University of Zurich
Zurich Open Repository and Archive

Winterthurerstr. 190
CH-8057 Zurich
http://www.zora.uzh.ch

Year: 2009

Age Differences in the Underconfidence-With-Practice Effect

Rast, P; Zimprich, D

Age Differences in the Underconfidence-With-Practice Effect

Abstract

In two verbal learning experiments, the authors examined the accuracy of memory monitoring and the underconfidence-with-practice (UWP) effect in younger and older adults. Memory monitoring was operationalized as judgements of learning (JOL). An open issue is whether UWP can also be found in older adults. In the first experiment, both younger and older adults overestimated their memory performance in the first trial, but the older group differed from the young group in the second trial. The JOLs given by older participants matched, on average, their recall performance. In fact, the UWP effect was not observed in any of several conditions in older participants. In the second experiment involving five study-test cycles and two age groups, the same basic pattern of results was present: Older adults did not show an UWP effect. These findings appear to fit into a framework of dual factors affecting JOLs, which posits that the magnitude of JOLs derives both from an anchoring point and from on-line monitoring of items.

Running head: Age differences in Immediate and Delayed JOLs

Age Differences in the Underconfidence-With-Practice Effect

Philippe Rast, Daniel Zimprich

Dept. of Gerontopsychology, University of Zurich

Abstract

In two verbal learning experiments, we examined the accuracy of memory monitoring and the underconfidence-with-practice (UWP) effect in younger and older adults. Memory monitoring was operationalized as judgements of learning (JOL). An open issue is whether UWP can also be found in older adults. In the first experiment, both younger and older adults overestimated their memory performance in the first trial but the older group differed from the young group in the second trial. The JOLs given by older participants matched, on average, the correctly recalled words. In fact, the UWP effect was not observed in any of several conditions in older participants. In the second experiment involving five study-test cycles and two age groups, the same basic pattern of results was present: Older adults did not show an UWP effect. Our findings appear to fit into a framework of dual factors affecting JOLs, which posits that the magnitude of JOLs derives both from an anchoring point and from on-line monitoring of items.

Age Differences in the Underconfidence-With-Practice Effect

A pertinent issue of research on metamemory is the degree to which individuals can accurately predict their memory performance. An accurate appraisal of one's own memorizing abilities seems desirable because, in subject to such an appraisal, more or less effort could be allocated to attain a certain level of mastery (Koriat, Sheffer, & Ma'ayan, 2002; Nelson & Dunlosky, 1991). That is, based on their metacognitive judgments, individuals might be able to use self-monitoring to more efficiently control and regulate their strategies for learning and retrieving information from memory (Schneider & Pressley, 1989). The importance of monitoring may gain even more weight when cognitive resources and memory performance is declining, as it is the case with older adults, because this requires an optimized allocation of memory resources. An accurate appraisal of one's own memory functioning may essentially facilitate the appropriate allocation of cognitive resources in order to achieve a desired level of mastery and to avoid unnecessary overlearning. Findings regarding the changes in memory and metacognitive monitoring across the lifespan evidence that even though memory performance is impaired by aging, monitoring appears to be spared from cognitive decline (Connor, Dunlosky, & Hertzog, 1997; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Shaw & Craik, 1989).

With respect to assessing one's own memorizing ability, two different kinds of predictions have frequently been elicited: *Global predictions*, in which people judge how many items of an entire study list they will subsequently recall, and *item-by-item predictions*, which requires people to predict the likelihood of subsequent recall separately for each item (Nelson, Dunlosky, Graf, & Narens, 1994; Perlmutter, 1978). These latter, so-called judgments of learning (JOLs), are typically elicited by presenting a set of paired associates, for example, Swahili cue words, and English translation equivalents as target words (e.g. Adha – Trouble). After the presentation of each word pair, participants are asked to give a JOL by judging the probability that they will remember the target word a few minutes later when prompted with the cue word.

Metamemory accuracy is typically conceptualized in two ways: One way to define accuracy refers to *relative accuracy or resolution* (Koriat et al., 2002; Nelson & Dunlosky, 1991), which is commonly indexed by an average within-participant gamma correlation between JOLs and actual memory performance (Nelson, 1984). By contrast to resolution, *absolute accuracy or calibration* pertains to the correspondence between mean JOLs and mean recall performance in a memory test (see Metcalfe, 1998). In both experiments described in the present paper, we will focus on absolute accuracy measures because we aimed at examining the underconfidence-with practice effect which will be described in some detail later and which is reported for this type of measure only (Koriat, 1997).

Findings from studies using relative and absolute accuracy indicate that the ability to predict one's own memory performance is moderate (Koriat, 1997; Koriat et al., 2002; Mazzoni & Nelson, 1995; Scheck, Meeter, & Nelson, 2004; Schneider, Visé, Lockl, & Nelson, 2000). These findings can be generalized across different age groups. A characteristic regarding older adults memory predictions, however, seems to be the finding that they "overpredict" their own memory performance (Bruce, Coyne, & Botwinick, 1982; Mazzoni & Nelson, 1995; Schneider et al., 2000). Connor et al. (1997), for example, compared the metamemory accuracy of younger and older adults. A mixed list of high- and low-association pairs was created to enable the evaluation of whether there were age differences in the sensitivity of predictions to the level of relatedness. All participants were able to differentiate high- and low-association pairs, producing lower mean JOLs for the latter. At the same time, older participants predicted, on average, higher levels of recall than younger adults, which led to a substantial overprediction of performance in the elderly. Although some researchers have found that older adults' predictions are calibrated (e.g., McDonald-Miszczak, Hunter, & Hultsch, 1994; Rebok & Balcerak, 1989), the predominant finding is that older adults overestimate their memory performance compared to younger adults (Bruce et al., 1982; Coyne, 1985; Devolder, Brigham, & Pressley, 1990; Murphy, Sanders, Gabriesheski, & Schmitt, 1981; Perlmutter, 1978; Rebok & Balcerak, 1989). Lovelace (1990) hypothesized that, in relation to younger adults, older persons are more prone to prediction errors, generally in the direction of overestimating memory performance because older adults may actually be expecting or demanding more from their memory system than younger adults are.

It is, however, possible to raise the relative and absolute accuracy of JOLs in young and old substantially by eliciting JOLs with a certain *delay* after the presentation of the paired associates. The first to report this effect were Nelson and Dunlosky (1991), who asked participants to memorize paired associates. After a given delay (between 10 and 33 items), respondents were prompted with the cue word and were asked to give a JOL. Both, the absolute and relative accuracy of these delayed JOLs were found to be superior as opposed to that of immediate JOLs. Possibly, when JOLs are delayed, participants rely more heavily on cues pertaining to the ease with which the target can be retrieved from memory and, hence, the accuracy of predicting the recall probability is enhanced, that is, a delayed JOL instantiates a first recall attempt which offers the respondent a first impression of the subjective item difficulty (Dunlosky & Nelson, 1994; Koriat, Ma'ayan, Sheffer, & Bjork, 2006). This delayed-JOLs effect was consistently found in a number of studies and to practically the same extent in younger and older adults (Connor et al., 1997; Dunlosky & Nelson, 1992; Koriat et al., 2006; Scheck et al., 2004). Note, however, that the delayed-JOLs effect has been qualified as reflecting a memory, rather than a metamemory effect because retreived items receive a boost in recall from spaced study (Kimball, Metcalfe, 203). Similarly, Spellman and Bjork (1992) argued that the delayed JOLs effect may represent a psycholigcal analog of the Heisenberg Uncertainty Principle where the very effort to provide a current JOL by making covert recall attempts alters the state of learning, enhancing the accuracy of the judgment.

Underconfidence-With-Practice effect

Another way to increase the familiarity with the stimulus material is to present the items in more than one learning and recall trial. Giving participants more than one learning occasion leads to an increase in the relative accuracy of judgments, that is, the resolution typically increases over the trial course (Koriat et al., 2002). This finding, however, contrasts with the observation that calibration seems to be impaired with practice. Koriat (1997) reported a discrepancy between JOLs and recall performance which arose with repeated presentation of the stimuli: In two studies participants memorized a list of paired associates in several learn test cycles and, following the study of each pair, provided JOLs. A comparison of the effects of practice on JOLs and actual memory performance, in terms of absolute accuracy, disclosed a pattern the author referred to as the *underconfidence-with-practice (UWP) effect*: While the recall performance increased from the first to the second learning occasion, the effects of practice did not lead to more accurate predictions. Instead of improving calibration, that is, the difference between JOL and actual recall, average JOLs for the second occasion became markedly lower (underconfident) than recall performance. Briefly, UWP thus refers to a loss in accuracy in calibration across practice trials. This is also true for delayed JOLs, in which a smaller, but still relevant UWP effect was reported. In a subsequent review of several studies requiring participants to give JOLs, the UWP effect proved to be robust against a number of experimental manipulations (see, e.g., Koriat et al., 2002). The UWP effect has also been replicated by other authors who investigated JOLs under different conditions (e.g., Meeter & Nelson, 2003; e.g., Serra & Dunlosky, 2005). Investigations of the UWP effect in older adults, however, have not been conducted up to date.

In the next section we briefly discuss three theoretical perspectives which are geared to explain the UWP effect: (1) The cue-utilization framework, (2) the anchoring-and-adjustment effect, and (3) the dual-factors hypothesis:

Koriat (1997) interpreted the UWP effect within a *cue-utilization framework*, which distinguishes three types of cues for JOLs: Intrinsic, extrinsic and mnemonic. Intrinsic cues involve characteristics of the study items pertaining to its perceived difficulty (e.g., degree of associative relatedness between the members of a pair). Extrinsic cues relate to the conditions under which stimuli are learned and to the encoding operations applied by the learner (e.g., the number of times an item was studied and the amount of time the item was presented). The third type of cues comprises internal, mnemonic indicators that signal to the participant the extent to which an item has been learned and will be recalled in the future (e.g., cue familiarity). According to the cue-utilization approach, in making JOLs participants do not monitor directly the strength of the memory trace of the item in question, but use a variety of cues that are generally predictive of subsequent memory performance. Because JOLs are based on inferences and heuristics, accuracy judgments depend on how the learner weights the importance of the cues for decision-making. Koriat (1997) argued that the UWP effect might arise due to an insufficient regard to the contribution of extrinsic factors relative to that of intrinsic factors. That is, the effect of list repetition, representing the extrinsic factor, might be underweighted in the formation of the JOL (Carroll, Nelson, & Kirwan, 1997; Koriat et al., 2002).

