in keeping with the instructions given to subjects, i.e. PR subjects focused on reviewing their progress while PI subjects explained what they would be attempting in the future. While no in-depth analyses of the protocols were made these excerpts are representative of the entirety of subjects' protocols in each condition.

Together with the difference in content, there was a notable difference with regard to the average length of speaking time for subjects in the PR condition of Experiment 1 (3 min and 36s) as compared to subjects in the PI condition of Experiment 2 (2 min and 5s). [The average number of speaking episodes for subjects was virtually identical in both conditions; PR = 6 episodes (S.D. 0.54) whereas PI = 7 episodes (S.D. 0.49)]. This difference in length of speaking time is perhaps not surprising since one might naturally have more to report if describing what one had just been doing (PR) than if reporting on plans for the future (PI).

In terms of performance at test, it appears that the Prompted Intentions intervention is less effective than the Prompted Reviews intervention. However, this comparison between Experiments 1 and 2 may be somewhat misleading, because variations between cohorts of subjects were noted; for example, compare the number of key presses made by subjects in the unstructured exploration conditions

TABLE 1

*Excerpts of verbal protocols generated by two subjects in the Prompted Reviews and two subjects in the Prompted Intentions conditions*

---

### **Prompted Reviews**

Subject 4 (Errors at test: 0)

- I learned that when you press the Set button, when it's in the Timer, that square indicates . . . kind of colours one of the digits and indicates the Set mode. By pressing the Lap/reset button . . . and then you . . . I can't remember how to do this. [*laughs*]
- The Set button highlights one of the digits and you can move that little highlighted square from hours, minutes, seconds, on the date and press the Start/stop button and depending on what it's on, you can then alter it using the lap/reset button. And that seems to work on setting the Timer.

Subject 8 (Errors at test: 9)

- Been setting the alarm again, and even though I managed to set it and it went off, I didn't know how to reset it because at one stage there was a musical note on the left hand side and there was the alarm signal on the right hand side . . . and then I pressed something and it went and I didn't know how to get it back. But not I've got it back and I don't know how I did it.
  - I've been looking at the . . . it's called the Chrono, or Chrono . . . and I've discovered that different laps come up on the right hand side and I'm not quite sure why because I went through what I've been doing before, using Start/stop button and the Lap/reset button, just stopping and starting and then I get set times for laps 1, 2 and 3.
- 

### **Prompted Intentions**

Subject 4 (Errors at test: 10)

- Um, there's a function on the mode which says Chrono which I'm trying to work out what that does.
- I'm still trying to find out about the Timer. I can't work out the function of it. I thought the Chrono was the timer but it wasn't.
- I've given up on the Chrono for now. I'm trying to make the Alarm to go off.
- I want to find out what that beeping noise is . . . that I've just found by pressing the Lap/reset button.

Subject 7 (Errors at test: 20)

- I'm just going to work out how get the alarm . . . I've worked out how to set it, I'm going to work out how to get it to go at that time.
  - I got the alarm clock to work, I think, I'm just seeing if I can change the time and the date, the present time and the present date.
  - I've worked out the four different modes and how to use them and now I'm just sort of practising
  - I'm trying to work out how to use the two different tones for the alarm.
- 

of Experiment 1 (483) and 2 (670). These variations may be due to the fact that the experiments were run at different times during the university term, and may therefore have attracted volunteers from differing cohorts. Given such variations it seemed wise to compare the two verbalization conditions within one experiment.

## **5. Experiment 3: prompted reviews, prompted intentions, describe screen and unrelated questions**

The aims of this experiment were firstly to find out more about the relative efficacy of reviews and intentions, as mentioned above, and secondly to test a minimalist explanation of what it is about interruptions during exploration which makes the

learner more successful. Could it be that simply taking a break functions as an “incubation” period during which what has been learned has a chance to consolidate? Wallas (1929) defined incubation as an increase in the probability of finding the solution to a problem after a pause, as compared with continued problem-solving. If something like an incubation effect was at work, then interrupting the explorer to verbalize about any topic may yield beneficial effects.

To test this, subjects explored under one of four conditions, where the verbalizations required were differentiated with regard to the content generated and the relevance of that content to the task of finding out how to operate the device. The experimental conditions are described in terms of these factors below.

