Investigating Interruptions in the Context of Computerised Cognitive Testing for Older Adults

Matthew Brehmer[1]**, Joanna McGrenere**[1]**, Charlotte Tang**[1]**, Claudia Jacova**[2]
[1]Department of Computer Science, [2]Department of Medicine
University of British Columbia, Vancouver, Canada
{brehmer, joanna, scctang} @cs.ubc.ca, cjacova@mail.ubc.ca

ABSTRACT

Interruptions in the home pose a threat to the validity of self-administered computerised cognitive testing. We report the findings of a laboratory experiment investigating the effects of increased interruption workload demand on older adults' computerised cognitive test performance. Related work has reported interruptions having a range of inhibitory and facilitatory effects on primary task performance. Cognitive ageing literature suggests that increased interruption workload demand should have greater detrimental effects on older adults' performance, when compared to younger adults. With 36 participants from 3 age groups (20-54, 55-69, 70+), we found divergent effects of increased interruption demand on two primary tasks. Results suggest that older and younger adults experience interruptions differently, but at no age is test performance compromised by demanding interruptions. This finding is reassuring with respect to the success of a self-administered computerised cognitive assessment test, and is likely to be useful for other applications used by older adults.

Author Keywords

Interruptions; experiment; task resumption; computerised cognitive assessment; older adults.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces - Evaluation/methodology;

General Terms

Experimentation, Human Factors.

INTRODUCTION

Interruptions are common in everyday life, occurring in all contexts, affecting all people, young and old. Interruptions can have detrimental effects on ongoing tasks, incurring costs to productivity [21] and increases in errors [13].

Effects of interruptions have been studied in many naturalistic and experimental settings, resulting in implications for designing applications to support productivity [2], decision-making [25], and vigilance [1].
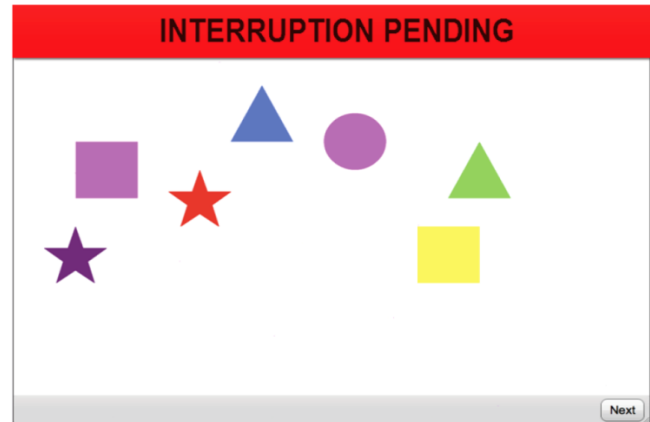
Figure 1. Notification of a pending interruption in a verbal memory task, adapted from a cognitive assessment task.

These implications largely focus on younger adults in workplace contexts. They have only minimally addressed how the ageing mind is affected by interruptions to ongoing primary tasks, which is the focus of our current work.

Our motivation stems from an initiative to develop Cognitve Testing on a Computer (C-TOC), a self-administered web-based computerised cognitive assessment that individuals will be able to take independently from their home [18]. With ongoing advances in modern medicine, people are living longer. This is associated with an increase of older individuals (55+) experiencing cognitive decline and presenting concerns regarding cognitive health. Currently, the screening for pathological cognitive decline as a result of Alzheimer's disease and related dementias are conducted using paper-based tests including the Mini-Mental State Examination [12] and the Montreal Cognitive Assessment [20]. They are administered by professionals in clinical settings during a visit. There is currently no opportunity to identify potential impairments before a visit. At our clinic for Alzheimer's and related dementias, which is representative of those in major urban centres in Canada, wait times for in-depth diagnoses and consultation regarding cognitive concerns ranges between 6 and 24 months. Thus, innovation in cognitive testing is an urgent yet unmet need due to the growing demand for diagnostic services.