A different perspective was adopted by Scheck and Nelson (2005). They hypothesized that the UWP effect might be an instantiation of a more general *anchoring-and-adjustment effect*. Briefly, anchoring may be described as a pervasive judgment bias in which decision makers are systematically influenced by arbitrary starting points (Chapman & Johnson, 1999). After having examined earlier studies where participants were required to judge recall probabilities for paired associates (Connor et al., 1997; Richards & Nelson, 2004), Scheck and Nelson (2005) concluded that a possible anchor is located around JOLs of 30% to 50%. Other than the cue-utilization approach, Scheck and Nelson assumed JOLs to be mostly unaffected by item difficulty, that is, no matter how easy or difficult items are, people tend to locate JOLs between 30% and 50%. In order to test their assumptions, the authors conducted an experiment with two learning and recall trials using easy and difficult items. The JOLs indeed seemed to be influenced by an anchor: In the first trial, JOLs given for difficult items (recall was around 5%) were pulled towards the anchor of 30% and lead to markedly overconfident judgments whereas JOLs for easy items (recall around 25%) were less overconfident. In the second trial, only easy items (now recall was above 50%) displayed the UWP effect and JOLs for difficult items were practically identical with the recall level, at 30%. The authors concluded that the hypothesis that underconfidence arises in part from an anchoring mechanism receives support.

Another, recently presented approach is the *dual-factors hypothesis*, which bears on both the anchoring and the cue-utilization approach (Scheck et al., 2004). The hypothesis states that the magnitude of JOLs derives both from an anchor point and from the on-line monitoring of items. The magnitude of JOLs is expected to change according to item difficulty, but not to the same extent as the corresponding recall level. The authors compared the adjustment process with the notion of the regression-toward-the-mean. Hence, difficult items with recall levels lower than the anchor or easy items with recall levels higher than the anchor will elicit JOLs somewhere between the recall level and the anchoring point. Items with medium difficulty elicit JOLs with comparable magnitudes to the recall level. In fact, JOLs given immediately after the presentation of paired associates resulted in large anchoring effects (Scheck et al., 2004). Conversely, delayed JOLs changed directly with item difficulty and were minimally affected by the anchor. Hence, for immediate JOLs the results were consistent with the anchoring hypothesis. In turn, delayed JOLs seemed to relate more on monitoring processes which means that participants rely more on cues pertaining to the ease with which the target can be retrieved than on an anchor.

In sum, the three explanatory accounts for the UWP effect appear to be complementary rather than disjunctive, as they describe similar processes from different perspectives. While the anchoring process and the dual-factor hypothesis are mainly geared to explain how JOLs are generated without prior task-related knowledge or cues about actual recall level, the cue-utilization approach explains the formation of JOLs including feedback from prior learning occasions

The present study

Viewed from an aging perspective, the reported empirical data does not allow drawing strong conclusions about the UWP effect in older adults or about the impact of a psychological anchor on the formation of JOLs because the available data is mainly restricted to studies examining young adults. Hence, the presence of an UWP effect in absolute accuracy judgments remains to be tested in older adults, whereby, at the same time, we aimed at gauging the UWP effect. There is, however, evidence that the UWP effect observed in young persons is not necessarily replicable in older adults. Connor et al. (1997) pointed out that older persons tend to overestimate their memory performance to a greater extent than younger adults, which may also results from lower recall performance in older persons. Lower recall performance and larger overestimation might influence the UWP effect as well. The relation between the recall performance and the over- or underestimation can be varied by using easy or difficult word pairs (cf., Koriat et al. 2006; Scheck & Nelson, 2005). Note that the UWP effect can only result from JOLs which are markedly lower than the actual recall performance, that is, in order to instantiate the effect, participants are required to underestimate the benefit from additional learning occasions. Older adults, however, recall fewer words and learn less during repeated presentation compared to young adults (Kausler, 1994)

Thus, the overarching goal of the present study was to examine the UWP effect in older persons and provide further empirical data to accuracy judgments in older and younger adults. More specifically, we (I) compared absolute accuracy judgments elicited by younger and older adults for easy and difficult stimulus material, and (II) compared the effects of immediate versus delayed JOLs in both age groups. Finally, we (III) addressed the UWP effect in older adults across two and five study test cycles.

Experiment 1

Experiment 1 was designed to investigate differences in the calibration of immediate and delayed JOLs between young and older adults using easy and difficult word pairs across two trials.

Method

Participants, Design, and Items. Thirty-six young adults (M = 25.7 years, SD = 3.9) and thirty-six older adults (M = 65.8 years, SD = 4.3) from the city of Zurich participated in this study. The experiment was a 2 (JOL timing: Immediate vs. delayed) × 2 (trial: Trial 1 vs. trial 2) × 2 (difficulty: Difficult vs. easy items) × 2 (age group: Young vs. old participants) × 2 (measure: JOL vs. recall performance) design with age group as a between-subjects factor and JOL Timing, Trial, Difficulty, and Measure as within subjects factors. The difficulty of items was determined by combining a German cue word with another German target word (easy: Kitchen - Car) or a Turkish word with its German translation equivalent (difficult: Mesnet - Agency). Note that first the Turkish words were selected to control for word length and syllables and then the German equivalent was matched.

Apparatus and Procedure. The experiment was programmed and executed in Inquisit (Version 1.33) on a Dell personal computer running Microsoft Windows XP Pro system software. Stimuli were presented on a 17" LCD display set at 1024 x 768 pixels.

Participants saw, in the center of the screen, 18 German word pairs and 12 Turkish-German word pairs for 3.5 s each. To rule out position effects, in each presentation cycle items were randomized anew for each participant, within six blocks containing five items (three easy and two difficult items). Half of the easy items and half of the difficult items were randomly allocated to the immediate or to the delayed JOL condition but the process of randomization was manipulated such that words from the first third remained in that third for the second presentation. The same manipulation was applied on words which were presented in the last third. This allocation remained the same across trials. The procedure for the second trial was the same as it was in trial one and word pairs remained in the same JOL-timing condition (immediate vs. delayed). Immediate JOLs were given after the presentation of the cue word by asking participants to answer the following question: "With what probability will you remember the target word in about five minutes from now if you see the cue word? (0 = *will definitely not recall*, 20 = 20% probability of recalling the word, 40 … 100 = *will definitely recall.*"). In the delayed JOLs condition, participants were asked to answer the same query as given for immediate JOLs, but the cue word appeared with a delay of, on average, 45 s after the presentation of the word pair in question. All JOLs were self paced and the participants' response prompted the next item. Finally, a self-paced recall test was administered, where the cue word was presented and the participants were asked to recall the corresponding target word. After the recall test, the second trial started using the same items. The second presentation cycle was again completed with the self-paced recall test.

Results

Mean JOLs (dashed line) and mean recall levels (solid line) are depicted in Figure 1, where each JOL timing condition (immediate vs. delayed) paired with each item difficulty (easy vs. difficult) is plotted in four panels. Further, the two age groups are represented by gray (young) and black (old) lines. Over- and underconfidence, defined by the difference between JOLs and recall level, is shown in Figure 2 using the same scheme as in Figure 1. Overconfidence is represented by positive bars and underconfidence is represented by negative bars. In Table 1, means and standard deviations of JOLs and recalled words are reported.

Insert Table 1 about here

The main effects of trial (Easy × Immediate: $F(1,70) = 18.85$, $p < .01$, $\eta^2 = .21$; Difficult × Immediate: $F(1, 70) = 5.98$, $p < .02$, $\eta^2 = .08$; Easy × Delayed: $F(1,70) = 88.93$, $p < .01$, $\eta^2 = .56$; Difficult × Delayed: $F(1, 70) = 65.39$, $p < .01$, $\eta^2 = .48$) and age (Easy × Immediate: $F(1,70) = 11.34$, $p < .01$, $\eta^2 = .14$; Difficult × Immediate: $F(1, 70) = 11.86$, $p < .01$, $\eta^2 = .15$; Easy × Delayed: $F(1,70) = 28.56$, $p < .01$, $\eta^2 = .29$; Difficult × Delayed: $F(1, 70) = 24.63$, $p < .01$, $\eta^2 = .26$) were statistically significant in all four conditions, implying higher overall means in trial two, that is, higher means for JOLs and correctly recalled words in both groups compared to trial one, and higher means for younger participants. The main effect of measure was statistically significant in both immediate conditions (Easy × Immediate: $F(1,70) = 18.20$, $p < .01$, $\eta^2 = .21$; Difficult × Immediate: $F(1, 70) = 25.63$, $p < .01$, $\eta^2 = .27$) as well as in the Difficult × Delayed condition ($F(1, 70) = 5.64$, $p < .02$, $\eta^2 = .08$). The significant main effect of measure implies, on average, higher means of JOLs compared to the average number of correctly recalled words across both trials and both groups.

Insert Figure 1 about here

Relevant for the examination of the UWP effect are the interaction terms between Trial × Measure: A flatter or opposite slope for the JOLs, compared to the slope of correctly recalled words between the first and the second trial are a prerequisite for the UWP effect to occur. In fact, the two-way interaction effects of Trial × Measure were significant in all four conditions (Easy × Immediate: $F(1, 70) = 117.75$, $p < .01$, $\eta^2 = .63$; Difficult × Immediate: $F(1, 70) = 95.88$, $p < .01$, $\eta^2 = .58$; Easy × Delayed: $F(1, 70) = 50.88$, $p < .01$, $\eta^2 = .42$; Difficult × Delayed: $F(1, 70) = 44.47$, $p < .01$, $\eta^2 = .39$). As can bee seen from Figure 1, the slope of the recall performance across both trials was positive and larger compared to the slope of the mean JOLs. However, it remains to be tested if mean JOLs are significantly lower than mean recall performance. Further, the interactions of Trial × Age (Easy × Immediate: $F(1, 70) = 13.52$, $p < .01$, $\eta^2 = .16$; Difficult × Immediate: $F(1, 70) = 9.67$, $p < .01$, $\eta^2 = .12$; Easy × Delayed: $F(1, 70) = 22.51$, $p < .01$, $\eta^2 = .24$; Difficult × Delayed: $F(1, 70) = 26.96$, $p < .01$, $\eta^2 = .28$) were statistically significant in all four conditions, indicating that older participants showed lower increase in response levels between Trial 1 and Trial 2 compared to the young group. The two-way interaction of Measure × Age was only significant in the Easy × Immediate condition ($F(1, 70) = 4.56$, $p < .05$, $\eta^2 = .61$) implying that older participants showed a greater discrepancy between JOLs and recall level. Additionally, in the Difficult × Delayed condition the triple interaction of Trial × Measure × Age ($F(1, 70) = 4.46$, $p < .05$, $\eta^2 = .06$) reached statistical significance.

UWP. In order to test for the presence of the UWP effect, *t*-tests of the relevant effects were calculated, that is, means of the JOLs and recalled words from the second trial were compared. In Figure 2 the differences between mean JOLs and mean recall performance are depicted with the corresponding estimates of effect size.

In order to instantiate the UWP effect the mean of confidence judgments must be significantly lower than mean recall. In fact, in the Easy × Immediate condition, the UWP effect was found in the younger ($t(35) = -2.67$, $p < .05$, $\eta^2 = .17$, $\Delta_{\text{JOL-Recall}} = -10.9$), but not in the older group ($t(35) = 0.58$, $p > .05$, $\eta^2 = .01$, $\Delta_{\text{JOL-Recall}} = 2.6$) where, on average, JOLs almost matched mean recall performance. In the Difficult × Immediate condition, the mean of JOLs and recalled words did not differ significantly in either group and, hence, the UWP effect was not observed in this condition (see top right panel in Figure 2).