One of the two new conditions in this experiment. Describe Screen (DS) will require that subjects describe the appearance of the screen they happen to be working on when the interruption is signalled. We think that this type of verbalization will require that subjects expend less cognitive processing about the operation of the device than the PR and PI conditions and thus led to poorer learning. However, since subjects will nevertheless need to focus on the device, albeit at a very superficial level, it should produce some benefits for learners. Finally, in the other condition, Unrelated Questions (UQ), subjects will be asked a series of questions about topics unrelated to the task. It is possible that this may function as a disruption task; hence we predict that subjects in this condition will learn less about the device when compared to subjects in the three conditions described above.

## 5.1. METHOD

### 5.1.1. *Subjects*

Subjects were 32 undergraduates (18 females, 14 males). With regard to previous computer experience, six subjects had none, two subjects had 6 months or less, 16 subjects had 3 years or less, six subjects had between 4 and 8 years, one subject had 10 years (mature student) and one subject had 15 years (mature student). All subjects in the latter five categories listed word processing, and of these subjects, 13 also listed games, six listed statistics, four listed graphics and one listed spreadsheet as packages they had used in the past. One subject had limited programming experience.

### 5.1.2. *Design*

In this experiment we manipulated the content and degree of relevance of the interruption/verbalizations with regard to the task (i.e. discovering how to operate the device). More specifically four groups of eight subjects were interrupted every 2 min in order to do the following:

- (1) Report on what they had learned in the preceding 2 min (PR), or
- (2) Report on what they were planning to work on in the following 2 min (PI), or
- (3) describe the screen they happened to be working on (DS), or
- (4) answer a series of questions unrelated to the task (UQ).

The questions asked of the latter group were of the type; “What type of music do

you prefer?”, “What is your favourite recent/all-time film?”, “Have you made plans for the week-end yet?”, “When do you feel you do your best work: morning or evening?”, “Which country would you most enjoy visiting?”, etc. (see Appendix 2).

## 5.2. RESULTS

The mean times for which subjects in each condition spoke were as follows: PR: 4 min and 30 s (S.D. 1 min and 39 s), PI: 2 min and 14 s (S.D. 1 min and 10 s), DS: 2 min and 53 s (S.D. 55 secs), UQ: 4 min and 9 s (S.D. 1:01). We did not count the number of speaking episodes for each group. The time they spent speaking was part of the 20 min allotted for exploration of the device.

We analysed the data by carrying out a  $1 \times 4$  ANOVA (Prompted Reviews vs. Prompted Intentions vs. Screen Description vs. Unrelated Questions).

### 5.2.1. Number of key presses: exploration

On average, subjects in the PR group employed 432.4 (S.D. 162.9), PI group employed 467.4 (S.D. 84.0), DS group employed 548.8 (S.D. 145.6) and the UQ group employed 382.4 (S.D. 120.3). There was no significant difference between the four groups with regard to how many key presses they employed while they explored the device,  $F(3, 28) = 2.26$ ,  $p = 0.10$ .

### 5.2.2. Post-exploration questionnaire scores

Subjects in the PR group listed an average of 11.1 (S.D. 2.2) functions, PI group listed 8.9 (S.D. 2.5), DS group listed 8.9 (S.D. 1.1) and the UQ group listed 7.9 (S.D. 3.8). ANOVA indicated that we could not argue for a difference between groups on the basis of this test ( $F(3, 28) = 2.26$ ,  $p = 0.10$ ).

### 5.2.3. Number of errors at test

Figure 4 illustrates the significant difference between conditions with regard to how many errors subjects made, ( $F = (3, 28) = 10.6$ ,  $p < 0.0001$ ).

Pairwise comparisons (Newman's Keul) revealed that the differences between groups were that the PR group made significantly fewer errors (8.5, S.D. 3.7) than the PI (errors 32.5, S.D. 11.3,  $p < 0.05$ ), DS (errors 62.2, S.D. 18.4,  $p < 0.01$ ) and UQ (errors 60.1, S.D. 38.4,  $p < 0.01$ ) groups. There was also a significant difference between the PI and both the DS ( $p < 0.05$ ) and UQ ( $p < 0.05$ ) groups.

## 5.3. DISCUSSION

Again there was no significant difference between subjects in the four conditions with respect to how many items they recalled about the device in the Questionnaire. In light of the very large differences at the performance test, this confirms our suspicion that the questionnaire is a rather insensitive instrument. The online test indicated that the both the review and planning subjects had acquired more robust knowledge of procedures than the screen description and unrelated questions subjects, and furthermore, that Prompted Review was a more effective intervention than Prompted Intentions.