The intent of C-TOC is to make preliminary cognitive testing more widely available and to triage individuals that exhibit pathological decline to fuller diagnostic testing as promptly

as possible. This novel testing tool brings together existing non-computerised tests and new test paradigms.

Since users will be accessing C-TOC at home, it is imperative to address the issue of interruptions and distractions that are pervasive in home environments. Interruptions may hinder older adults' progress in completing the test, potentially affecting their task performance which will in turn affect the validity of test results. Our research aim is to understand these effects, which will help inform designs for detecting and mitigating interruptions, specifically in C-TOC, and more generally in other applications designed for the ageing population.

The cognitive ageing literature suggests that interruptions will affect older adults to a greater extent than young adults [5,11,14,29]. We conducted an experiment to determine the effects of interruption on older adults' performance on two primary tasks adapted from C-TOC tests. A group of old adults (70+) was compared against two other age groups (19-54, 55-69). We did not include individuals with diagnosed cognitive impairments, as the effects of interruption must be gauged first when cognition is not impaired. We expected to find an increased cost of interruption for older adults, with disproportionally worse performance as interruption workload increases. We also expected a greater cost of interruption on a verbal working memory (WM) task than on a spatial problem solving task, as the former places a greater demand on WM.

Our findings indicate a cost to task resumption time incurred by interruptions, as well as possible compensatory behaviour in older adults' verbal memory task performance following an interruption: older adults were disproportionally slower than young adults to resume the task, but not disproportionally slower to complete the task. In a spatial problem solving task, we found evidence that interruptions were experienced differently by the different age groups. For both tasks, test scores were not compromised by interruptions.

The contributions of this research include the finding of divergent effects of increased interruption demand between different age groups for different primary tasks. These findings led to implications for the design of C-TOC; many of these implications are also promising for the design of other applications used by older adults.

RELATED WORK

We situate our work within a well-established body of research investigating the effects of interruptions on primary task performance. We review empirical costs of interruption and discuss factors for predicting these costs, including the age of the interrupted individual.

Measuring the COI

The cost of interruption (COI) on an ongoing primary task has been examined in naturalistic and experimental settings, and can be defined by several measures. A coarse COI measurement is the frequency of non-resumption of a primary task following an interruption, leaving the primary task uncompleted [21]. When primary task performance following an interruption is considered, the COI could be measured as the difference in task completion time between uninterrupted and interrupted conditions [33]. This measurement may not capture variation in behaviour immediately following an interruption [19]. Task resumption lag, the time elapsed when switching from an interrupting task back to the primary task, can be a precise local COI measurement [1,17,27]. Finally, COI can be measured in terms of the difference in a primary task's error rate between uninterrupted and interrupted conditions [13,22,25,26]. Many studies, including our own, report on several measures of COI.

Predicting the COI

Experimental approaches have attempted to isolate factors of primary and interrupting tasks predictive of the COI, however many divergent findings exist in the literature.

Interrupting task demand. Increased workload demand of the interrupting task has been shown to be predictive of the COI in some cases [13,19], but not in others [22]. However, low-demand interrupting tasks can sometimes improve performance on a primary task [33]. This phenomenon of improved performance on an interrupted primary task, compared to an uninterrupted task, was first reported by Zeigarnik [32] in the 1920s, and is now known as the `Zeigarnik effect'. If primary task representations can be encoded in long-term WM, increased interruption demand has no effect on primary task performance, otherwise increased interruption workload demand is disruptive [22].

Primary task demand. Bailey [2] examined the COI for two interrupting tasks on six primary tasks, finding memory demand in the primary task at the point of interruption to be most predictive of COI, rather than any demand incurred by the interrupting task. Similarly, an effect of primary task workload demand was also found by Speier [25], wherein for highly-demanding primary tasks, interruptions can inhibit performance; conversely, they also found that interruptions can actually improve performance on low-demand primary tasks, another *Zeigarnik effect* [32].

Similarity of primary & interrupting tasks. Gillie [13] found that a high degree of similarity between interrupting and primary tasks was predictive of greater COI (i.e., when tasks interfere with one another, engaging the same cognitive processes). However, later work failed to find an effect of similarity between primary and interrupting tasks [2].

