BEHAVIOUR & INFORMATION TECHNOLOGY, 2001, VOL. 20, NO. 1, 23–35

The role of age, foreknowledge and complexity in learning to operate a complex device

DANIEL FREUDENTHAL†‡

†IPO Center for User–System Interaction, Eindhoven University of Technology, PO Box 513, 5600 MB, Eindhoven, The Netherlands

‡Present address: School of Psychology, University Park, Nottingham NG7 2RD, UK; e-mail: DF@psychology.nottingham.ac.uk

Abstract. Old (60 – 70 years) and young (18 – 25 years) adults performed two experiments in which they had to learn to operate a simulated device. It was assessed whether age differences in performance were comparable over groups with different levels of foreknowledge. Of particular interest was the question how age differences between groups with limited foreknowledge compared to those between groups with foreknowledge. The role of complexity of operating procedures was studied by using tasks which differed in the number of actions that was needed for completion. In experiment 1, which employed three complexity levels, no effects of age were found as a function of presence of foreknowledge or complexity of procedures. Experiment 2, however, which employed an additional complexity level, showed that the performance of old adults with foreknowledge was poorer than that of young adults with foreknowledge. Complexity had no differential effect with respect to either age or amount of foreknowledge.

1. Introduction

Older adults appear to be at a disadvantage when learning to use interactive equipment. Research by Elias *et al*. (1987) and Zandri and Charness (1989) has shown that, when training people of different age groups to use interactive devices, older people need more time and request more help than young people. A study by Czaja and Sharit (1993) showed that, after they were trained on text-editing tasks, the older participants were slower than the young ones and made more errors, especially on tasks requiring what they called 'more information processing components'. A survey concerning Automatic Teller Machine (ATM) usage by adults of varying ages (Rogers *et al*. 1996a) showed that older adults are less likely to use ATMs, use them less often and rate their use as being more difficult. Research by Birdi and Zapf (1997) has also shown that older people react with stronger negative emotions to errors during computer-based work.

One possible cause for the slower learning and poorer performance of elderly people may be the fact that the elderly have less experience in using interactive equipment. Another possible cause may lie in changes in cognitive processing associated with increasing age. Increasing age has been shown to be associated with poorer working memory capacity (Craik and Bosman 1992), reasoning ability (Salthouse 1992), and learning (see Kausler 1994 for a review).

2. The role of foreknowledge and expertise in learning to operate interactive equipment

Though foreknowledge and expertise may seem equivalent at first, the two concepts do denote something different. Foreknowledge refers to specific knowledge elements concerning a specific application. Foreknowledge is different from expertise in that the latter includes aspects of skill and general (background) knowledge that has been acquired over a longer period of time.

Few studies have assessed the influence of age-related differences in foreknowledge or expertise on the performance of interactive tasks. Foreknowledge is likely to have an effect on performance, as has been shown by, for instance, Kieras and Bovair (1984) and Halasz and Moran (1983). In both these studies, participants who received a metaphor for the workings of a device were better able to derive operating procedures on this device. With respect to age-related differences in (computer) experience, Czaja and Sharit (1993), assessed the influence of age and (number of

Behaviour & Information Technology
ISSN 0144-929X print/ISSN 1362-3001 online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01449290010020666

years of) computer experience on performance of interactive tasks following training. Czaja and Sharit found older participants to perform more poorly than young ones. However, older participants also had less experience in using a computer. Correlations between performance measures and computer experience were generally higher than between age and performance measures. Age did explain variance after computer experience was taken into consideration though. As the age range used by Czaja and Sharit was limited (mean age 48.04, s.d. 13.84), age effects may have been underestimated. Furthermore, it is difficult to draw conclusions on the basis of an observed (rather than manipulated) variable. A variable such as computer expertise is difficult to measure and may be mediated by learning and problem-solving ability. Thus, a clearer way of addressing the relative performance of old and young persons as a function of their expertise is to study how well they are able to derive operating procedures on a device which is equally novel to young and old adults, while the amount of foreknowledge they have is manipulated.

Attempts to manipulate the training regime in order to find age differential effects have yielded mixed results, and were usually concerned with training participants on existing applications. Expertise differences are therefore likely to influence the results.

Caplan and Schooler (1990) studied age differences in learning to use the Fullpaint program (a computer drawing program) in relation to the presence of a mental model. A group of middle-aged participants who were trained using a 'desktop-metaphor' of the device performed worse than middle-aged participants without this model. For young participants, the reverse was true; young participants with the model outperformed those without. Charness *et al.* (1992) assessed to what extent older and younger adults being trained in word processing were aided by the provision of an advance organizer, but found no main effect of the presence of the advance organizer, nor an interaction with age. Similarly, Czaja *et al.* (1989) found no age differential effects of training type (instructor-based, online or manual) in learning a text-editing system. Rogers *et al.* (1996b) did find differential effects for different training regimes on ATMs, but employed only one age group.

In general, it thus appears that expertise mediates age differences in the performance of computer-based work, and that foreknowledge can be beneficial to users. However, it is not clear whether old and young users will benefit to the same extent from foreknowledge. The studies described in this paper were aimed at assessing the relative importance of age and foreknowledge in learning to perform interactive tasks under conditions in which the amount of foreknowledge is manipulated.

3. Age differences in knowledge utilization

An area in which age differences in knowledge utilization has been studied is that of memory for text in relation to the presence of a thematic subject or schema. The term schema was originally coined by Bartlett (1932) and denotes a mental structure of past experiences. Bartlett showed that recall of a narrative could be distorted to fit with general expectations and knowledge. In the 1970s, schema theory was developed as a more comprehensive theory of memory (see, for example, Rumelhart and Ortony 1977). According to schema theory, the knowledge that a person possesses is structured in schemas, which are subsumed under each other. These schemas are used to actively process and encode incoming information. The study of how more general knowledge can affect memory for a text has been referred to as *schematic influences on memory* (Hess 1990). In a typical experiment on schematic influence on memory, participants might read a piece of text that is either preceded or not preceded by a thematic subject (e.g. washing clothes). Their memory is then tested on the piece of text. In general, memory of the text is better when a thematic subject is present. As Arbuckle *et al.* (1990) point out, three hypotheses have been put forward regarding age differences in schematic influence on memory. According to the *age invariance hypothesis*, young and old persons use schematic knowledge to a similar extent. The age invariance hypothesis predicts that age differences in memory for a text are equally large when thematic support is present as when it is absent. The *developmental shift hypothesis* posits that old adults utilize schematic support more than young adults do. This hypothesis predicts that the provision of a thematic subject will reduce age differences relative to the situation where a thematic subject is absent. Finally, according to the *production deficiency hypothesis*, young people are better able to use schematic support; age differences in memory for a text will be largest in the situation where a thematic subject is present.