One outcome not anticipated was that the performances of subjects in the screen description and unrelated questions conditions would be similarly poor. In fact,

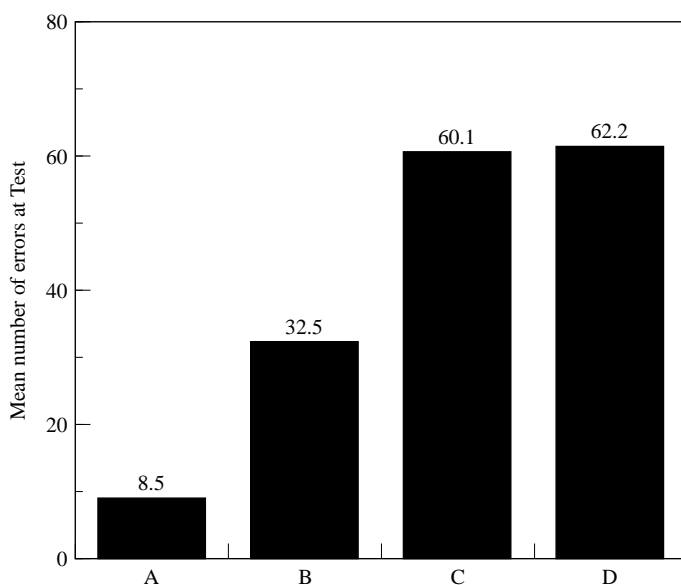


FIGURE 4. Mean number of errors at test. A: Prompted Reviews, B: Prompted Intentions, C: Describe Screen, D: Unrelated Questions conditions.

both groups made more errors on test than subjects in the constrained exploration conditions of Experiments 1 and 2. We expected that the Screen Description group would do better than the Unrelated Questions group because they spent considerable time (14% of the exploration session) describing the appearance of the function buttons, the appearance of the display screens in the four different watch modes. It seemed likely that because their focus remained on the device they would have greater success than subjects who regularly had to abandon what they were doing to talk about unrelated issues. Evidently though, describing the screen was as disruptive to learning as having to answer questions. An explanation may lie in the fact that there were not large variations in the appearance of the screens. The four function buttons were always in the same position, and as can be viewed from Figure 1, the four screens were not that complex. The repetitiveness of the task, especially when describing a screen for a second or third time may be what interfered with the learning.

The possibility that any of the advantages gained by the review and intention subjects may have been due to the quantity of the protocols generated (i.e. total length of time they spoke during the interruptions) can be rejected. The review group spoke for an average of 4 min and 30 s, which represents 22.5% of their exploration time. Yet they were significantly better than the unrelated questions group who spoke for an average of 4 min and 9 s, or 20.8% of the exploration session. We can also compare the intentions group, who spoke for an average of 2 min and 14 s, or 11.2% of the time, to the screen description group whose mean was 2 min and 53 s, 14.4%. Thus we have two sets of groups who spoke for almost the same length of time, yet whose performance was significantly different. In fact a correlational analysis across the four groups, which compared the length of time



they spoke to their performance scores at test, demonstrated that these two factors were not correlated ( $r = 0.224$ ,  $df = 30$ , n.s.). This evidence strengthens the argument that it is the content and not the verbalizations *per se* which affect performance.

The results outlined here allow us to discount the argument that the beneficial effects observed in the review and future intention conditions were due to an “incubation” period—where any type of verbalization pause would have functioned as period where new knowledge had a chance to be processed and stored in long term memory. Subjects were all interrupted at regular intervals yet we witness significant differences in how this affected their performance. We have put forth possible reasons for why reviewing and planning facilitated exploratory learning of the device, while the other two conditions were detrimental to this task.

## 6. Experiment 4: self-regulated reviews

Some of the literature reviewed in the general introduction suggested that many students require external prompting in order to improve their learning (e.g. cues, instructor interruptions, etc.). Having shown that experimenter prompting is indeed *sufficient* to produce a self-monitoring learning benefit for exploratory learners of interactive devices, we were curious to investigate whether prompting was *necessary*. If subjects are allowed to determine *for themselves* when to pause and self-monitor, will learning benefits still be observed?

From the data of the control groups (UE) in Experiments 1 and 2, as well of similar groups in our previous work (Trudel & Payne, 1995) it is evident that most subjects, exploring when no constraints are imposed, tend to engage in high levels of interaction at the expense of reflecting about their learning. We therefore tested whether employing a quite subtle intervention would be enough to effect a change in subjects' behaviour. Subjects were simply told that, in past experiments, it had been found that people learned a lot more about the device if they stopped every few minutes to report what they had learned so far.