Insert Figure 2 about here

In the delayed condition only the young group underestimated the actually recalled level of the target words (Easy: $t(35) = -3.25$, $p < .01$, $\eta^2 = .23$, $\Delta_{\text{JOL-Recall}} = -8.3$; Difficult: $t(35) = -2.44$,

$p < .01$, $\eta^2 = .15$, $\Delta_{\text{JOL-Recall}} = -6.0$), the JOLs of the older persons were, on average, not significantly different from their mean recall performance (Easy: $t(35) = -1.47$, $p > .05$, $\eta^2 = .06$, $\Delta_{\text{JOL-Recall}} = -3.2$; Difficult: $t(35) = -0.09$, $p > .05$, $\eta^2 = .00$, $\Delta_{\text{JOL-Recall}} = -0.2$). Hence, the UWP effect was found for the young group only in the delayed condition.

Anchoring. In order to test for the presence of an anchor in forming JOLs during the first trial, we compared JOLs and recall performance from both age groups in all four conditions.

In the immediate condition young (Easy: $t(35) = 6.00$, $p < .01$, $\eta^2 = .50$, $\Delta_{\text{JOL-Recall}} = 21.5$; Difficult: $t(35) = 6.86$, $p < .01$, $\eta^2 = .57$, $\Delta_{\text{JOL-Recall}} = 22.7$) and old participants (Easy: $t(35) = 7.55$, $p < .01$, $\eta^2 = .62$, $\Delta_{\text{JOL-Recall}} = 29.3$; Difficult: $t(35) = 7.23$, $p < .01$, $\eta^2 = .60$, $\Delta_{\text{JOL-Recall}} = 29.1$) overestimated their recall performance significantly. Effect sizes were comparable across the immediate condition, implying that the difficulty of items did not substantially affect the discrepancy between both measures in the first trial, that is, JOLs elicited after the first trial seemed to be independent of item difficulty. In addition, both age groups showed comparable JOL levels (Easy: $t(70) = 0.47$, $p > .05$, $\eta^2 = .00$, $\Delta\text{JOL}_{\text{Young - Old}} = 2.04$; Difficult: $t(70) = 0.70$, $p > .05$, $\eta^2 = .01$, $\Delta\text{JOL}_{\text{Young - Old}} = 3.52$) but the recall performance was significantly lower in older participants (Easy: $t(70) = 2.53$, $p < .05$, $\eta^2 = .08$, $\Delta\text{Recall}_{\text{Young - Old}} = 9.88$; Difficult: $t(70) = 3.76$, $p < .01$, $\eta^2 = .17$, $\Delta\text{Recall}_{\text{Young - Old}} = 9.88$) implying more overconfidence in the older group. The results seemed to support the notion of an anchor determining largely the location of the JOLs in the first trial.

In the delayed condition, young (Easy: $t(35) = 2.51$, $p < .05$, $\eta^2 = .15$, $\Delta_{\text{JOL-Recall}} = 7.2$; Difficult: $t(35) = 4.81$, $p < .01$, $\eta^2 = .40$, $\Delta_{\text{JOL-Recall}} = 10.3$) and old participants (Easy: $t(35) = 4.31$, $p < .01$, $\eta^2 = .35$, $\Delta_{\text{JOL-Recall}} = 8.5$; Difficult: $t(35) = 5.09$, $p < .01$, $\eta^2 = .43$, $\Delta_{\text{JOL-Recall}} = 8.2$) significantly overestimated the mean of recalled words. Note, however, that effect sizes were considerably smaller than they were in the immediate conditions. Further, old and young differed in both, the JOL and recall levels. That is, in the easy (JOL: $t(70) = 3.25$, $p < .01$, $\eta^2 = .13$, $\Delta\text{JOL}_{\text{Young - Old}} = 11.67$; Recall: $t(70) = 3.77$, $p < .01$, $\eta^2 = .17$, $\Delta\text{Recall}_{\text{Young - Old}} = 12.96$) as well as in the difficult condition (JOL: $t(70) = 2.14$, $p < .05$, $\eta^2 = .06$, $\Delta\text{JOL}_{\text{Young - Old}} = 6.42$; Recall: $t(70) = 2.59$, $p < .05$, $\eta^2 = .09$, $\Delta\text{Recall}_{\text{Young - Old}} = 4.32$) older participants showed both smaller JOLs and lower recall levels than young participants. Hence, the anchor still appeared to be present, its strength, however, was largely attenuated by the delayed condition.

Discussion

Experiment 1 was designed to investigate and compare the UWP effect in younger and older adults. The young and the older group showed comparable *immediate* JOLs in the *first trial* within both difficulty levels, hence, analogous to the findings from Scheck and Nelson (2005), mean JOLs were within the range of a psychological anchor of 30% to 50% correct recall. At the same time, the recall level in both groups was lower than 30%, which lead to substantially overconfident JOLs. Note that older adults recalled markedly fewer items than young participants and, consequently, the predictions of the older participants were more overconfident than those of the young. These findings are in line with results from earlier studies where older adults accuracy judgments were more biased toward overconfidence than those of the younger adults (Connor et al., 1997; Murphy, Sanders, Gabriesheski, & Schmitt, 1981; Touron & Hertzog, 2004). In sum, the JOLs in the first trial across both groups seemed to be influenced by a psychological anchor because these judgments appeared to be unaffected by actual recall performance, that is, it appears that if the respondents have no information about item difficulty they seem to rely on an anchor when giving JOLs (Scheck & Nelson, 2005).

For *delayed* JOLs in the *first trial*, we expected JOLs to be better calibrated (Nelson & Dunlosky, 1991). In fact, participants were able to give more precise JOLs compared to the immediate condition. Accordingly, JOLs differed between both age groups and also across both difficulty conditions, that is, other than immediate JOLs, delayed JOLs were influenced by item difficulty. Even though JOLs were closer to the recall level, they still did not predict the recall probability correctly. Hence, delayed JOLs seemed to rely on monitoring processes, which delivered more precise informations about the probability of recalling a specific item, *and* on anchoring mechanisms, which upward-biased the judgements to some degree. These findings fit in well with the results from Scheck et al. (2004) who investigated the dual-factors hypothesis and reasoned that immediate JOLs are based on an anchor while delayed JOLs rely also on informations about item difficulty stemming from monitoring processes.

In the *second trial*, the accuracy of JOLs seemed to be boosted by the additional learning trial with JOLs being fairly close to the actually recalled words. The increasing influence of monitoring on the formation of *immediate* JOLs relative to that of an anchor most probably enhanced the accuracy of judgments. However, there was a difference regarding the pattern of accuracy judgments between both age groups: The young group displayed the UWP effect for easy word-pairs, as described by Koriat (1997), even though the effect size of the underconfident judgements was just a third compared to the overconfidence effect reported in the first trial. The older group, in turn, predicted the recall level correctly. Hence, items given in the second trial were judged by both groups more precisely but older participants seemed to be more successful in rating their recall performance accurately. In the *delayed* condition, a similar pattern was observed for older participants. The second learning trial increased the accuracy of JOLs, which then matched the actual recall level. The young group, in comparison, did not benefit to the same extent from the additional presentation. For both difficulty levels, the young displayed the UWP effect, that is, the overconfidence from Trial one turned into underconfidence in Trial two. This effect was greater for easy items than it was for difficult items. An apparent difference between both groups was the level of memory performance, also in terms of learning effects across both trials: Apart from the higher initial level, younger participants seemed to benefit more from the additional learning trial than older participants. As a consequence, older participants' performance did not exceed the level expected from their JOLs.

Thus, with respect to the UWP effect, the results from both age groups were disparate. Young participants showed an UWP effect in most conditions, except for immediate JOLs given for difficult word pairs. Overall, the UWP effect in the young group can be explained by anchoring *and* monitoring processes (Koriat et al., 2002; Scheck & Nelson, 2005). Results from the older group yielded a different pattern: Both age groups overestimated their performance in the first trial, but the older group differed from the young group mainly in the second trial, where JOLs given by older participants matched, on average, the correctly recalled words. In fact, the UWP effect was not observed in any of the four conditions in older participants. The complete absence of the UWP effect in the older age group appears remarkable when considering that the effect has been shown to withstand several manipulations and, thus, was thought to be very robust (see Koriat et al., 2002). Several reasons may have contributed to the non-occurrence of the effect in the older group: In the *immediate* condition, the initial overconfidence in older participants was larger than in the young group which, in turn, implies that older participants would have needed to downgrade their judgements in the second trial to a larger amount than young participants in order to underestimate their performance in the second trial. In fact, the older group showed an interaction effect between JOLs and recall performance and downgraded substantially their JOLs in the second trial, but probably the relatively flat learning trajectory prevented judgments from being underconfident. In the *delayed* condition, again, older adults were almost perfect at predicting their recall performance. As in the immediate condition they showed a very flat learning trajectory which may have contributed to the absence of the UWP effect. Another explanation bears on the distinction proposed by Koriat (1997) between a theory-based analytic inference and an experience-based nonanalytic heuristic in the formation of a JOL. He proposed that the improvement in the accuracy of the judgments with repeated practice derives specifically from a change in reliance on theory- to experience-based cues. This would suggest that older adults' judgments were less biased and relied more on an experience-based heuristic compared to younger adults. This view might be supported by studies showing that, in the face of limited memory capacities, older adults are successful in selecting items which they need to remember in a subsequent recall (Castel, Benjamin, Craik, & Watkins, 2002; Castel, Farb, Craik, 2007). Hence, the better calibrated judgments provided by older adults may also be a consequence of an efficient selection of recallable items.

Older and younger adults differed in the benefit in recall performance drawn from learning which resulted in lower mean recall for older adults. In light of the UWP effect poor performance in mean recall can be problematic: If recall performance is low, *under*confidence in JOLs can hardly be achieved. A straightforward approach to deal with the problem of low mean performance in recalled words in older participants would be to increase the number of learning occasions in order to raise the recall performance above the JOL level and finally elicit the UWP effect also in older adults. As a consequence, we conducted a second experiment by administering five, instead of two, learning trials.

Experiment 2

Experiment 2 can be seen as an expansion of Experiment 1. Hence, it was mainly designed to maintain the first two trials comparable with Experiment 1 but give participants the possibility to learn and recall the paired associates in five trials – instead of two. The primary aim was test the assumption made in Experiment 1 that the absence of the UWP effect in the older group stems from too few presentation cycles, i.e., older participants may need more than two learning trials until their recall level exceeds their JOLs. We expected the UWP effect in younger participants to be unaffected by extending the number of learning trials. In order to expand our view on JOLs we further focused on (dis-)similarities in younger and older adults' trajectories in JOLs, independent from JOLs being over- or underconfident. This trajectory based perspective might offer a new look on age-differences in JOLs.