Though some authors have argued that elderly people benefit more from schematic support (Hess 1990; Craik and Bosman 1992), it appears that the hypothesis which best describes the data depends on the amount of encoding which is required for the type of information that is to be recalled. In favour of the age-invariance hypothesis, Zelinski and Miura (1988) found that young and old adults benefited to the same extent from the presence of a thematic subject to a passage. Arbuckle *et al.* (1990) report three studies which showed that 'schema access facilitated memory equally across age levels' (p. 305). In support of the developmental shift hypothesis, Stine and Wingfield (1987) found age differences in memory for sentences to increase when prosodic and semantic cues were removed. Apparently, older persons rely on these cues to a larger extent than young ones do. Hess *et al*. (1989) had participants read 'scripted stories', and tested their recall and recognition. Though young participants exhibited better memory overall, 'age differences in performance decreased as the relevance and typicality of target information increased' (Hess *et al*. 1989, p. 89). It should be noted though, that semantic and prosodic information is likely to be processed relatively automatically. Similarly, encoding of typical information is probably less demanding than encoding of untypical information.

Finally, evidence for the production deficiency hypothesis has come from studies by Charness (1981) and Smith *et al*. (1983). In both these studies, however, the condition which showed the elderly to use schematic support to a lesser extent than young adults required the participants to impose their own organization on the materials (which is different from the studies providing support for the age invariance and the developmental shift hypotheses). This might be caused by elderly people being less likely to elaborate on the information to be remembered (Rankin and Collins 1985).

It might be argued that the size of age differences in memory for scripted information increases as a function of the amount of elaboration that has to be performed, or that it may decrease with the automaticity of encoding the information which has to be recalled (Hasher and Zacks 1979). This suggestion is supported by experiments concerning automatic and controlled memory processing (Jennings and Jacoby 1993; Titov and Knight 1997), which showed that age differences in automatic memory processes are relatively small compared to controlled memory processes. With respect to expertise, similar findings have been reported by Morrow *et al*. (1992, 1994), or see Morrow and Leirer (1997) for an overview of ageing, pilot performance and expertise. Morrow *et al*. (1994) found that age differences on reading back Air Traffic Control messages were smaller for pilots than for non pilots. Expertise did not reduce age differences in a study where subjects had to choose referents for sentences in aviation narratives.

It thus appears that expertise or the availability of knowledge can reduce age differences in memory performance, at least in situations where the information to be remembered is encoded relatively automatically. The task of learning to operate a device given a certain amount of foreknowledge seems slightly more complex. Since this task is essentially a problem-solving task, it appears that participants must impose an organization on the materials. It can therefore be expected that young participants are better able to benefit from the availability of foreknowledge. The situation in which foreknowledge is absent, however, is one that has hardly been studied with respect to age differences. It is thus not clear how age differences in performance with foreknowledge will compare to those without foreknowledge. Such a comparison can shed light on the causes of age differences in performance of interactive tasks.

4. Complexity

Another issue that seems relevant in the domain of learning to use interactive devices is the complexity of operating procedures. Czaja and Sharit (1993) showed that, on trained tasks, older adults made more errors on the tasks requiring more 'information processing components'. According to the complexity hypothesis (Cerella *et al*. 1980), age differences in task performance increase with increasing task complexity. The complexity hypothesis has received support in a study on reasoning and spatial abilities by Salthouse *et al*. (1989) and in a divided attention study by McDowd and Craik (1988), though Charness and Campbell (1988) found only a weak relation between complexity and age differences in performance. Unfortunately, in the ageing literature, the issue of complexity has not been studied in relation to expertise or foreknowledge. The studies discussed here aimed to answer the following questions.

- When participants are learning to use interactive devices, are age differences in performance between participants without foreknowledge equal, smaller or larger than between participants with foreknowledge?
- To what extent are effects of complexity of procedures mediated by age and amount of foreknowledge?

In order to answer these questions, subjects were asked to perform tasks of varying complexity on a simulated medical laser. Task complexity was manipulated by varying the number of preconditions that had to be met in order to complete a task. Foreknowledge was manipulated by varying the extensiveness of the instructions. Half the participants were given instructions that only stated the overall functionality of the device. The other half received instructions that gave more detailed information about operating procedures. The major performance measure used in the experiment was the number of actions that participants needed to perform the different tasks.

It seems appropriate to point out at this point that the manipulation of foreknowledge does not rule out differences in expertise. Though the application used in the reported experiments was new to all participants, it cannot be ruled out that differences in background knowledge concerning technology in general affect performance. By using an unknown application and manipulating foreknowledge, it was hoped to minimize the potential effect of age differences in expertise.

5. Experiment 1

5.1. *Method*

5.1.1. *Subjects*: Forty people participated in the experiment; 20 young (M = 21.9, SD = 2.2), and 20 old (M = 66.6, SD = 4.0). All older participants had received higher vocational training, and all young participants were undergoing higher vocational training. Participants were paid NLG 8.- for their participation. All participants were drawn from the IPO subject pool.

5.1.2. *Apparatus*: A device similar to that of Kieras and Bovair (1984) was developed using Visual Basic. The device, which is shown in figure 1, represents a simulated medical laser with three power levels for scanning and treating illnesses. The device consists of six control elements and five indicator lights. The switch on the left is a power switch. When this is switched to '1', the indicator light above it comes on. On the right are three push buttons labelled 'red', 'yellow' and green' and three lights in corresponding colours. These lights reflect the three different power levels (green is lowest, red is highest). The push buttons are used to activate the different power levels. When the laser is active, the light corresponding to the power level being employed flashes three times. The two rotary switches in the middle are used to enable operation of the middle and high power levels (which 'require more energy'). The top rotary can be viewed as a safety switch. When this is in the 'Sc'

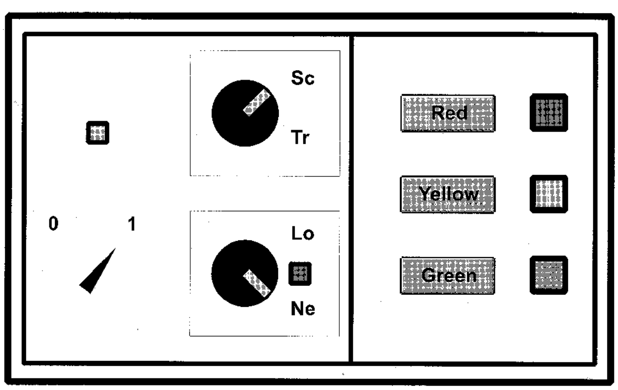

Figure 1. The experimental device for experiment 1.