### 6.1. METHOD

#### 6.1.1. Subjects

Subjects were 16 undergraduates (seven females, nine males). It is important to highlight the different composition of this sample with regard to subjects in the previous experiments reported in this paper. Because it was run at the end of summer term, none of the subjects participated in this experiment for credit. All subjects were paid £3 for the hour during which they participated. With regard to previous computer experience, three subjects had none, one subject had 6 months or less, three subjects had 1 year or less, four had 4 years or less and five subjects had 7 years or less. All subjects in the latter four categories listed word processing and of these subjects, six also listed games, one listed spreadsheets and four listed graphics as packages they had used in the past. Only one subject had limited programming experience.

### 6.1.2. Design

We compared the performance Unstructured Exploration (EU) subjects to that of Self-Regulated Reviews (SR) subjects. The latter were asked to stop working whenever they felt they had learned something of importance and to state aloud what that was. Therefore, the major difference with Experiment 1 was that *subjects* were the ones who decided when and how often they interrupted their work to record their reviews. The instructions given to this group also stated that “in previous experiments we found that people learned a lot more about the watch if they were required to stop working every few minutes and reported what they had learned so far. People who worked without such interruptions were not as successful as the group who evaluated their progress”.

## 6.2. RESULTS

Subjects in the Self-Regulated condition spoke for an average of 3 min (S.D. 2 min and 9 s). The time they spent speaking took up part of the 20 min allotted for exploration of the device. The vast individual differences in the amount of time subjects spent speaking is quite striking when compared to Experiments 1 and 2. If the mean is calculated without including the time taken by one subject, whose score was 2 standard deviations away from the mean, we obtain a mean of 2 min and 23 s (S.D. 1 min and 18 s). The average number of speaking episodes, i.e. how often subjects chose to stop working in order to report on their progress was 10. Again if we remove the score of the same outlier subject, whose score was again 2 standard deviations from the mean, we obtain a mean of 8 episodes.

### 6.2.1. Number of key presses: exploration

There was no significant difference with regard to how many key presses subjects employed while they explored the device. On average, the SR subjects made 375.9 (S.D. 139.2) key presses while the EU subjects made 341.9 (S.D. 130.6), ( $t = 0.504$ ,  $df = 14$ ,  $p = 0.31$ , one tailed).

### 6.2.2. Post-exploration questionnaire scores

There was a significant difference in recall between the SR and the EU groups. While subjects in the SR group remembered, on average 9.9 (S.D. 2.1) items, the UE group's average was 7 (S.D. 2.5), ( $t = 2.49$ ,  $df = 14$ ,  $p < 0.05$ ).

### 6.2.3. Number of errors at test

Figure 5 illustrates that subjects who interrupted their work during exploration and stated what they had learned so far made significantly fewer errors (SR = 7.4, S.D. 3.5) at Test than the group who explored the device without interruptions (EU = 35.8, S.D. 17.3), ( $t = 4.55$ ,  $df = 14$ ,  $p < 0.0005$ , one tailed).

## 6.3. DISCUSSION

In contrast to the first three experiments reported in this paper, we found a difference between conditions with respect to how many items about the device were recalled. Comparing questionnaire performance across experiments, it appears that the Unstructured Exploration group in Experiment 4 performed particularly

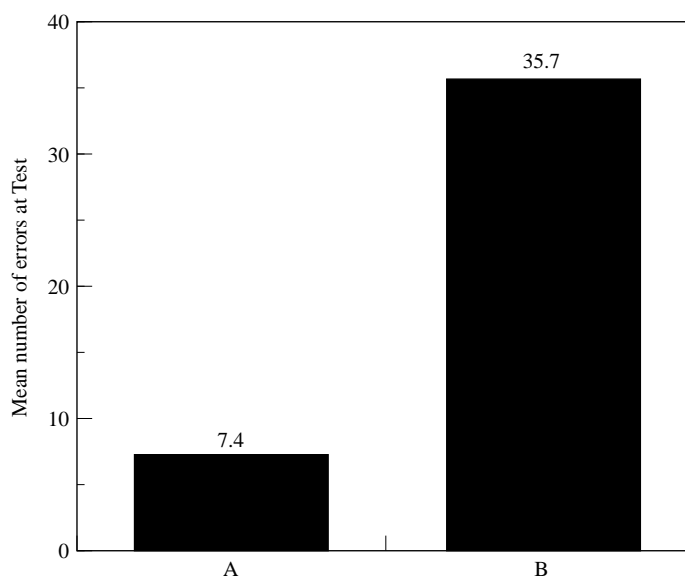


FIGURE 5. Mean number of errors. A: Self-Regulated Reviews, B: Unstructured Exploration.