The general hypotheses remain the same as in Experiment 1: If the recall level is lower than the anchor, overconfidence is expected in the first trial. For each consecutive trial, we expect an increasing approximation of JOLs and previous recall performance. In older adults, underconfidence, and the UWP effect, may result after the second trial.

Method

Participants, Design, and Items. Thirty-four young (M = 26.1 years, SD = 3.1) and thirty-four older persons (M = 68.6 years, SD = 3.6) participated in Experiment 2. The design of the study was basically the same as in Experiment 1 with the difference that participants were given five learning occasions instead of two. This lead to a 2 (JOL timing: Immediate vs. delayed) × 5 (trial: Trial 1 to trial 5) × 2 (difficulty: Difficult vs. easy items) × 2 (age group: Young vs. old participants) × 2 (measure: JOL vs. recall performance) design.

Apparatus and Procedure. The apparatus and the procedure used here were the same as in Experiment 1. To increase precision of measurement, the number of stimuli was doubled to 36 German word pairs and 24 German – Turkish word pairs. Hence, the randomization procedure was extended to five presentation cycles consisting of six blocks containing ten items (six easy and four difficult items).

Results

As in Study 1, a repeated measures ANOVA with the between-subject factor age (young vs. old group) was computed. Here, participants were given five trials which lead to a 5 × 2 × 2 (trial, measure, age group) design. The mean JOLs and the mean recall performance are shown in Figure 3 while the difference between JOLs and the recalled words is illustrated in Figure 4. We start with reporting main and interaction effects, afterwards we focus on the UWP effect and finally we revise the anchoring effect.

As can be seen from Figure 3, the mean JOLs and the average recall performance increased from trial to trial. The level of the mean responses (JOLs and recall performance) of the young group was, on average, higher as the one from the old group and, in addition, the main effect of measure was significant, because the participants overestimated their performance markedly in the first trial. Older participants judged, on average, their recall performance lower and remembered on average fewer words than the younger group did. Figure 3 also reveals that the greatest changes in level and slope of the mean JOLs and the mean recall performance occurred between trial one and trial two. The two main effects of trial (*Easy Immediate*: $F(4, 264) = 208.99$, $p < .01$, $\eta^2 = .76$; *Difficult Immediate*: $F(4, 264) = 193.60$, $p < .01$, $\eta^2 = .75$; *Easy Delayed*: $F(4, 264) = 242.74$, $p < .01$, $\eta^2 = .79$; *Difficult Delayed*: $F(4,264) = 202.73$, $p < .01$, $\eta^2 = .75$) and age (*Easy Immediate*: $F(1, 66) = 55.28$, $p < .01$, $\eta^2 = .46$; *Difficult Immediate*: F(1,

66) = 67.95, $p < .01$, $\eta^2 = .51$; *Easy Delayed*: $F(1, 66) = 82.31$, $p < .01$, $\eta^2 = .56$; *Difficult Delayed*: $F(1,66) = 108.50$, $p < .01$, $\eta^2 = .62$) were statistically significant in all four conditions, which lead to essentially the same conclusion as in Experiment 1: The means of recalled words and JOLs increased significantly over the five learning trials and young participants recalled more words and judged their recall performance higher than older participants did. Additionally the main effect of measure was statistically significant in the two immediate conditions (*Easy Immediate*: $F(1,66) = 14.08$, $p < .01$, $\eta^2 = .18$; *Difficult Immediate*: $F(1, 66) = 10.81$, $p < .01$, $\eta^2 = .14$). That is, the mean of recalled words was considerably higher compared to the mean of JOLs in the immediate but not in the delayed conditions (cf. Figure 3).

Insert Figure 3 about here

With respect to the interaction terms, in all four conditions the interaction of Trial × Age (*Easy Immediate*: $F(4, 264) = 27.66$, $p < .01$, $\eta^2 = .30$; *Difficult Immediate*: $F(4, 264) = 60.75$, $p < .01$, $\eta^2 = .48$; *Easy Delayed*: $F(4, 264) = 25.23$, $p < .01$, $\eta^2 = .28$; *Difficult Delayed*: $F(4, 264) = 61.23$, $p < .01$, $\eta^2 = .48$) and the UWP-relevant interaction term of Measure × Trial (*Easy Immediate*: $F(4, 264) = 53.31$, $p < .01$, $\eta^2 = .45$; *Difficult Immediate*: $F(4, 264) = 43.91$, $p < .01$, $\eta^2 = .40$; *Easy Delayed*: $F(4, 264) = 18.53$, $p < .01$, $\eta^2 = .22$; *Difficult Delayed*: $F(4, 264) = 12.46$, $p < .01$, $\eta^2 = .16$) reached statistical significance. Further, the interaction of Measure × Age was significant in both immediate conditions (Easy: $F(1, 66) = 11.10$, $p < .01$, $\eta^2 = .14$; Difficult: $F(1, 66) = 13.77$, $p < .01$, $\eta^2 = .17$) and in the *Difficult Delayed* condition ($F(1, 66) = 4.93$, $p < .05$, $\eta^2 = .07$).

Across all four conditions, effect sizes were largest after the first two trials, and then an attenuation of the effects occurred across the five trials. This is also a typical finding from research on verbal learning, that is, the return in recall performance from every additional trial is continually diminishing, leading to growth curves with an asymptotic trajectory (Zimprich, Rast, & Martin, in press). Compared to the *immediate* condition, however, the effect sizes of JOLs provided in the *delayed* condition were markedly smaller implying that the discrepancy between JOLs and recall performance was attenuated by the delayed judgments.

UWP. As reported above, the significant interaction between measure and trial indicated the possible presence of the UWP effect in the responses of both age groups. A visual inspection of Figure 4 reveals a strong resemblance to the results from Experiment 1, namely, the older group never displayed the UWP effect whereas the young group consistently underestimated its performance in the immediate conditions.

More precisely, in the *immediate* conditions the young group underestimated its recall performance significantly in the second trial (Easy: $t(33) = -2.27$, $p < .05$, $\eta^2 = .14$, $\Delta_{\text{JOL-Recall}} = 6.9$; Difficult: $t(33) = -2.10$, $p < .05$, $\eta^2 = .12$, $\Delta_{\text{JOL-Recall}} = 6.8$) which indicated the presence of the UWP effect. From trial three on, the mean JOLs in the easy condition were not statistically different from the mean of recalled words. In the difficult condition, in turn, underconfidence persisted throughout trial three ($t(33) = -4.14$, $p < .01$, $\eta^2 = .34$, $\Delta_{\text{JOL-Recall}} = 12.4$), four ($t(33) = -3.05$, $p < .01$, $\eta^2 = .22$, $\Delta_{\text{JOL-Recall}} = 7.4$) and five ($t(33) = -2.09$, $p < .05$, $\eta^2 = .12$, $\Delta_{\text{JOL-Recall}} = 4.5$). Accordingly, the criterion for the UWP effect was met. The older group started with a large and statistically significant overconfidence effect for easy words in the first trial which persisted in trial two ($t(33)= 2.51$, $p < .05$, $\eta^2 = .16$, $\Delta_{\text{JOL-Recall}} = 13.7$), three ($t(33)= 2.28$, $p < .05$, $\eta^2 = .14$, $\Delta_{\text{JOL-Recall}} = 11.4$), and five ($t(33)= 2.10$, $p < .05$, $\eta^2 = .12$, $\Delta_{\text{JOL-Recall}} = 8.7$). Similarly, in the difficult condition older adults overestimated their performance significantly from the first to the fourth trial on (Trial 2: $t(33) = 3.13$, $p < .01$, $\eta^2 = .23$, $\Delta_{\text{JOL-Recall}} = 18.9$; Trial 3: $t(33) = 2.59$, $p < .05$, $\eta^2 = .17$, $\Delta_{\text{JOL-Recall}} = 15.3$; Trial 4: $t(33) = 2.43$, $p < .05$, $\eta^2 = .15$, $\Delta_{\text{JOL-Recall}} = 13.6$). The effect sizes in the immediate condition emphasized this finding with large effects for the first trial and smaller effect sizes for trials two to five. Hence, due to overconfident judgments in most of the trials of the immediate conditions, older adults did not display the UWP effect.

In the *delayed* conditions the results were more ambiguous with respect to the UWP effect. As can be seen from Figure 4 (lower panels), after the initial overconfidence both groups improved calibration of JOLs, that is, the UWP effect was not present at the second trial. Again, older adults did not underestimate their performance and, hence, the UWP effect was not observed for the older group. The JOLs of the young group, however, were underconfident in the third trial ($t(33) = -2.42$, $p < .05$, $\eta^2 = .15$, $\Delta_{\text{JOL-Recall}} = -4.4$) of the easy condition and in the fourth trial of the difficult condition ($t(33) = -2.18$, $p < .05$, $\eta^2 = .13$, $\Delta_{\text{JOL-Recall}} = -6.2$). These results are less evident with respect to the UWP effect because the underconfidence did not occur in the second trial. Still, if all five trials are taken into consideration, the UWP effect can be observed in the young group.

Insert Figure 3 about here

Anchoring. The presence of an anchor was indicated by the overconfidence in the first trial which was present in all four conditions in both age groups.[1] Similar to Experiment 1 the largest overconfidence arised in both immediate, easy (Young: $t(33) = 4.73$, $p < .01$, $\eta^2 = .40$, $\Delta_{\text{JOL-Recall}} = 19.2$; Old: $t(33) = 8.52$, $p < .01$, $\eta^2 = .69$, $\Delta_{\text{JOL-Recall}} = 40.8$) and difficult (Young: $t(33)$

= 5.76, $p < .01$, $\eta^2 = .50$, $\Delta_{\text{JOL-Recall}} = 27.5$; Old: $t(33) = 7.07$, $p < .01$, $\eta^2 = .60$, $\Delta_{\text{JOL-Recall}} = 37.6$) conditions. Accordingly, effect sizes of the discrepancy between JOLs and recall performance were largest in the first trial. The mean JOL levels in the immediate condition were almost identical in both age groups (Easy: $t(66) = -0.14$, $p > .05$, $\eta^2 = .00$, $\Delta\text{JOL}_{\text{Young - Old}} = -0.82$; Difficult: $t(66) = -0.75$, $p > .05$, $\eta^2 = .01$, $\Delta\text{JOL}_{\text{Young - Old}} = -4.95$) at the same time, older adults recalled significantly fewer words than younger participants (Easy: $t(66) = 6.46$, $p < .01$, $\eta^2 = .39$, $\Delta\text{Recall}_{\text{Young - Old}} = 20.78$; Difficult: $t(66) = 3.82$, $p < .01$, $\eta^2 = .18$, $\Delta\text{Recall}_{\text{Young - Old}} = 7.35$). These results corroborated the findings from Experiment 1 where older adults overestimated their performance to a larger degree and supported the notion of a psychological anchor which biases responses in the first trial.