(Scan) mode, only the lowest (green) power level can be used. In order to use the higher power levels, this rotary needs to be switched to the 'Tr' (Treatment) position. The lower rotary switch controls the battery, the state of which is signalled by the indicator light directly to the right of it. When the lower rotary is in the 'Ne' (Neutral) position, the highest power level cannot be used, since the battery is 'disconnected'. In order to use this highest power level, the rotary needs to be switched to 'Lo' (Load). This results in the battery being charged, which is signalled by the red 'battery light' turning yellow (charging) and then green (charged). Only when the battery is charged *and* the top rotary is in the 'Tr' position, can the highest power level be used. The procedures for using the three power levels are as follows:

> Low power: 1. Power switch to '1'. 2. Press 'green'
> Medium Power: 1. Power switch to '1'. 2. Top rotary to 'Tr'. 3. Press 'yellow'
> High Power: 1. Power switch to '1'. 2. Top rotary to 'Tr'. 3. Lower rotary to 'Lo' (wait for light to turn green). 4. Press 'red'.

The operating procedures for the three power levels thus differ in the number of actions that need to be performed (or number of preconditions that need to be met) in order to use the power levels. All control elements are operated by positioning a mouse cursor over them and pressing the mouse button. When the power is switched off, the device does not respond to any actions, though the two rotary switches in the middle can be moved. As a result, the order of using the power switch and the rotary switches can be interchanged. In order to use the medium power level, for instance, one can also start by setting the top rotary to 'Tr', then switch the device on and then press 'yellow'.

Though it is not clear whether a single optimal strategy can be defined for this task, some measures of inefficiency can be defined. One measure concerns the relevance of the buttons that are pressed. Not all buttons are relevant for all complexity levels. For the green complexity level, for instance, the position of the limiter and the state of the battery are irrelevant. For each complexity level, the 'laser activation' buttons for all the other complexity levels are irrelevant.

5.1.3. *Procedure*: Participants were given instructions that stated that they were to learn the operating procedures for a medical laser with three differing power levels. Half the subjects were given no information regarding the device's operating procedures. This group is referred to as the *no model* group. The instructions for the no model group are shown in table 1.

Table 1. Instructions for the no model group (translated).

In this study you have to try to operate a simulated device. The device in question is a medical laser. This laser has three levels that differ in strength. When the laser is active, this is signalled by a flashing light. For every power level of the laser there is a light of a different colour. For safety reasons there are differences in the actions you have to perform in order to use the different levels.

Your task in this study is to try to use the different power levels of the laser, i.e., make the different colours of lights flash. You can do this by operating controls on the device. In the experiment you will be asked to make lights of a specific colour flash. You are to try to do this with as few actions as possible. The time you take is not important. Before the experiment starts, you will be given the opportunity to practice with the controls used in the device.

If you have any questions, please ask them now.

Table 2. Instructions for the model group (translated).

In this study you have to try to operate a simulated device. The device in question is a medical laser. This laser has three levels that differ in strength. When the laser is active, this is signalled by a flashing light. For every power level of the laser there is a light of a different colour. For safety reasons there are differences in the actions you have to perform in order to use the different levels.

Your task in this study is to try to use the different power levels of the laser, i.e., make the different colours of lights flash. A brief explanation of the procedures is given below.

For the lowest power level, little energy is required. This level can always be used immediately.
For the medium and highest levels more energy is needed. Because these levels are potentially harmful, an extra safety switch (limiter) has been added. When the limiter is *on*, the maximum amount of energy that can be used is limited. The medium and highest level can therefore only be used when the limiter is *off*.
When the limiter is off, the medium level can be used immediately.
In order to use the highest level, a battery has to be charged as well. In order to do this, the device has to be put in charge mode. During the charging, a yellow light comes on.
This turns green when the battery is charged. After the laser has been used, the battery is empty and has to be recharged. This happens automatically if you leave the device in charge mode.

In the experiment you will be asked to make a light of a specific colour flash. You are to try to do this with as few actions as possible. The time you take is not important. Before the experiment starts, you will be given the opportunity to practice with the controls used in the device.

If you have any questions, please ask them now.

Note: when participants asked questions concerning the experimental device they were referred to the instructions. Only questions concerning the procedure were answered.

The instructions for the other participants contained a description of the device's operating procedures as well as some metaphorical information on the level of the different control elements (e.g. the instructions that the device contained a battery). This group is referred to as the *model* group. All instructions stressed that, in completing the tasks, it was important to minimize the number of actions rather than finishing the task as fast as possible. The instructions for the model group are shown in table 2.

After reading the description of the procedures, participants in the model group were informed they would be tested on their knowledge, and it was suggested that they read the part containing the operating procedures once more. After re-reading this description, participants filled out a questionnaire to test their knowledge of the procedures. The questionnaire consisted of nine multiple choice questions where they had to indicate in what position certain control elements have to be in order for a particular power level to be used. The nine questions covered the relevant position of the rotary for the limiter and battery, as well as the actual status of the battery (loaded or not), for all three power levels. For the steps following the questionnaire, the procedure was identical for all participants. Participants were shown a screen containing all the control elements that were used in the experimental device. This could be used to practice using all the control elements. After they had tried all elements, the experimental device was started. All control elements were again pointed out to the participants and they were informed that these were all (and the only) elements that were needed to interact with the device. The instructions were not available to participants when they performed the tasks.

The tasks were displayed in the top part of the screen and remained visible until the task was completed. All tasks were of the form: 'Attempt to use the highest power level; the red light should flash'. After a task was completed, the device was taken off the screen, and participants pressed a button to start the next task. All tasks started with the controls in the positions shown in figure 1. Tasks were organized into four rounds of three trials each. In every round subjects were to use the three different power levels (so every task was completed a total of four times by every participant). The order in which the power levels were to be used was randomized over participants, but fixed over rounds. After every round, participants were tested on their knowledge of operating procedures. This was done by showing them eight pictures of the experimental device with the controls in different positions (this corresponded to all possible states of the device with the power on). For each picture, subjects were to indicate whether or not they could use the 'green, yellow and red' power level. The order of presentation of the pictures was randomized over participants and rounds; answers were recorded by the experimenter. While participants interacted with the device, the reaction times and contents of all their mouse presses were recorded in a log file. The major dependent variable was the number of actions performed by the participants.

5.2. *Results*

Based on the results of an exploratory analysis one outlier was removed from the data. This young participant in the model group had an extremely high score on the first trial in the first round (this score was more than five standard deviations removed from the average score for the entire population for that trial). This score was extremely high, both with respect to this participant's other data points and to all data points of the other participants. All data from this participant were removed. The number of actions was considered the most important dependent measure here. Latency data were considered less relevant because elderly people are likely to have longer latencies for all tasks and because latencies will increase more actions are needed to complete a task.