poorly. The level of questionnaire performance of the Self-Regulated Review condition is in line with that of subjects in the earlier experiments. As for the online test, we again found that the group who reviewed their learning outperformed the group who engaged in uninterrupted exploration. Simply informing subjects at the beginning of the experiment of the benefits of reviewing has prompted them to regularly interrupt themselves and to reap the benefits of this practice.

## 7. General discussion

### 7.1. SUMMARY OF MAIN FINDINGS

Subjects who were required to stop working at regular intervals during exploration, to report on what they had learned so far, outperformed subjects in a control condition who explored without such interruptions (Experiment 1). We then investigated whether “looking ahead”, i.e. pausing to report on what one would work on next, would be as beneficial to learning as we had found reviewing to be. Subjects who stated future intentions indeed made fewer errors, when their ability to operate the device was tested, as compared to subjects in a control condition where exploration was unconstrained (Experiment 2).

In Experiment 3, these conditions were directly compared with each other, and with two conditions in which subjects interrupted their explorations to verbalize content unrelated to their previous or next exploration episode. Reviewing and stating intentions both produced benefits over unrelated verbalization, with reviewing being the more effective. Unrelated verbalization evidently depressed performance compared with the unstructured exploration conditions of Experiments 1 and 2.

Finally, Experiment 4 showed that subjects need not rely on external prompting to effectively use reviewing in order to benefit exploratory learning. Despite large individual differences in the amount of time subjects spent reviewing their knowledge these subjects outperformed control subjects who did not pause to review aloud.

## 7.2. WHY DOES PROMPTED REVIEWING FACILITATE LEARNING?

We reported, in the introduction, several studies which have demonstrated that summarizing what has been learned so far can exert an important positive influence on incremental learning. The majority of these studies were carried out with subjects reading text. Perhaps the major contribution of this article is to establish that this type of intervention can also be successfully applied to situations where one is discovering how to operate an unfamiliar, interactive device by exploring without the benefit of assistance or instructional texts.

That *prompted* reviewing is beneficial has two very separate implications. First, it shows that most learners do not spontaneously review optimally. As well as signalling a dispositional bias (in common with the literature on learning from text) this perhaps emphasizes a downside in the highly interactive properties of modern computer-based devices. The great benefit of interactive devices is that their display-based nature lessens demands on memory (e.g. Payne, 1991). This however may also contribute to the lack of reflection witnessed in learners. When subjects rely on the fact that the display prompts are always available they may be less attentive to the fact that the system is in fact constantly changing and that one needs to focus on what action has brought about what changes (i.e. cause–effect relationships).

An additional aspect of modern interactive systems is that the cost of carrying out an action is low. This means that the subject can rapidly switch from one feature, mode or goal to the next without taking the time to learn anything about any of them.

Secondly, the success of prompted reviewing shows that reviewing *per se* is beneficial to incremental learning. There are several established comprehension and memory phenomena that would predict such an effect. Most straightforwardly, *rehearsal* of to-be-remembered items leads to superior long-term memory (Atkinson & Shiffrin, 1968). Further, reviewing requires explicit integration of so-far-encountered material, and may well engender “deeper” processing. Again, this will lead to more robust long-term encoding. All these effects are likely to be just as important in exploratory learning of devices as they are in learning from text. These memory effects show why the reviewing activity *per se* might have learning benefits. A different kind of effect that is very likely to be in operation is the secondary effect of having to do a review on the between-review behaviour. One can imagine that if a learner knows that she will have to articulate everything that she learns in a particular episode, this might work to focus attention. In the case of exploratory device-learners, the demand of reviewing might encourage the explorer to stick at a problem, so as to learn something reportable, rather than drift around from failure to failure.