In the delayed conditions, JOLs were better calibrated in the first trial which resulted in smaller effect sizes compared to the immediate conditions (cf. Figure 4). Still, the young (Easy: $t(33) = 2.85$, $p < .01$, $\eta^2 = .20$, $\Delta_{\text{JOL-Recall}} = 10.3$; Delayed: $t(33) = 2.68$, $p < .05$, $\eta^2 = .18$, $\Delta_{\text{JOL-Recall}} = 8.8$) and the old group (Easy: $t(33) = 4.18$, $p < .01$, $\eta^2 = .35$, $\Delta_{\text{JOL-Recall}} = 17.5$; Delayed: $t(33) = 3.92$, $p < .01$, $\eta^2 = .32$, $\Delta_{\text{JOL-Recall}} = 14.4$) overestimated markedly their recall performance in the first trial. In the *easy* condition the age groups differed both, in mean JOLs ($t(66) = 2.44$, $p < .05$, $\eta^2 = .08$, $\Delta\text{JOL}_{\text{Young - Old}} = 12.84$) and in mean recall ($t(66) = 5.85$, $p < .01$, $\eta^2 = .34$, $\Delta\text{Recall}_{\text{Young - Old}} = 20.10$), indicating that older adults JOLs and recall performance was significantly lower than those from the young group. These findings replicated largely the results from Experiment 1 which signifies that, besides anchoring, monitoring processes are also involved in the formation of delayed JOLs. In the *difficult* condition, however, the JOLs did not differ between the age groups ($t(66) = 0.48$, $p > .05$, $\eta^2 = .00$, $\Delta\text{JOL}_{\text{Young - Old}} = 2.30$) although older adults recalled significantly fewer words compared to the young group ($t(66) = 4.14$, $p < .01$, $\eta^2 = .21$,

$\Delta\text{Recall}_{\text{Young - Old}} = 7.84$). The young group, however, appeared to react more sensitively on the increased difficulty of items by providing lower JOLs compared to the old group.

Discussion

Experiment 2 was designed to investigate and compare the trajectories of JOLs and recall performance in two age groups across five learning occasions. In Experiment 1 we hypothesized that older adults would need more than two learning trials to achieve a recall level that might instantiate the UWP effect, hence, we required participants to complete five learning occasions. The difficulty for both types of items (easy and difficult) was comparable to the difficulty in Experiment 1. Note, however, that the amount of items in Experiment 2 was doubled to a total of 36 easy and 24 difficult word pairs to avoid ceiling effects.

As in Experiment 1, in the *immediate* condition both age groups started out at almost identical JOL levels, while their average recall performance was very different which lead to substantial overconfidence in JOLs. In the first trial the mean JOLs were within 30% to 50% predicted recall, which replicated findings from Experiment 1 and from earlier studies (Scheck & Nelson, 2005) and substantiated the anchoring hypothesis that the magnitude of mean JOLs in the first trial depends largely on a psychological anchor if no prior information about the item difficulty is available. At the same time, the larger overconfidence in JOLs of older adults probably stemmed from the anchoring mechanism, that is, if young and old participants judge items almost independently of their difficulty, this leads to larger overconfidence in older adults simply because their recall level is lower.

In the first trial of the *delayed* condition we also expected to find results similar to Experiment 1. In fact, delayed JOLs were generally more calibrated compared to the immediate condition. Accordingly, JOLs given for easy items by older adults were smaller than JOLs from young participants. This again substantiated the notion of enhanced accuracy in delayed judgments, probably due to monitoring processes, which was also reported in Experiment 1 and in earlier studies (Koriat, 1997; Nelson & Dunlosky, 1991). An exception seemed to be JOLs given for difficult items, where both age groups had comparable levels.

Regarding the UWP effect, in the young group the results from Experiment 1 were only in part replicated. Participants showed the UWP effect in the Easy × Immediate condition but, contrary to the previous experiment, they also underestimated the difficult words in the Difficult × Immediate condition, which corroborates findings from Scheck and Nelson (2005). In the delayed condition, results from the first two trials did not elicit the UWP effect neither for easy nor for difficult word pairs. Hence, when considering only the first two trials, the UWP effect was found in the immediate but not in the delayed condition. However, if all five trials are taken into consideration, underconfidence can also be observed for word pairs in the delayed condition (see Figure 4, Easy × Delayed and Difficult × Delayed). Note that young participants were overconfident only in the first trial. From the second trial on, the mean JOLs remained underconfident or correct throughout consecutive trials.

In the old group, JOLs elicited in the last three trials still did not underscore the recall level and, hence, the UWP effect could not be found: As in Experiment 1, the older group did not show the UWP effect in any of the four conditions and in none of the five learning trials. Our assumptions that older adults would need more trials to display the UWP effect could not be verified in Experiment 2, in this respect, adding three trials did not lead to the expected underconfident judgments in later trials. For immediate JOLs, older participants overestimated their recall performance in most of the trials. Apart from the initial and large overconfidence, in the consecutive four trials older participants seemed to overrate their performance by a rather stable amount. When giving delayed judgments, the estimated recall probability was generally closer to the actual recall performance compared to immediate JOLs. Contrary to the young group, the older participants never underestimated their recall level. Even at the second trial, where the UWP effect typically is found, older adults still overestimated or judged correctly their recall level. Note, however, that the recall performance in the delayed × difficult condition was very low in the first trials. Hence, overconfidence might also be in part due to a floor effect in recall performance, that is, recall levels lower than 10% almost inevitably led to overconfident judgments, because participants hardly downgraded their judgments lower than 10%.

Given that the presence or absence of the UWP effect is also a matter of definition it is worth pointing out that Koriat and colleagues (2006) argued for a more liberal understanding of the effect. That is, underconfidence should not be a prerequisite for the UWP effect anymore. On p. 606 the authors state: “This interactive pattern is most often instantiated in the form of a UWP pattern in which initial overconfidence gives way to subsequent underconfidence. Sometimes, however, the underconfidence segment may be missing…” Hence, in terms of this liberal definition it is sufficient for the instantiation of the UWP effect if the slope of JOLs is increasing with a lower rate as the slope of the learning performance – independent of its absolute relation to the recall performance. In view of this definition older adults would have displayed the UWP effect as well in Experiment 1 and 2 (see e.g., Figure 3). By loosening the requirements for the UWP effect, however, a paradoxical situation may occur: If initially overconfident judgments are, due to effects of practice, correctly calibrated to a lower degree of overconfidence they can still be indicative of underconfidence with practice (e.g., see Figure 4 top panels and lower right panel). From another perspective, however, one might simply state that calibration increased with practice. In fact, older adults JOLs remained overconfident across most of the trials in the immediate condition. Given this unambiguity, we preferred to remained with the definition of the UWP effect given by Koriat et al. (2002).

To sum up, in both groups the greatest changes in JOLs took place in the first two trials. Independent of item difficulty or timing of the JOL, the overconfidence effect was always largest in the first trial, followed by a marked decrease in the second trial. This corroborates the notion of a strong influence of an anchor in the first trial which decreases in favor of monitoring in consecutive trials. In the second trial, young participants downgraded JOLs to underconfidence but older adults adjusted JOLs to be slightly overconfident or correct. In the consecutive three trials the trajectories of both measures (JOLs vs. correct recall performance) appeared to converge, that is, the absolute accuracy increased across all trials. Note that the shape of the subjective accuracy trajectories was comparable across both age groups. Only in relation to the objective learning trajectory the pattern of over- and underconfidence differed across groups, that is, the tendency of younger adults to remain underconfident and of older adults to remain overconfident from the second trial on persisted throughout almost all conditions.

General Discussion

The UWP effect has been examined exclusively in young populations (see, e.g., Koriat et al., 2002) and, hence, it remained an open issue whether this effect could be replicated in older adults. As has been hypothesized by Scheck and Nelson (2005), part of the UWP effect might be due to the impact of anchoring on the formation of JOLs. First of all, our results seem to support the notion of a psychological anchor which determines JOLs at the first trial. The anchoring effect appeared to be very pronounced when no prior information was available, as it was the case for immediate JOLs: Both age groups rated the probability of recalling the same items at almost identical levels. Simultaneously, the difficulty of items was very different for both groups as young recalled up to four times more items than older participants (as seen in Experiment 2 with easy words after the first trial). We interpret this in favour of an anchoring mechanism which determined largely the location of JOLs – almost independent of item difficulty. Note that the higher accuracy of delayed JOLs is not contradictory to this view. When judgments are delayed, JOLs are, presumably, based on a retrieval attempt, which provides a mnemonic cue that might be utilized in forming a better calibrated JOL (Nelson & Dunlosky, 1991). A similar mechanism leads to increasingly accurate judgments when participants are given more than one trial, that is, the monitoring process following the first trial reduces the bias toward the anchor to a considerable extent and results in an increase in absolute accuracy (Koriat et al., 2002; Nelson & Dunlosky, 1991).

Earlier studies have demonstrated that repeated practice leads to JOL levels which typically fall below the level of memory performance in the second trial and consequently instantiate the UWP effect. In fact, in these studies the UWP effect appeared to be very robust, at least against several experimental manipulations, and we therefore expected it to appear in older adults as well (for a summary, see Koriat et al., 2002). The most striking finding from both our experiments was the non-appearance of the UWP effect in older adults at all, while the results in the younger group replicated earlier findings (Scheck & Nelson, 2005). As both groups received the same stimulus material and the same procedure, it seemed unlikely that methodological differences were responsible for the unexpected results in the older group. The presented paired-associates, however, appeared to be much more difficult for older than for younger participants. Hence, we concluded that older adults did not display the UWP effect as found in young people.