5.2.1. *Number of actions*: Mouse presses that followed the previous mouse press within 500 ms were discarded. Instances where the mouse was pressed twice in such a short time were thought to reflect a mouse operating problem rather than an attempt to press a control element twice (in all instances where two presses were less than 500 ms apart, the same button had been pressed). The number of discarded mouse-button presses amounted to less than 4% of the total number.

Since the minimum number of actions to complete a task is different for the three different power levels, raw scores are not an appropriate measure. Deviation scores calculated as the number of actions minus the minimum number required to complete the task were therefore computed for every trial. Figure 2 shows the data in a graphical format, averaged over rounds and goals (separately).

Before analysis, data were averaged over rounds and the distribution of data was plotted. This plot showed the distribution to deviate from normality, which constitutes a violation of assumptions for analysis of variance. To accommodate these assumptions, data were log-transformed before analysis. As there were many zeros in the data, a value of 0.1 was added to each score to offset data from zero before it was log-transformed. The value of 0.1 was chosen because it appeared to provide the best approximation of normality (through visual inspection of the distribution). Log-transforming the data has the additional advantage of decreasing the inequality of variances that was also apparent in the data.

The transformed data were submitted to a $2 \times 2 \times 3 \times 4$ repeated measures ANOVA with age (2) and model (2) as between subject factors and complexity (3) and round (4) as within subject factors. The main effect of model was significant ($F(1,35) = 31.83$, $p < 0.001$; participants with a model needed fewer actions than participants without a model). There also was a main effect of round ($F(3,33) = 19.36$, $p < 0.001$, all participants learn over rounds), as well as complexity ($F(2,34) = 3.45$, $p < 0.05$, all participants need more actions with increasing complexity). The only significant interaction was the model × round interaction ($F(3,33) = 3.73$, $p < 0.05$). Separate analysis of the model group and the no model group showed the effect of round to be more pronounced in the model group ($F(3,16) = 21.27$, $p < 0.0001$) than in the no model group ($F(3,15) = 3.60$, $p < 0.05$). Participants with a model thus appear to learn faster than those without a model. No other interactions were significant.

5.2.2. *Questionnaires*: Average scores on the questionnaires are shown in table 3.'Pretest' score refers to the questionnaires filled out by participants in the model condition after they read the instructions. A score of 3.0 refers to chance level for this variable. Participants in the no-model condition did not receive this questionnaire, and thus have no score on this measure. To test whether young and old participants in the model group differed on their pretest scores, a t-test for independent samples was performed. This test showed no significant difference between the groups ($t(18) = 1.26$, $p < 0.2$, two-tailed test). Young and old participants thus seemed to equal each other with respect to how much they learned from the instructions.

The knowledge scores after round 1,2,3 and 4 refer to the questions concerning the pictures of the device that (all) participants answered after each round. As this questionnaire consisted of 24 yes-no questions, a score of 12.0 corresponds to chance level. Knowledge scores were submitted to a $2 \times 2 \times 4$ repeated measures ANOVA with age (2) and model (2) as a between subjects factor and round (4) as a within subjects factor. As scores are proportional in nature, they were logit-transformed before analysis. The ANOVA revealed a main effect of model, ($F(1,34) = 16.19$, $p < 0.001$) and round ($F(3,32) = 8.40$, $p < 0.001$). Participants in the model condition had higher knowledge scores than those in the

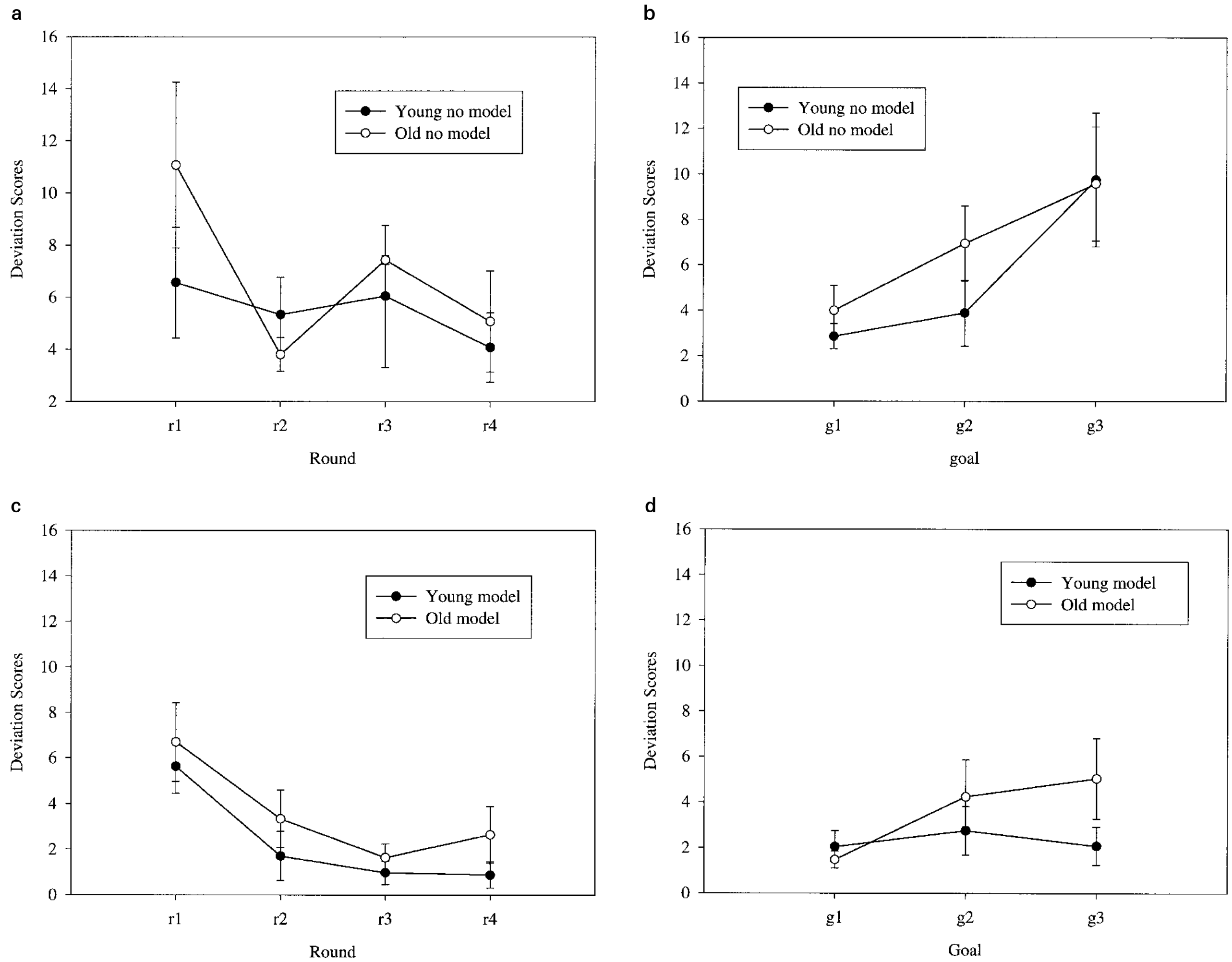

Figure 2. Deviation scores for experiment 1. (a) Deviation scores for the no model groups, averaged over goals. (b) Deviation scores for the no model groups, averaged over rounds. (c) Deviation scores for the model groups, averaged over goals. (d) Deviation scores for the model groups, averaged over rounds.