### 7.3. WHY DOES PROMPTED PLANNING FACILITATE LEARNING?

As with reviewing, the *success* of prompted planning requires two kinds of explanation. First, it shows that ordinarily users might not plan at optimal levels. As Young and Simon (1987) and others have argued, an HCI environment creates special conditions with regard to planning future actions. In contrast with many domains, the planner simply does not have all the information that is needed to play a sequence of actions from the current state to some goal state; rather, the information becomes available in the course of interaction. In such situations it makes more sense to produce partial plans, to take into account the resulting state and the information gained after any action, and to re-plan from there. The fact that a lower memory load results if only the outline of a plan is created also biases people toward making only partial plans. The current experiments suggest that people may be seduced by the necessity of partial planning into producing sub-minimal plans, at least for the purpose of exploratory learning.

The second implication of the benefits of prompted planning is that planning, or the articulation of plans, must be beneficial to the process of exploratory learning. Unlike the case of reviewing, there is no widely received set of theoretical ideas to explain this phenomenon. Rather than speculate, we prefer at this stage to simply flag this phenomenon as raising an important issue for future research.

### 7.4. OVERCOMING OVERACTING

The above discussion supports our previous conclusion (Trudel & Payne, 1995) that exploratory learners, if left to their own devices (apologies for the pun), will tend to make too many moves, and insufficiently consider each move. We suggest that this is a consistent cognitive bias, that we might call “overacting”.

The literature on judgement and decision making is replete with examples of cognitive biases, such as overconfidence. Fischhoff (1982) has discussed how it is almost impossible to rid people of a such biases. Even when people are expressly warned of the proclivity of these biases they still exhibit a disposition toward making judgmental errors about their knowledge or behaviour.

In contrast, in our work we have been rather successful in devising manipulations which force people to work in a more reflective manner and which have beneficial effects on their learning (e.g. limiting the total number of key presses subjects are allowed to make during exploration, forcing them to interrupt their work to report on their progress, having them work from a list of goals, Trudel & Payne, 1995). The first three experiments in this article extend this set: adding prompted reviewing and prompted planning. However, all these manipulations were externally imposed constraints to which subjects had no choice but conform.

More striking, in view of the robustness of biases reported by Fischhoff (1982), is the fact that with a simple set of instructions (i.e. letting subjects know that in previous experiments people learned a lot more about the device if they stopped working every few minutes and report what they had learned so far) apparently has a very similar effect, overcoming the overacting bias at a stroke.

Perhaps the reason we were able to effect this change in subjects' exploration is because our instructions offered subjects a specific behavioural suggestion, that is a

suggestion to stop and review, as opposed to a more general exhortation to overcome a cognitive bias.

## References

- AHLUM-HEATH, M. E. & DI VESTA, F. J. (1986). The effects of conscious controlled verbalization of a cognitive strategy on transfer in problem-solving. *Memory and Cognition*, **14**, 281–285.
- ANDERSON, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- ASCHEP, C. (1984). Improving the mathematical skills of low achievers. *Urban-Review*, **16**, 187–191.
- ATKINSON, R. C. & SHIFFRIN, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. SPENCE & J. T. SPENCE, Eds. *The Psychology of Learning and Motivation*, Vol 2. New York: Academic Press.
- BALLSTAEDT, S. P. & MANDL, H. (1984). Elaborations: assessment and analysis. In H. MANDL, N. L. STEIN & T. TRABASSO, Eds. *Learning and Comprehension of Text*. Hillsdale, NJ: Erlbaum, pp. 331–353.
- BERRY, D. C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology*, **35**, 39–49.
- BERRY, D. C. & BROADBENT, D. E. (1984). On the relationship between task performance and associated verbalisable knowledge. *Quarterly Journal of Experimental Psychology*, **36A**, 209–231.
- BROWN, A. L., CAMPIONE, J. C. & DAY, J. D. (1981). Learning to learn: on training students to learn from texts. *Educational Researcher*, **10**, 14–21.
- CHI, M., BASSOK, M., LEWIS, M. W., REIMANN, P. & GLASER, R. (1989). Self-explorations: how students study and use examples in learning to solve problems. *Cognitive Science*, **13**, 145–182.
- CHI, M., DE LEEUW, N., CHIU, M. & LEVANCHER, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, **18**, 439–477.
- DE JONG, F. P. C. & SIMONS, R.-J. (1988). Self-regulation in text processing. *European Journal of Psychology of Education*, **v.III**, 177–190.
- ERICSSON, K. A. & SIMON, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215–251.
- ERICSSON, K. A. & SIMON, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- FISCHHOFF, B. (1982). Debiasing. In D. KAHNEMAN, P. SLOVIC & A. TVERSKY, Eds. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, MA: Cambridge University Press.
- GAGNE, R. M. & SMITH, E. C. (1962). A study of the effect of verbalization on problem solving. *Journal of Experimental Psychology*, **63**, 12–18.
- GASKINS, I. G. (1988) Teachers as thinking coaches: creating strategic learners and problem-solvers. *Journal of Reading, Writing and Learning Disabilities International*, **4**, 35–48.
- HEINS, E. D., LLOYD, J. W. & HALLAHAN, D. P. (1986). Cued and noncued self-recording of attention to task. *Behavior Modification*, **10**, 235–254.
- LAN, W. Y., BRADLEY, L. & PARR, G. (1993). The effects of a self-monitoring process on college students' learning in an introductory statistics course. *Journal of Experimental Education*, **62**, 26–40.
- LAVOIE, D. R. & GOOD, R. (1988). The nature and use of prediction skills in a biological computer simulation. *Journal of Research in Science Teaching*, **25**, 335–369.
- METCALFE, J. & WIEBE, D. (1987). Intuition in insight and non insight problem solving. *Memory & Cognition*, **15**, 238–246.
- PAYNE, S. J. (1991). Display-based action at the user interface. *International Journal of Man-Machine Studies*, **35**, 275–289.
- PINTRICH, P. R. & DEGROOT, E. V. (1990). Motivational and self-regulated learning