Several factors may be responsible for the non-occurrence of the effect in older adults: As mentioned earlier, throughout both experiments and across all learning trials younger adults remembered more items than older adults implying that for older adults the same items were more difficult. This is not an unusual finding in laboratory test situations when paired associates are presented in non-self paced learning trials. If older and younger adults are compared in their recall performance, young participants generally recall more words than older participants and, moreover, they tend to learn faster than older adults (Kausler, 1994). This was also observed in both our experiments where younger adults recalled more items and learned faster which resulted in steeper learning trajectories compared to older participants' performance. Even though the stimulus was the same for both age groups, the difficulty appeared to be highly elevated for older adults. The older group started at very low recall levels and did not benefit much from additional learning trials. Given these circumstances, the preconditions for the UWP effect were not exactly the same for young and old participants: Equal JOL levels across both groups but lower recall levels in the older lead to a larger overconfidence effect in the old group after the first trial. Further, young participants benefited more from additional learning trials than older, resulting in steeper learning trajectories. Note that the UWP effect requires an interaction between the average JOL and recall trajectories additionally, JOLs in the second trial need to be significantly underconfident (but see Koriat et al., 2006), which, altogether, increases the demands with respect to the UWP effect for older adults: Older adults would have been required to adapt their JOLs in the second trial to a greater extent compared to younger adults in order to elicit underconfident judgments. In fact, older adults adjusted their JOLs, from the first to the second trial, to a greater extent than younger, but still not enough to fall below the recall level. The question remains why older adults did not adapt their JOLs to a larger portion? There might be two explanations: One is that JOLs can not be adapted arbitrarily away from a given level but only to a certain degree due to a persistence or inertia of JOLs. That is, if the average JOL in the first trial was at 40% one will not simply downgrade it by 35 points to 5% but, for example, maximally by 15 points to 25%. Hence, the adaptation of JOLs from the first to the second trial may have tapped the full range in older adults' adjustment possibilities. Even with the maximal downward adaptation of JOLs, this was not enough to elicit the UWP effect because recall was still lower. In addition, older adults' performance appeared to be limited by a floor effect across, at least, the first two trials which made underconfidence almost impossible. The second explanation bears on earlier findings where older adults tended to generally overestimate their cognitive performance (Connor et al., 1997; Murphy et al., 1981; Touron & Hertzog, 2004). If this was the case for JOLs in both of our studies, the instantiation of the UWP effect must have been additionally impeded simply because the required underconfident judgements are all biased with the general tendency to overestimate one's own performance and, finally, lead to higher average JOLs.

Instead of just focusing on the presence or absence of the UWP effect in older adults, our experiments have shown that it might be also fruitful to concentrate on (dis-)similarities in younger and older adults JOLs across all learning trials in order to learn more about monitoring. The JOL trajectories in older and younger adults were fairly similar in their shape: Young and old adults both showed the largest adaptation in JOLs from the first to the second trial, which was also underlined by significant interaction terms of Trial × Measure. From the second trial on, the adjustment of JOLs was much smaller but still lead to increasingly calibrated judgments. By and large, the adaptation process in JOLs was comparable in young and old participants. This finding also substantiates Koriat et al.'s (2006) suggestion to uncouple the UWP effect from the recall level. Furthermore both groups started out, in the immediate condition, at almost the same level which suggests that the anchor is not underlying much change in two very different cohorts.

One might speculate if younger adults would have shown an UWP effect if their learning performance had been as low as in older adults, but the basic process of JOLs adjustment over a number of trials appeared to be basically the same in the young and the old group. On the other hand, the relation between JOLs and recall performance from trials two to five was very different between both groups. As the young underestimated their performance, older adults systematically overestimated their recall level, especially when JOLs were given immediately. These results are similar to findings from global memory predictions where older adults tend to overestimate their memory performance as well. Murphy, Sanders, Gabriesheski, and Schmitt (1981), for example, had younger and older adults estimate their memory span for the number of common objects that they thought they could remember, followed by a recall task in which this span was actually measured. They found that younger adults tended to underestimate their memory span, whereas older adults tended to overestimate their memory span. Note, however, that the difference between both age groups regarding over- or underconfidence can not be interpreted unambiguously as long as the recall performance, and eventually the difficulty of items are different in young and old (cf. Hertzog et al., 2002). Given that older adults' performance was influenced by a floor effect, JOLs must have been necessarily correct or overconfident.

In view of the three explanatory approaches described in the introduction, the dual-factor hypothesis (Scheck et al., 2004) appeared to best catch the interplay between a strong anchor in the first trial and increasing importance of monitoring throughout the following trials. The higher accuracy for JOLs in the delayed condition can be deducted by the dual factors hypothesis as well: Delayed presentation of the cue word initiates a first retrieval attempt which delivers valuable information on the probability of recalling the item later. Consequently, monitoring outweighs the arbitrary anchor in the formation of the delayed JOL and leads to greater accuracy. Note, however, that in all three approaches, in the cue-utilization framework, in the anchoring hypothesis, and in the dual-factors hypothesis, the weighing of different cues is crucial for the outcome of a response. Hence, if one considers the anchor to represent an internal cue which may be used in absence of any other relevant cue of item difficulty, the anchoring hypothesis and the dual-task hypothesis may be recast in terms of the cue-utilization approach. Recently, Finn and Metcalfe (2007) provided an additional explanation for the UWP effect in immediate JOLs. In two experiments the authors examined the hypothesis that the Memory for Past Test (MPT) heuristic influences the magnitude of JOLs after a first learning and recall cycle, that is, if they remember getting the item right after the first recall, they will give it a high JOL – if not, they will provide a low JOL. Hence, underconfidence should be mainly reported for items which were not recalled in the previous test. In fact, both experiments provided evidence for the MPT heuristic, which makes our findings even more puzzling regarding the large amount of unrecalled items in older adults. Apparently, most unrecalled items still received high JOLs which stands at odds with the MPT heuristic.

An issue which remained untouched in this paper concerns the relation between JOLs and concomitant variables such as, for example, working memory span (e.g., De Bruin, Rikers, Schmidt, 2005) or processing speed (cf., Nelson & Leonesio, 1988, Salthouse, 1996) which may account for a considerable amount of variance in recall and in JOLs. Especially in the light of the MPT heuristic the working memory span may determine largely the capacity to remember correctly how one performed on individual items in the past test (cf., Finn & Metcalfe, 2007). Given that older adults, on average, did not underestimate their performance, future research on older adults will have to include explanatory variables in order to clarify the origin of these age effects.

In the introduction to this paper we emphasized the importance of monitoring memory, especially in old age, and argued that monitoring is spared from cognitive decline. In fact, the process of memory monitoring did not seem to be different from monitoring in the younger group except for its tendency to be overconfident. This, however, puts into perspective the advantage of an intact monitoring in older persons. If one overestimates his own memory performance, the effort invested in future learning trials is probably smaller compared to someone who underestimates his or her performance. Unfortunately, the older group with low recall performance and small learning rates overestimated their performance and, maybe, this contributed to even lower recall levels. If monitoring is to be used in future, as an intact resource for memory enhancement, older adults may benefit from it when their judgments become underconfident. Hence, the presence of the UWP effect could be seen as an indicator of an intact and self-propelling memory system.

References

Bruce, E. R., Coyne, A. C., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology, 37*, 354-357.

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and on long-term retention. *Acta Psychologica, 95*, 239-253.

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition, 30,* 1078-1085.

Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition, 35,* 689-700.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes, 79*, 115-153.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metameory accuracy. *Psychology and Aging, 12*, 50-71.

Coyne, A. C. (1985). Adult age, presentation time, and memory performance. *Experimental Aging Research, 11*, 147-149.

De Bruin, A. B., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology, 19,* 167-181.

Devolder, E. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging, 5*, 291-303.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (jols) and the delayed-jol effect. *Memory & Cognition, 20*, 373-380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (jols) to the effects of various study activities depend on when the jols occur? *Journal of Memory and Language, 33*, 545 -565.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*, 209-225.

Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego: Academic Press, Inc.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 595-608.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147-162.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-invain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676-686.

Lovelace, E. A. (1990). Aging and metacognition concerning memory function. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness and interventions* (pp. 157-187). Amsterdam: North-Holland.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1263-1274.

McDonald-Miszczak, L., Hunter, M. A., & Hultsch, D. E. (1994). Adult age differences in predicting memory performance: The effects of normative information and task experience. *Canadian Journal of Experimental Psychology, 48*, 95-118.

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113*, 123-132.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2*, 100-110.

Murphy, M. D., Sanders, R. E., Gabriesheski, A. S., & Schmitt, F. A. (1981). Metamemory in the aged. *Journal of Gerontology, 36*, 185-193.

Murphy, M. D., Sanders, R. E., Gabriesheski, A. S., & Schmitt, F. A. (1981). Metamemory in the aged. *Journal of Gerontology, 36*, 185-193.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (jols) are extremely accurate at predicting subsequent recall: The "delayed-jol effect." *Psychological Science, 2*, 267-270.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*, 207-213.

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology, 14*, 330-345.

Rebok, G. W., & Balcerak, L. J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology, 25*, 714-721.

Richards, R. M., & Nelson, T. O. (2004). Effect of the difficulty of prior items on the magnitude of judgment of learning for subsequent items. *American Journal of Psychology, 117*, 81-91.

Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language, 51*, 71-79.

Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology: General, 134*, 124-128.

Schneider, W., & Pressley, M. C. (1989). *Memory development between 2 and 20*. New York: Springer-Verlag.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring. Evidence from a judgment-of-learning (jol) task. *Cognitive Development, 15*.

Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1258-1266.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4*, 131-135.

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19*, 452-466.

Zimprich, D., Rast, P., & Martin, M. (in press). Individual differences in verbal learning in old age. In S. M. Hofer & D. F. Alwin (Eds.), *The handbook of cognitive aging: Interdisciplinary perspectives*. Thousand Oaks: Sage Publications.

Authors' Note

Philippe Rast & Daniel Zimprich, Gerontopsychology, Department of Psychology, University of Zurich, Switzerland.

The research described in the present paper was supported by a grant of the University of Zurich (Forschungskredit 2004) given to the second author.

Correspondence concerning this article should be addressed to

Philippe Rast

Gerontopsychology

Department of Psychology

University of Zurich

Binzmühlestr. 14/24

CH-8050 Zurich

p.rast@psychologie.uzh.ch

Tel. +41 44 6357420

Footnotes

[1] In order to investigate the influence of the serial position on the magnitude of the respective JOLs we calculated linear regressions across 12 or 18 items and correlations between the serial postion of the JOL and the judged recall probability in the first trial of each difficulty × timing condition. The correlations for JOLs given immediately were both negative and statistically significant ($r = -.12$, $p < .01$). The unstandardized regression weights were negative and statistically significant as well (-0.685, $p < .01$, in the easy, and -0.929, $p < .05$, in the difficult condition) indicating that with each additional item the estimate of a JOL was reduced by 0.7% in the easy, and 0.9% in the difficult condition, which corroborates the hypothesis of an anchoring effect. The estimates were not different for both age groups. In terms of effect sizes, however, the correlation and the estimate from the linear regression were practically negligible ($R^2 = 0.01$). In the delayed condition neither the correlations between sereial position and magnitude of JOLs, nor the regression weights were statistically significant. The estimates of effect sizes even smaller compared to those found in the immediate condition.