Table 3. Knowledge scores as a function of age, model condition and round (standard deviation are shown in parentheses).

Age	Condition	Pretest	Round 1	Round 2	Round 3	Round 4
Young	No Model	n.a.	11.25 (3.33)	11.63 (2.50)	14.25 (3.24)	14.25 (4.20)
	Model	6.60 (1.58)	17.70 (4.45)	18.50 (5.58)	20.50 (3.84)	19.80 (5.79)
Old	No Model	n.a.	11.00 (1.15)	11.10 (2.92)	12.20 (2.97)	12.90 (1.97)
	Model	5.70 (1.77)	15.20 (4.69)	15.80 (5.77)	16.50 (5.84)	16.50 (6.02)

no-model condition. Knowledge scores increased over rounds for all groups. No interactions were significant.

5.2.3. *Efficiency*: As was mentioned earlier, it is difficult to define optimal strategies for the subjects' task in this experiment. The number of irrelevant actions does, however, give some indication of the participants' efficiency. As the total number of actions increases with irrelevant actions, analysis was performed on the proportion of irrelevant actions. The proportions of irrelevant actions (depicted in table 4) were logit-transformed before analysis. To avoid proportions of 0.00, 1 was added to the number of irrelevant actions and 2 was added to the total number of actions. This provides a Bayesian estimate of the occurrence of irrelevant actions. Analysis of proportions showed main

Table 4. Experiment 1: proportion of irrelevant actions (averaged over complexity levels) as a function of age and presence of model and round (standard deviations are shown in parentheses).

Age	Condition	Round 1	Round 2	Round 3	Round 4
Young	No Model	0.36 (.14)	0.23 (.13)	0.26 (.14)	0.26 (.16)
	Model	0.31 (.14)	0.11 (.12)	0.09 (.14)	0.08 (.13)
Old	No Model	0.38 (.15)	0.30 (.12)	0.36 (.17)	0.28 (.12)
	Model	0.30 (.15)	0.20 (.15)	0.15 (.15)	0.15 (.18)

effects of model ($F(1,35) = 11.12$, $p < 0.005$) and round ($F(3,33) = 16.30$, $p < 0.001$). The model × round interaction was significant as well ($F(3,33) = 3.60$, $p < 0.05$). As can be seen in table 4, the proportion of irrelevant actions decreases over rounds for all groups. The decrease in proportion of irrelevant actions is substantial for the model group, but slight for the no model group.

5.3. *Discussion*

Analysis of experiment 1 showed main effects of model, round and complexity only. Since there was no significant effect of age, it appears that expertise differences related to age did not play a role in this experiment. Apparently young participants were not helped by their greater technological expertise. For both age groups the presence of foreknowledge was equally helpful in decreasing the number of actions needed to complete the tasks. As such, these data would support an age-invariance hypothesis. All participants needed more actions with increasing complexity, and they all learned over the rounds, though this effect seemed more pronounced for participants with foreknowledge. Participants with foreknowledge were better able to avoid control elements irrelevant to the task. No effects of age (or interactions involving age) were apparent in this experiment. This lack of age differences may seem surprising given the general literature on cognitive ageing. Inspection of figure 2 does, however, suggest some potential effects associated with age. Specifically, in the group with foreknowledge, age differences in performance seemed to increase with increasing complexity. A more detailed analysis also suggested that on the highest complexity level, old participants appeared to learn less quickly than young ones (especially for participants with foreknowledge). The high variance in the data and the deviation from normality may have obscured these effects. A second experiment was therefore run. In this second experiment, some measures were taken in an attempt to lower the variance, and an additional complexity level was employed.

6. Experiment 2

6.1. *Method*

6.1.1. *Subjects*: Forty-eight participants, of whom 24 old (M = 62.5, SD = 2.4) and 24 young (M = 20.3, SD = 2.0) participated in experiment 2. Educational requirements were identical to those in experiment 1. Participants were recruited through an advertisement in a local newspaper. None of the participants in experiment 2 had participated in experiment 1.

6.1.2. *Experimental device*: Experiment 2 aimed to investigate the following issues. Data from experiment 1 seemed to indicate that old adults learn more slowly than young adults with the higher complexity levels, especially in the model group. However, this effect was not significant. This effect may have failed to reach significance due to the high variance. In order to assess whether older people have more difficulty dealing with this type of complexity, two modifications were made to experiment 1.

Firstly, an extra complexity level was added. Operating procedures for the power levels used in experiment 1 remained unchanged. The new complexity level constituted a higher power level. A calibration button and a 'purple' button and light were added to the device. In order to use the new power level, the same actions had to be performed that were needed to use the red laser in experiment 1. In addition, the calibration button had to be pressed. This resulted in a purple light next to the calibration button being lit. This light stayed on for 5 seconds, during which time the purple laser could be fired by pressing the 'purple' button.

Another change that was made to the device was that the labels of the buttons were changed to ones that were considered more informative. This change was made in an attempt to lower the variance by decreasing participants' uncertainty about the function of the different controls. Figure 3 shows the device for experiment 2. The instructions for experiment 1 were changed to incorporate the new power level. This did not change the descriptions of the other power levels (though

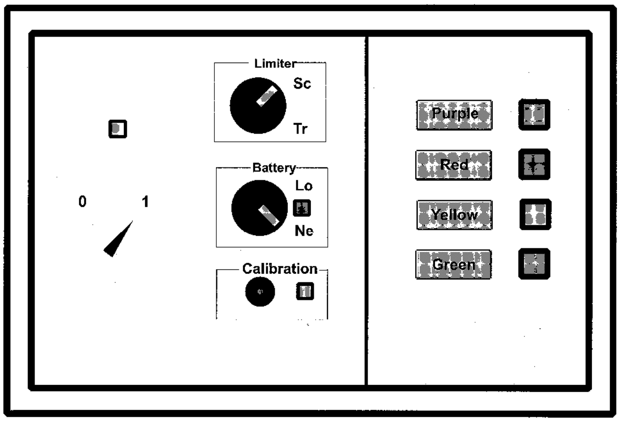

Figure 3. The experimental device for experiment 2.

the new instructions referred to power levels 1, 2, 3 and 4, rather than low, medium and high.). The experimental device for experiment 2 is shown in figure 3.