- components of classroom academic performance. *Journal of Educational Psychology*, **82**, 33–40.
- PRESSLEY, M. & GHATALA, E. S. (1990). Self-regulated learning: monitoring learning from text. *Educational Psychologist*, **25**, 19–33.
- SCHOOLER, J. W., OHLSSON, S. & BROOKS, K. (1993). Thoughts beyond words: when language overshadows insight. *Journal of Experimental Psychology: General*, **122**, 166–183.
- SCHUNK, D. H. (1983). Progress self-monitoring: Effects on children's self-efficacy and achievement. *Journal of Experimental Education*, **51**, 89–93.
- SCHUNK, D. H. & RICE, J. H. (1987). Enhancing comprehension skill and self-efficacy with strategy value information. *Journal of Reading Behaviour*, **3**, 285–302.
- STANLEY, W. B., MATHEWS, R. C., BUSS, R. R. & KOTLER-COPE, S. (1989). Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *The Quarterly Journal of Experimental Psychology*, **41**, 553–577.
- STINESSEN, L. (1985). The Influence of verbalization on problem-solving. *Scandinavian Journal of Psychology*, **26**, 342–347.
- TRUDEL, C. I. & PAYNE, S. J. (1995). Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies*, **42**, 307–339.
- VEENMAN, M. V. J., ELSHOUT, J. J. & BIERMAN, D. J. (1989). Structured and unstructured approaches to simulation: what do the students think? In *Proceedings of the Sixth Canadian Symposium on Instructional Technology*, pp. 457–461. Halifax, Canada: NRCC.
- VEENMAN, M. V. J., ELSHOUT, J. J. & BUSATO, V. V. (1994). Metacognitive mediation in learning with computer-based simulations. *Computers in Human Behaviour*, **10**, 93–106.
- VEENMAN, M. V. J., ELSHOUT, J. J. & GROEN, M. G. M. (1993). Thinking aloud: does it affect regulatory processes in learning? *Tijdschrift voor Onderwijsresearch*, **18**, 322–330.
- WALLAS, G. (1926). *The Art of Thought*. New York: Harcourt Brace.
- WILDER, L. & HARVEY, D. J. (1971). Overt and covert verbalization in problem-solving. *Speech Monographs*, **38**, 171–176.
- YOUNG, R. M. & SIMON, T. (1987). Planning in the context of human-computer interaction. In D. DIAPER & R. WINDER, Eds. *Proceedings of the 3rd conference of the British Computer Society—HCI Specialist Group*. University of Exeter, 7–11 September.