Table 1

Means and standard deviations of JOLs and recalled words across both experiments

Experiment 1

			JOL Timing			
			Immediate		Delayed	
			Easy	Difficult	Easy	Difficult
Young	Trial 1	JOLs	44.0 (*18.3*)	34.8 (*20.2*)	29.4 (*18.1*)	16.8 (*15.2*)
(n=36)		Recall	22.5 (*16.9*)	12.0 (*14.9*)	22.2 (*17.4*)	6.5 (*9.0*)
	Trial 2	JOLs	41.6 (*21.5*)	30.6 (*23.9*)	44.1 (*24.7*)	28.9 (*21.0*)
		Recall	52.5 (*24.0*)	35.2 (*27.4*)	52.5 (*25.1*)	34.9 (*22.6*)
Old	Trial 1	JOLs	42.0 (*18.6*)	31.2 (*22.6*)	17.7 (*11.6*)	10.4 (*9.7*)
(n=36)		Recall	12.7 (*16.2*)	2.2 (*5.2*)	9.3 (*11.1*)	2.2 (*4.5*)
	Trial 2	JOLs	29.8 (*23.1*)	16.9 (*20.0*)	19.3 (*16.1*)	10.6 (*12.4*)
		Recall	27.2 (*26.4*)	14.2 (*20.8*)	22.5 (*19.4*)	10.8 (*14.2*)

Experiment 2

			JOL Timing			
			Immediate		Delayed	
			Easy	Difficult	Easy	Difficult
Young	Trial 1	JOLs	47.8 (*21.0*)	33.4 (*22.0*)	36.4 (*20.5*)	17.2 (*18.4*)
(n=34)		Recall	28.6 (*16.6*)	8.1 (*10.9*)	26.1 (*18.5*)	8.3 (*10.9*)
	Trial 2	JOLs	53.2 (*20.3*)	32.1 (*17.4*)	51.5 (*21.5*)	31.4 (*18.5*)
		Recall	60.1 (*21.2*)	38.9 (*22.3*)	55.4 (*23.2*)	32.1 (*21.1*)
	Trial 3	JOLs	72.3 (*18.7*)	54.5 (*21.3*)	67.8 (*20.6*)	55.8 (*24.4*)
		Recall	75.3 (*20.2*)	66.9 (*19.6*)	72.2 (*20.2*)	60.0 (*23.9*)
	Trial 4	JOLs	81.3 (*17.1*)	75.5 (*18.8*)	78.2 (*18.3*)	71.7 (*22.9*)
		Recall	84.5 (*16.0*)	82.8 (*19.1*)	82.0 (*17.1*)	77.9 (*19.2*)
	Trial 5	JOLs	88.3 (*12.6*)	85.9 (*16.1*)	85.6 (*14.0*)	81.4 (*19.1*)
		Recall	89.4 (*11.8*)	90.4 (*14.1*)	88.1 (*13.6*)	85.3 (*17.5*)
Old	Trial 1	JOLs	48.6 (*26.9*)	38.4 (*31.3*)	23.6 (*22.8*)	14.9 (*21.2*)
(n=34)		Recall	7.8 (*8.7*)	0.7 (*2.4*)	6.0 (*7.7*)	0.5 (*2.0*)
	Trial 2	JOLs	36.5 (*27.6*)	26.3 (*30.7*)	23.7 (*20.0*)	12.5 (*13.8*)
		Recall	22.9 (*16.7*)	7.4 (*11.0*)	20.4 (*14.1*)	6.4 (*8.5*)
	Trial 3	JOLs	45.7 (*29.2*)	30.5 (*30.6*)	32.7 (*22.9*)	17.5 (*18.0*)
		Recall	34.3 (*19.3*)	15.2 (*18.5*)	31.4 (*19.5*)	12.0 (*13.5*)
	Trial 4	JOLs	52.2 (*29.5*)	37.8 (*33.4*)	37.9 (*25.4*)	22.6 (*21.2*)
		Recall	42.6 (*22.0*)	24.3 (*24.2*)	39.2 (*21.6*)	21.3 (*18.4*)
	Trial 5	JOLs	55.9 (*29.2*)	43.8 (*33.0*)	42.3 (*25.7*)	29.6 (*26.3*)
		Recall	47.2 (*21.5*)	34.1 (*26.6*)	44.1 (*20.6*)	30.4 (*26.1*)

Note. Standard deviations are in parentheses.

Figure Captions

Figure 1. Means of JOLs and recalled words are depicted in four panels representing two difficulty conditions and two timing conditions. Hatched lines represent JOLs and solid lines represent recall performance; grey symbolizes young adults and black old adults. The anchoring effect in the immediate condition in trial 1 is apparent upon inspection, where both age groups start at almost identical levels. In the delayed conditions, the JOLs are closer to the recall level and the UWP effect is smaller in young participants.

Figure 2. In order to make the UWP effect more apparent, the difference between JOLs and recall performance are displayed in all four conditions, represented by grey (young group) and black (old group) bars. Black stars represent statistically significant differences, at the $p < .05$ level, between JOLs and recall performance. The UWP effect is present if in trial 2 the average JOL level is significantly lower than the recall level, which is the case for young participants in both easy, and in the easy x difficult condition.

Figure 3. Means of JOLs and recalled words from Experiment 2 with five trials across all four conditions. Hatched lines represent JOLs and solid lines represent recall performance. The learning curves for older participants (black) are markedly lower compared to the young group (grey). Nonetheless, the average JOL in the first trial is practically identical in both groups for immediate JOLs which corroborates the notion of an anchoring mechanism. Delayed JOLs are much more calibrated and the anchoring effect is smaller in the first trial compared to immediately elicited JOLs.

Figure 4. The difference between mean JOLs and mean recall performance is represented by black and grey bars. Black stars represent statistically significant differences, at the $p < .05$ level, between JOLs and recall performance. Older adults did not show an UWP effect in any of the four conditions. Young adults, in contrast, displayed the UWP effect in the second trial, for immediately elicited JOLs and in the third or fourth trial for delayed judgments.

Figure 1

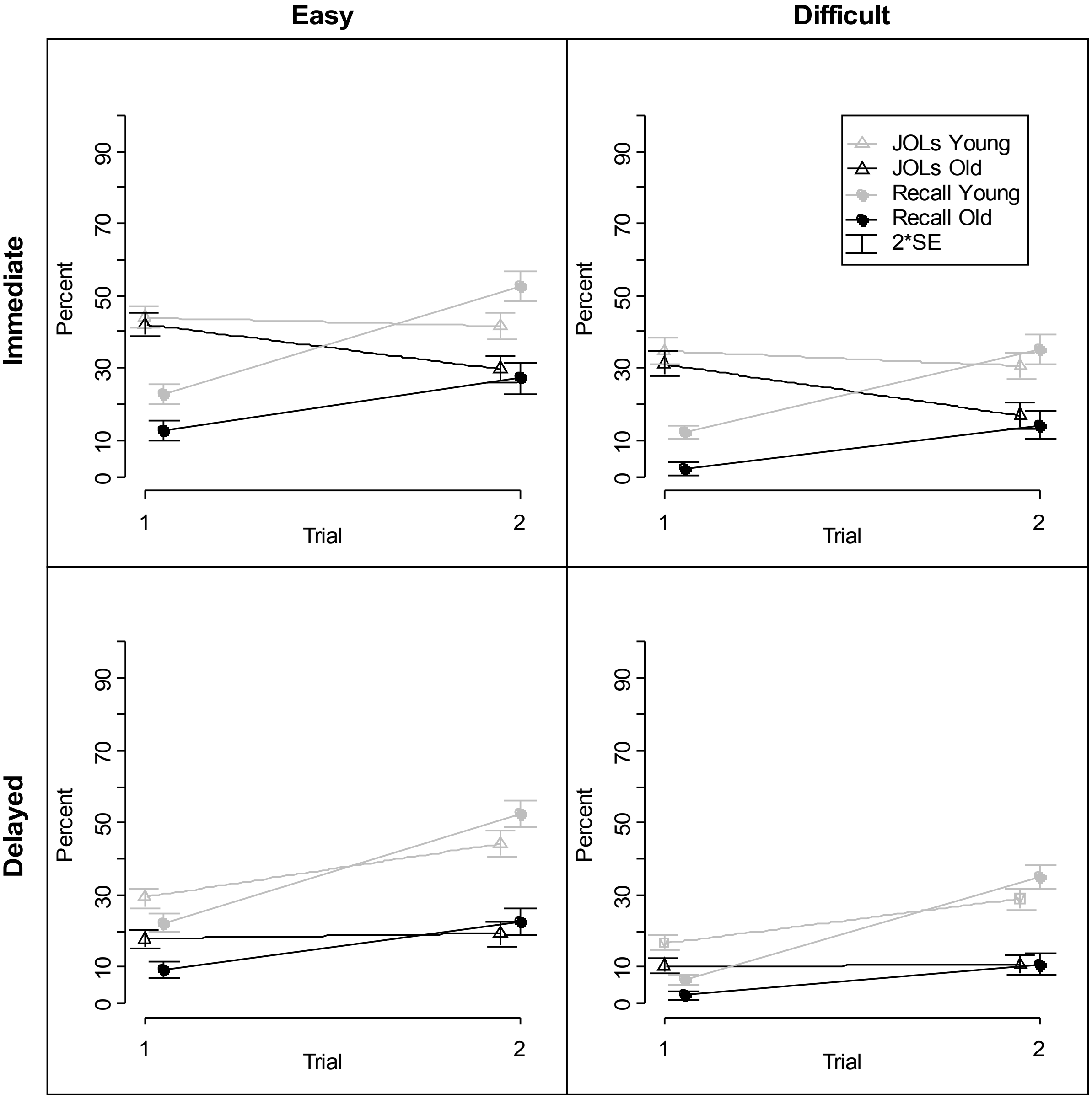

Easy
Difficult
Immediate
Delayed
Percent
0
10
30
50
70
90
Trial
1
2
JOLs Young
JOLs Old
Recall Young
Recall Old
2*SE