6.1.3. *Procedure*: As in experiment 1, participants completed four rounds, with every task occurring once in every round. Task order was randomized over particpants but not over rounds. Since the number of possible randomizations of four was higher than the number of participants per cell (i.e. 12), all 12 randomizations of *pairs* were used. Every possible randomization of pairs thus occurred once in each cell. In other respects the procedure for experiment 2 was identical to that of experiment 1. Due to experimenter error the data on questionnaires after the rounds are not available.

6.2. *Results*

As in experiment 1, all button presses that followed less than 500 ms after the previous press were discarded in order to calculate the number of actions needed. This resulted in less than 4% of all button presses being discarded. Figure 4 presents the raw data for experiment 2. As the distribution of data was skewed, data were transformed using the same transformation as in experiment 1 (Y = Log [X+ 0.1]).

Transformed data were submitted to a $2 \times 2 \times 4 \times 4$ repeated measures ANOVA with age (2) and model (2) as between subject factors and complexity (4) and round (4) as within subject factors. The ANOVA revealed main effects of age ($F(1,44) = 16.02$, $p < 0.0001$, old participants needing more actions than young ones), model ($F(1,44) = 40.56$, $p < 0.0001$), round ($F(3,42) = 44.71$, $p < 0.0001$) and complexity ($F(3,42) = 7.16$, $p < 0.005$). Neither the age $\times$complexity interaction ($p < 1$), nor the age $\times$model interaction ($F(1,44) = 2.80$ $p < 0.11$) were significant. Analysing the model group and the no model group separately, however, showed a main effect of age in the model group ($F(1,22) = 15.94$, $p < 0.005$), while the effect of age is not significant in the no model group ($F(1,22) = 2.74$, $p < 0.12$). The fact that the interaction between age and model is not significant is probably due to the variance still being rather large. This is supported by the finding that the age by model interaction is just significant when scores are averaged over rounds ($F(1,44) = 4.31$, $p < 0.05$) in a $2 \times 2 \times 4$ repeated measures ANOVA. The graphs also suggest that interactions between age and complexity and age and round show young participants with a model are less affected by complexity. However, the age $\times$model $\times$complexity interaction was not significant ($F(3,42) = 2.00$, $p < 0.13$). It is worth pointing out that for young participants with a model the deviation score increases from 0.25 for complexity level 1 to 1.17 for complexity level 2. For old subjects in the model group these figures are 1.21 and 5.60 respectively. Given the power of 0.48 for the age $\times$complexity interaction in the model group, this effect may well be significant in a larger sample. A similar story could be told for the age by round interaction. Thus, there is a suggestion that the age effect is more pronounced for earlier rounds and higher complexity levels, but these effects are not significant.

6.2.1. *Questionaires*: Due to experimenter error, the data concerning the knowledge scores after the rounds is not available. On the questionnaire, which was filled out at the pretest, young participants in the model group obtained an average score of 10.25; older participants scored 7.67 on average. Comparison in a simple t-test showed this difference to be significant ($t(22) = 2.66$, $p < 0.05$ in a two-tailed test). Given this difference, it cannot be discounted that older participants in the model group needed more actions simply because they learned less from the instructions. In order to test this possibility, an analysis was performed to test whether the main effect of age in the model group was still significant when using the score on the questionnaire as a covariate. The ANOVA with covariate still showed a significant main effect of age ($F(1,21) = 6.73$, $p < 0.05$), so the larger number of actions old particpants needed does not seem to be associated with their learning less from the instructions.

6.3. *Efficiency*

Proportions of irrelevant actions (depicted in table 5) were transformed to logits for analysis, using the same