## Appendix 1: questionnaire given to all subjects after exploration

(Answers appear below each question)

Without referring to the watch please write down your answers to the questions below:

1. What are the different **modes** of the watch?
  - Normal time and date display
  - Chronograph (stopwatch)
  - Countdown Timer
  - Alarm (wake-up and hourly beep)
2. What does the **set** button allow you to do?
  - Allows you to alter the different fields (ie. digit or day) by placing a grey “select” or “highlight” box onto the field to be altered
  - Removes “select” box from display once field(s) has been altered
  - Resets Chronograph to zero when in Laps function
3. What does the **start/stop** button allow you to do?
  - Moves “select” box around to different fields
  - Starts/stop Chronograph
  - Starts Countdown Timer

4. What does the **lap/reset** button allow you to do?

- Controls the icons (i.e. alarm chime and hourly beep); appear and disappear
- Advances digits (e.g. from 7 to 8, 9, 0, 1 etc.) and day (e.g. from Mon to Tues, Wed etc)
- Resets Countdown Timer to programmed time (e.g. if set to 10 s, pressing lap/reset will return bottom display to 10 s until countdown is started again).
- Resets Chronograph to zero—when NOT in laps function
- Records each individual lap (up to eight laps)
- Reviews individually recorded laps stored in memory.

## **Appendix 2: instructions for Experiments 1, 2, 3 and 4**

Subjects in every conditions (including Unstructured Explorations) were told:

Your task is to discover how to operate the functions of this watch (i.e. find out all that you can about what the different buttons do). You may try anything you want as you do this experiment.

### **Experiment 1: Prompted Reviews (PR)**

As a way of keeping track of what you are learning as you work with the device we would like you to stop every 2 min and say, out loud, what you have learned so far (in the preceding 2 min).

The experimenter will tell you when 2 min have elapsed so that you don't have to worry about keeping track of the time. However she is not permitted to give you any feedback or answer any questions.

### **Experiment 2: Prompted Intentions (PI)**

Every 2 min we will ask you to stop working so that you can say, out loud, what you are going to try and achieve in the next few minutes.

The experimenter will tell you when 2 min have elapsed so that you don't have to worry about keeping track of the time. However she is not permitted to give you any feedback or answer any questions.

### **Experiment 3: PR, PI, Describe Screen (DS) and Unrelated Questions (UQ)**

Instructions for PR and PI in this experiment were identical to those given in Experiment 1 and 2.

*For describe screen:* Every 2 min we will ask you to stop working so that you can describe, out loud, what the screen you are working on at that time looks like. As a guide to what we expect you to say here, imagine that you are describing the screen so that someone who listened to an audio tape of what you have described would be able to draw it.

*For Unrelated Questions:* Every 2 min we will ask you to stop working so that the experimenter can ask you questions about various topics. You need only give brief answers (i.e. talk between 30 s and 1 min)



### Experiment 4: Self-Regulated Reviews (SR)

In previous experiments we found that people learned a lot more about the device if they were required to stop working every few minutes and reported what they had learned so far. People who worked without such interruptions were not as successful as the group who evaluated their progress.

Now that you are aware of what strategy leads to better learning, we would like to encourage you to make use of it. Therefore we ask you to stop whenever you feel it is appropriate and report (out loud) what you have learned.

### Subjects in every conditions (including Unstructured Explorations) were also told:

**LIMIT:** You have **20 min** to do this exercise, so use your time efficiently. After that time you will be given a test which will draw on your knowledge about the different device functions.

Please remember that this is **not** an intelligence test and that I am not testing you personally. I am simply observing what people try out as they attempt to discover how the functions of the watch work.

### Appendix 3: instructions for test

You will now be presented with six tasks which you should try to carry out using the different functions of the watch. There is a maximum time for each task, so try to do each one as quickly and efficiently as possible—the experimenter will tell you how much time you have at the start of each exercise.

The tasks will be presented to you one at a time. Each one has to be carried out in the order presented—you will only have one chance to complete each and will not be able to return to tasks presented earlier.

Even if you are unsure how to carry out one of the tasks, it is better to try something than to skip it.

### On-line test

1. Program the watch so that the time which appears on the display is 4 h and 10 min later than what is currently indicated (i.e. add 4 h and 10 min to the current time).
2. Program the watch so that the day and date which appear on the display are 3 days later than what is currently indicated. (e.g. from Tuesday, 3rd to Friday, 6th.)
3. Program the watch so that it beeps every hour, on the hour. Then check to see that you have programmed it correctly. (Hint: don't wait for a full hour to go by!)
4. Program the watch so that the alarm rings at 7:15 am. Then check to see that you have programmed it correctly.
5. Program the watch so that it counts down from 10 s, until it reaches zero.
6. You want to measure the total time a friend spends swimming but do not want to include the time during which she takes two rest periods.

*The experimenter will indicate the start and when the two rest periods are taken. Tell the experimenter when you are ready to begin.*