Figure 2

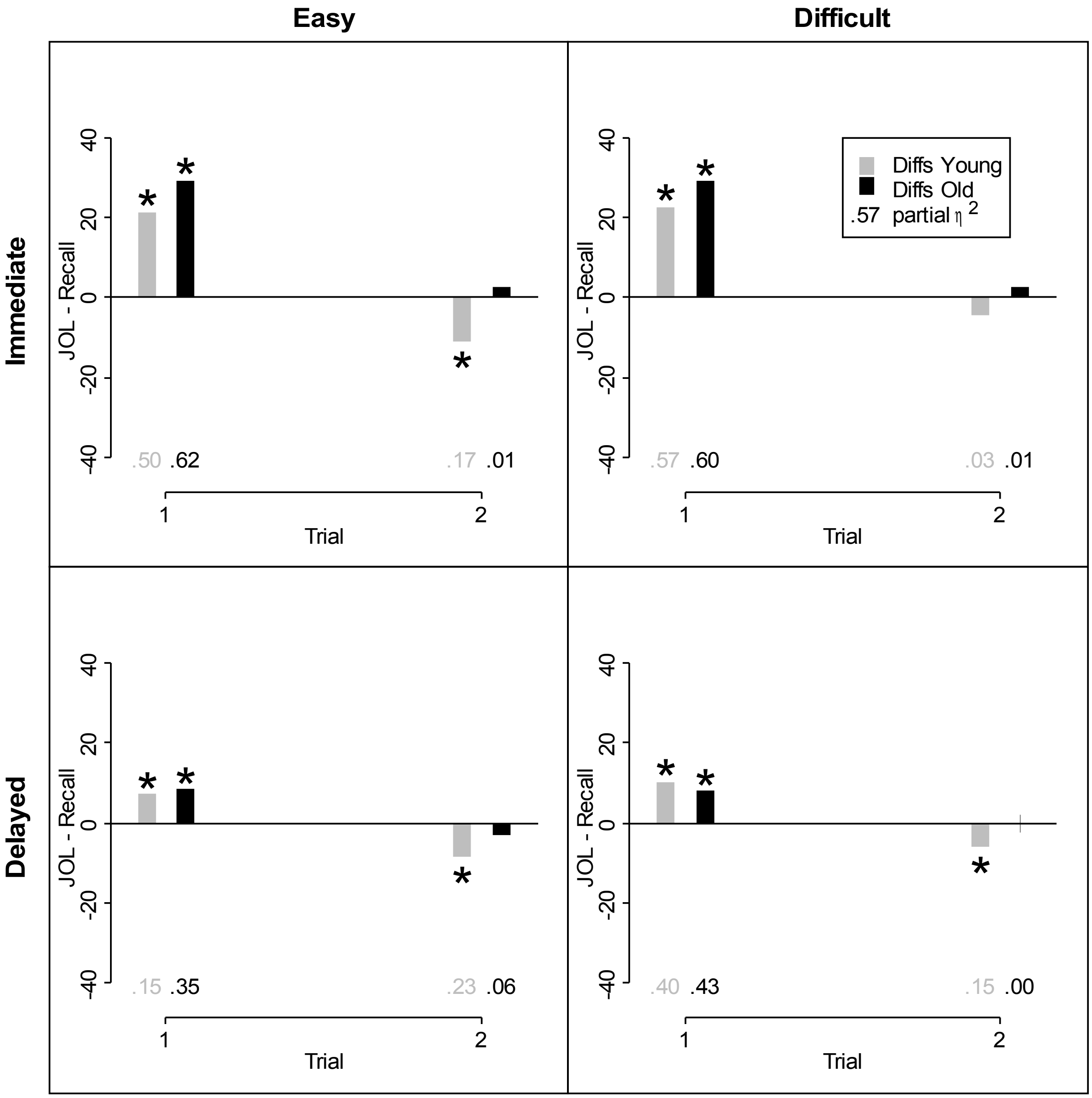

Easy
Difficult
Immediate
Delayed
JOL - Recall
40
20
0
-20
-40
*
.50 .62
.17 .01
.57 .60
.03 .01
.15 .35
.23 .06
.40 .43
.15 .00
1
2
Trial
Diffs Young
Diffs Old
.57 partial η 2

Figure 3

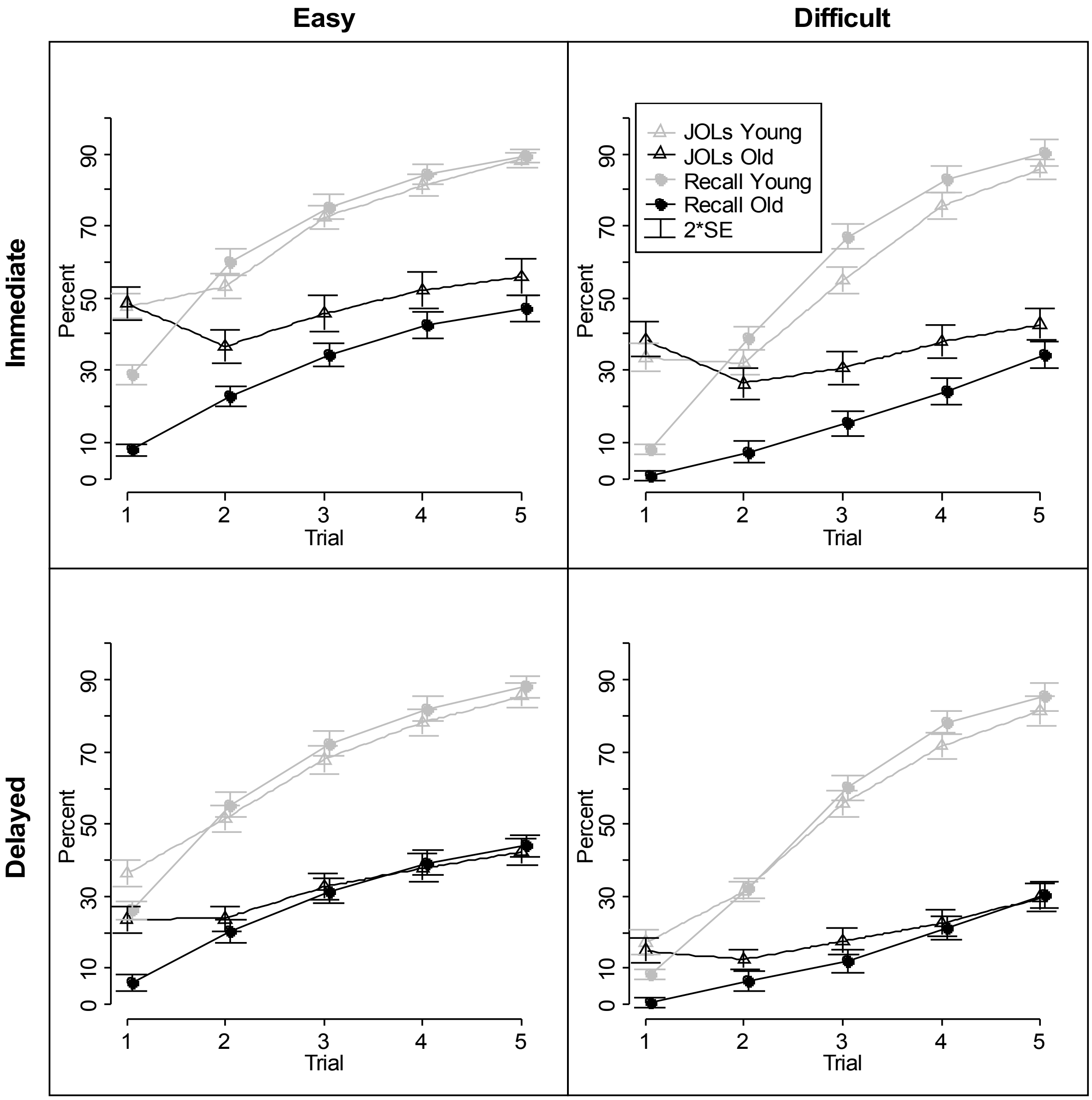
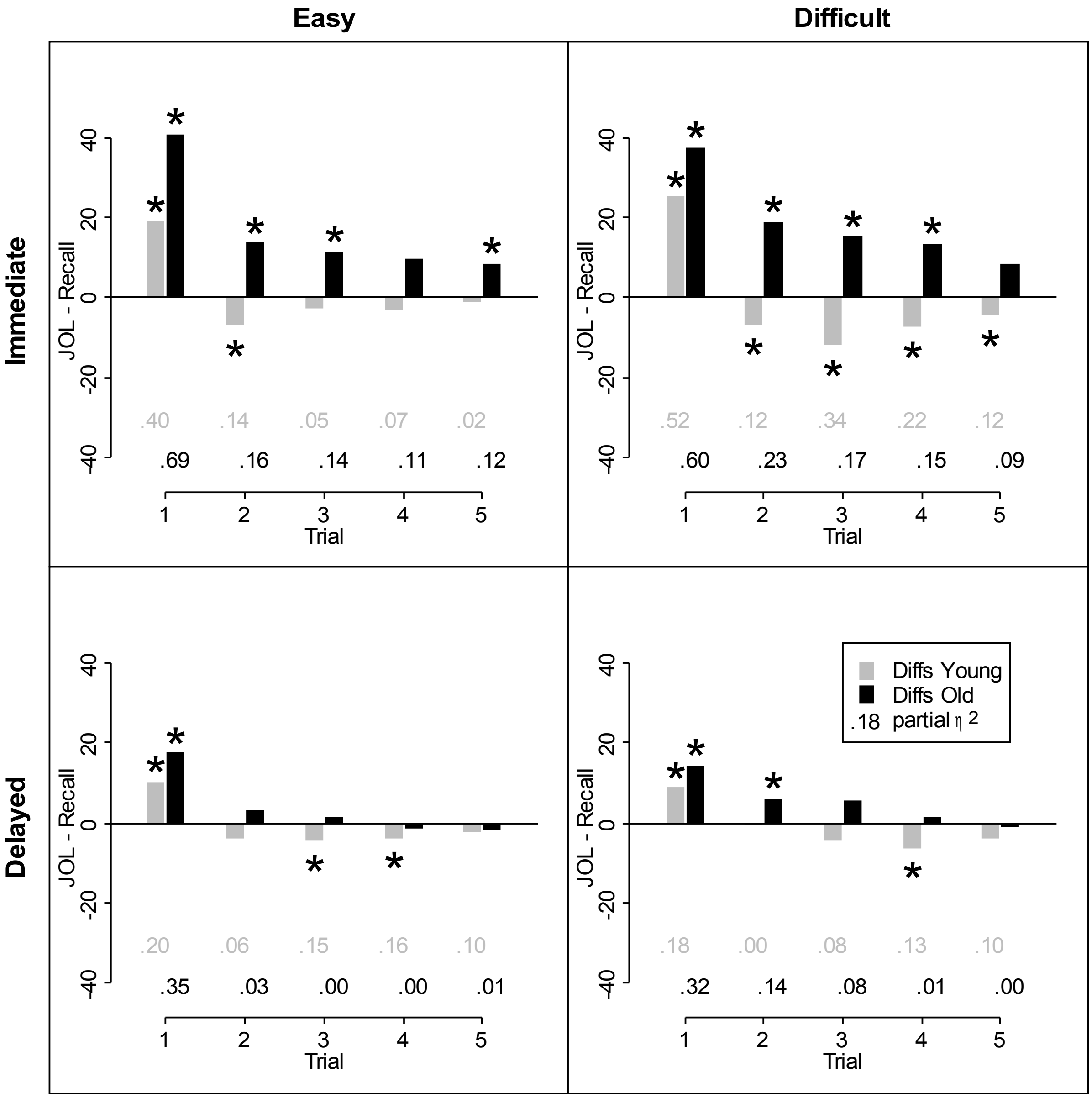

Easy
Difficult
Immediate
Delayed
Percent
0
10
30
50
70
90
1
2
3
4
5
Trial
JOLs Young
JOLs Old
Recall Young
Recall Old
2*SE

Figure 4


Easy
Difficult
Immediate
Delayed
JOL - Recall
40
20
0
-20
-40
Trial
1
2
3
4
5
*
.40 .14 .05 .07 .02
.69 .16 .14 .11 .12
.52 .12 .34 .22 .12
.60 .23 .17 .15 .09
.20 .06 .15 .16 .10
.35 .03 .00 .00 .01
.18 .00 .08 .13 .10
.32 .14 .08 .01 .00
Diffs Young
Diffs Old
.18 partial η 2