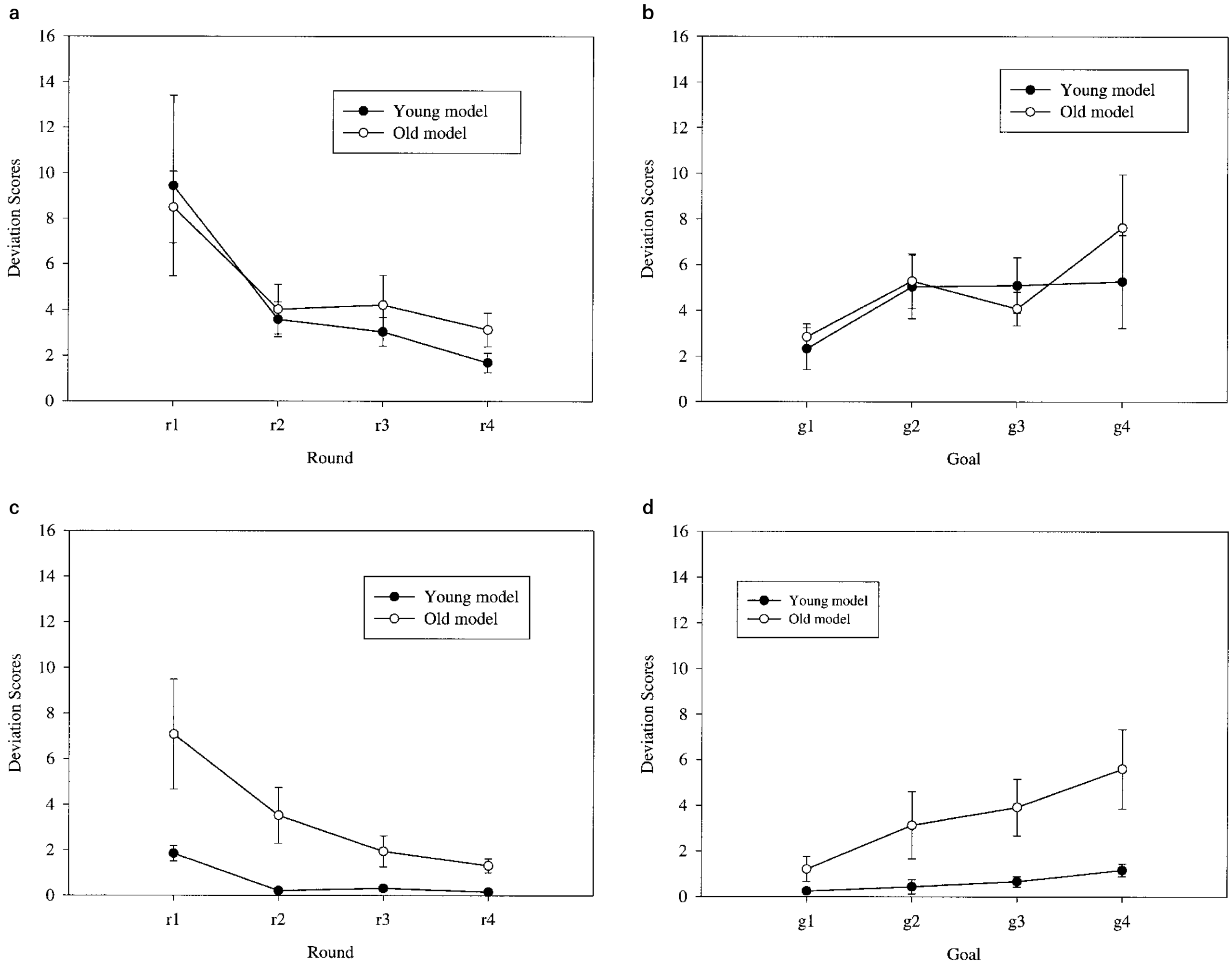

Figure 4. Deviation scores for experiment 2. (a) Deviation scores for the no model groups, averaged over goals. (b) Deviation scores for the no model groups, averaged over rounds. (c) Deviation scores for the model groups, averaged over goals. (d) Deviation scores for the model groups, averaged over rounds.

Table 5. Experiment 2: proportion of irrelevant actions (averaged over complexity level) as a function of age, presence of model and round (standard deviations are shown in parentheses).

Age	Condition	Round 1	Round 2	Round 3	Round 4
Young	No Model	0.31 (.12)	0.22 (.11)	0.22 (.13)	0.17 (.12)
	Model	0.13 (.13)	0.03 (.05)	0.02 (.04)	0.01 (.04)
Old	No Model	0.41 (.14)	0.29 (.11)	0.30 (.11)	0.28 (.09)
	Model	0.29 (.19)	0.17 (.14)	0.22 (.19)	0.15 (.13)

transformation as in experiment 1. Analysis of proportions of irrelevant actions showed main effects of age ($F(1,44) = 27.50$, $p < 0.001$, old participants have a higher proportion of irrelevant actions than young ones), model ($F(1,44) = 39.88$, $p < 0.001$), and round ($F(3,42) = 14.33$, $p < 0.001$). The analysis also revealed a significant age × model interaction ($F(1,44) = 4.48$, $p < 0.05$). Separate analysis of the model and the no model groups revealed that the effect of age was more pronounced in the model group ($F(1,22 = 23.10$, $p < 0.001$), than in the no-model group ($F(1,22) = 5.91$, $p < 0.05$).

6.4. *Discussion*

Comparing the results of experiments 1 and 2, participants in experiment 2 on average needed fewer actions for the three lowest complexity levels (which were identical to the ones in experiment 1), despite the fact that the device now had two more buttons. It thus appears that providing more informative labels made the task easier. The combination of the extra complexity level and the more informative labels resulted in the main effect of age becoming significant. Further analysis showed that this effect was due to young participants with foreknowledge outperforming older participants with foreknowledge, while young and old participants without foreknowledge did not differ. However, due to the high variance the age ×model interaction is significant only when averaging over rounds. Despite the fact that the device now had an additional complexity level, the age ×complexity interaction was again not significant. Regarding the efficiency of solutions, it was again found that participants in the model group performed a smaller proportion of irrelevant actions. Young participants perform a smaller proportion of irrelevant actions than old ones though, and the effect of age is more pronounced in the group with foreknowledge.

7. General discussion

The results of the two experiments show that both old and young adults performed better when they had foreknowledge. When such foreknowledge was lacking, old participants did not perform worse than young ones. The finding that young and old participants perform on the same level when they lack foreknowledge suggests that foreknowledge (or expertise) is a major influence in causing age related differences in learning to perform interactive tasks (after all, when young participants lack foreknowledge, they perform at the same level as older participants). As such, these results are consistent with previous studies claiming that computer experience is an important mediating variable in age differences in performance on tasks on interactive devices. Experiment 2 showed that when foreknowledge is present, young participants perform better than old ones. This finding is consistent with the production-deficiency hypothesis. Taken together, these findings point toward the importance of making foreknowledge available to older persons when they learn to use interactive equipment. It is also apparent, however, that old participants have more difficulty internalizing and applying this foreknowledge. Research on older people's learning and problem solving might point towards ways of optimizing this performance.

The age ×complexity interaction was not significant in either of the two experiments, though the results of experiment 2 did suggest such an interaction (as well as an age by round interaction). There is also a suggestion of an age ×complexity interaction between the experiments. The effect of age was significant in experiment 2, but not in experiment 1. Though a sampling effect cannot be discounted, the additional complexity level in experiment 2 seems a likely candidate for causing the significant age effect. The fact that the age ×complexity interaction is more apparent between experiments than within experiments suggests that in this type of (relatively difficult) task, overall complexity is more important than the complexity of separate procedures. With respect to the overall complexity, it should be noted that the addition of one complexity level resulted in the number of control elements being increased from six to eight. This relatively small increase may have posed a bigger selection problem for the old participants.

A final note concerns the absence of an age difference in performance in the no model group. Given the literature, this absence is surprising. It is possible, however, that young participants were using a sub-optimal strategy; such an effect was found by Denney *et al*. (1992). When they attempted to improve subjects' performance by stressing that subjects use optimal strategies, they found that only the young subjects improved. Trudel and Payne (1996), who studied discovery learning, provided one group of participants with a counter that recorded the number of actions they performed. Another group was not given this information. Trudel and Payne found that participants who were provided with the counter needed fewer actions than those without the counter to learn the same amount. Though Trudel and Payne employed young participants only, this effect may be differential with respect to age, as several researchers have argued that older people employ a more conservative criterion in speed-accuracy trade-offs.

In summary, results indicate that age differences in the performance of interactive tasks are most likely to be caused by differences in the amount of foreknowledge that old and young users possess. Moreover, young adults do not perform better than old adults when foreknowledge is absent. Participants with foreknowledge are better able to avoid irrelevant control elements. Only when the device is complex enough do performance differences between young and old participants with foreknowledge become apparent. This finding is consistent with the slower learning of elderly adults on 'training studies' of interactive devices. Taken together, the results indicate that both young and old users benefit from having foreknowledge, but these benefits seem to be larger for young users.

Regarding the design of interactive equipment, these findings suggest that it is important to specify the foreknowledge that is required to use a device and providing training or an advance organizer could be the answer. Designers should consider that older people would have more difficulty internalizing and applying this foreknowledge. A better strategy is to aim to have the relevant knowledge tie in with existing knowledge. Applying this general design principle is likely to benefit both older and younger users. The results from studies on schematic influences on memory suggest that these benefits will be larger for older users though. Since existing knowledge is likely to vary between people, the need to appeal to 'real world knowledge' appears particularly important in this case. Complexity appears to be an issue as well, though not as straightforward as was expected. Though there is a suggestion that older participants are more affected by the complexity of procedures than young participants, this effect was not significant. Moreover, the effects of complexity are more apparent between than within experiments; age effects become apparent with the addition of the extra complexity level in experiment 2. This finding suggests that designers should strive to limit the overall complexity of the device. Though this effect was not significant (and a little difficult to interpret given the use of the covariate), there is a suggestion that age differences are larger for the earlier rounds. This suggests that the results are especially important for applications that should be learned relatively quickly (e.g. applications in public places or 'walk-up-and-use interfaces').

References

Arbuckle, T. Y., Vanderleck, V. F., Harsany, M. and Lapidus, S. 1990, Adult age differences in relation to availability and accessibility of knowledge-based schemas. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16,** 305–315.

Bartlett, F. C. 1932, *Remembering*. (Cambridge: Cambridge University Press).

Birdi, K. S. and Zapf, D. 1997, Age differences in reactions to errors in computer-based work. *Behaviour and Information Technology*, **16,** 309–319.

Caplan, L. J. and Schooler, C. 1990, The effects of analogical training models and age on problem solving in a new domain. *Experimental Aging Research*, **16,** 151–154.

Cerella, J., Poon, L. W. and Williams, D. M. 1980, Age and the complexity hypothesis. In L. W. Poon (ed.) *Aging in the 1980's*. (Washington: American Psychological Association).

Charness, N. 1981, Aging and skilled problem solving. *Journal of Experimental Psychology: General*, **110,** 21–38.

Charness, N. and Campbell, J. I. D. 1988, Acquiring skill at mental calculation in adulthood: a task decomposition. *Journal of Experimental Psychology: General*, **117,** 115–129.

Charness, N., Schumann, C. E. and Boritz, G. M. 1992, Training older adults in word processing: effects of age, training technique and computer anxiety. *International Journal of Technology and Aging*, **5,** 79–106.

Craik, F. I. M. and Bosman, E. A. 1992, Age-related changes in memory and learning. In H. Bouma and J.A.M. Graafmans (eds) *Gerontechnology* (Amsterdam: IOS Press).

Czaja, S. J., Hammond, K., Blascovich, J. J. and Swede, H. 1989, Age-related differences in learning to use a text-editing system. *Behaviour and Information Technology*, **8,** 203–216.

Czaja, S. J. and Sharit, J. 1993, Age differences in the performance of computer-based work. *Psychology and Aging*, **8,** 59–67.

Denney, N. W., Lyons Tozier, T. and Schlotthauer, C. A. 1992, The effect of instructions on age differences in practical problem solving. *Journal of Gerontology*, **47,** P142–P145.

Elias, P. K., Elias, M. F., Robbins, M. A. and Gage, P. 1987, Acquisition of word-processing skills by younger, middle-age and older adults. *Psychology and Aging*, **2,** 340–348.

Halasz, F. G. and Moran, T. P. 1983, Mental models and problem solving in using a calculator. In *Proceedings of CHI'83 Human Factors in Computing Systems* (New York: ACM).

Hasher, L. and Zacks, R. T. 1979, Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, **108,** 356–388.

Hess, T. M. 1990, Aging and schematic influences on memory. In T.M. Hess (ed.) *Aging and Cognition: Knowledge Organization and Utilization* (Amsterdam: Elsevier Science Publishers).

Hess, T. M., Donley, J. and Vandermaas, M. O. 1989, Aging-related changes in the processing and retention of script information. *Experimental Aging Research*, **15,** 89–96.

Jennings, J. M. and Jacoby, L. L. 1993, Automatic versus intentional uses of memory: aging, attention and control. *Psychology and Aging*, **8,** 283–293.

Kausler, D. H. 1994, *Learning and Memory in Normal Aging* (San Diego: Academic Press).

Kieras, D. E. and Bovair, S. 1984, The role of a mental model in learning to operate a device. *Cognitive Science*, **8,** 255–273.

Morrow, D. and Leirer, V. 1997, Aging, pilot performance and expertise. In A.D. Fisk and W.A. Rogers (eds) *Handbook of Human Factors and the Older Adult* (San Diego, CA: Academic Press).

Morrrow, D. G., Leirer, V. O. and Altieri, P. A. 1992, Aging, expertise and narrative processing. *Psychology and Aging*, **7,** 376–388.

Morrrow, D. G., Leirer, V. O., Altieri, P. A. and Fitzsimmons, C. 1994, When expertise reduces age differences in performance. *Psychology and Aging*, **9,** 134–148.

McDowd, J. M and Craik, F. I. M. 1988, Effects of aging and task difficulty on divided attention performance. *Journal of Experimental Psychology: Human Perception and Performance*, **14,** 267–280.

Rankin, J. L. and Collins, M. 1985, Adult age differences in memory elaboration. *Journal of Gerontology*, **40,** 451–458.

Rogers, W. A., Cabrera, E. F., Walker, N, Gilbert, D. K. and Fisk, A. D. 1996a, A survey of automatic teller machine usage across the adult life span. *Human Factors*, **38,** 156–166.

Rogers, W. A., Fisk, A. D., Mead S. E., Walker, N. and Cabrera, E. F. 1996b, Training older adults to use automatic teller machines. *Human Factors*, **38,** 425 – 433.

Rumelhart, D. E. and Ortony, A. 1977, The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro and W. E. Montague (eds) *Schooling and the Acquisition of Knowledge* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Salthouse, T. A. 1992, Reasoning and spatial abilities. In F. I. M. Craik and T. A. Salthouse (eds) *The Handbook of Aging and Cognition* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E. and Babcock, R. L. 1989, Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15,** 507 – 516.

Smith, S. W., Rebok, G. W., Smith, W. R., Hall, S. E. and Alvin, M. 1983, Adult age differences in the use of story structure in delayed free recall. *Experimental Aging Research*, **9,** 191 – 195.

Stine, L. S. and Wingfield, A. 1987, Process and strategy in memory for speech among younger and older adults. *Psychology and Aging*, **2,** 272 – 279.

Titov, N. and Knight, R. G. 1997, Adult age differences in controlled and automatic memory processing. *Psychology and Aging*, **12,** 565 – 573.

Trudel, C. and Payne, S. J. 1996, Self-monitoring during exploration of an interactive device. *International Journal of Man – Machine Studies*, **45,** 723 – 747.

Zandri, E. and Charness, N. 1989, Training older and younger adults to use software. *Educational Gerontology*, **15,** 615 – 631.

Zelinski, E. M. and Miura, S. A. 1988, Context effects on script memory in young and old adults. *Psychology and Aging*, **3,** 292 – 299.