Interruption duration. Earlier studies did not find interruption duration to be predictive of the COI [2,13], however this finding has recently come into question. Oulasvirta [22] found that increased interruption duration was found to incur a greater COI when primary task representations cannot be encoded into long-term WM . Monk [19] have also found that increased interruption duration contributes to lower primary task performance, at

odds with earlier findings [2,13]. They resolved that activation of primary task goals decays as a function of interruption duration.

Interruption lag. Prior research [1, 27] has examined the role of the interruption lag, the brief period of time in which an individual is alerted of an imminent interruption but is still focused on the primary task. This addresses the observation that switching to an interrupting task is seldom an immediate action (e.g., a ringing telephone preceding a conversation). A short interruption lag, as brief as 1-2s, may be sufficient for encoding primary task cues and prospective goals [1]. Retrospective rehearsal following an interruption can retrieve these goals [27].

Contextual factors. Observational and simulated naturalistic studies of interruptions have addressed the many contextual, temporal, and social factors found to be predictive of COI. These have included task structure [16], primary task visibility during an interruption [17], the frequency of interruption [33] and the source and modality of an interruption [26]. Incorporating these factors was outside the scope of our laboratory experiment, but we plan to assess the impact of interruption frequency and modality on COI in the near future.

The COI for Older Adults
Many cognitive processes change as we age, and each may contribute to the COI for older adults. Normal age-related changes in cognition can be attributed in part to slower processing speed [24]. Changes in cognition may also be attributed to reduced activation of WM [7]. Prospective memory, the ability to remember intentions, is also inhibited in older adults [29], and can be compromised by interruptions [11]. The ability to suppress reactions to distracting or irrelevant information appears to be reduced [15]. Given these changes, it is no surprise to find that older adults have a reduced capacity for multitasking [30] and attention switching [5]. However, age-related cognitive changes are not always marked by losses in functioning; there is evidence for compensatory brain activity in old adults, activating additional areas of the brain such that they perform as well as young adults on WM tasks [4].

Related work discussed in the previous section was largely carried out with younger adults, and did not analyse participant age as a factor for predicting COI. While older adults have been found to be susceptible to distractions [15] and have a reduced capacity for task switching [30], few studies have directly addressed the COI for older adults. Clapp [5] conducted an experiment to compare WM performance between young and old adults in interrupted conditions, where they found that older adults perform disproportionately worse than young adults on a short facial comparison task when distracted or interrupted. In terms of attentional modulation, they asserted that older adults attend to interrupting stimuli more than younger adults. In a naturalistic prospective memory task, older adults' ability to remember intentions was inhibited when it was interrupted

with a demanding verbal fluency task [11]. To our knowledge, no previous studies have directly compared the COI in the setting of computer applications involving multiple task types in an older population.

EXPERIMENT
We conducted an experiment to investigate the effects of age, primary task, and interrupting task demand on C-TOC test performance. As in previous work, we measured COI on primary task performance in terms of task resumption lag, completion time, and accuracy. We manipulated primary task type and interruption workload demand as independent variables. We maintained fixed levels of interruption duration, frequency, and lag visibility.

Methodology
Primary Tasks
Two primary tasks were used in this study, adapted from C-TOC test components: a verbal WM sentence comprehension task and a spatial problem solving task, hereby referred to as the VERBAL task and the SPATIAL task. In the VERBAL task, participants arrange geometric figures according to an instruction. In each trial, an single puzzle instruction (Figure 2a) is read before advancing to the execution step (Figure 2b), at which point the instruction is no longer accessible; thus the participant must hold the task instruction in verbal WM. In the SPATIAL task (Figure 2c), the participant is given vertical and horizontal lines, and is instructed to arrange them to create a number of complete squares in a specified number of moves (Figure 2d). Instructions remain visible throughout a trial, but the original layout or number of moves remaining must be held in WM. Comparitively, less information is held in WM this task than in the VERBAL task. Unlike previous studies of interruptions and older adults [5,11], our tasks are open-ended, and allow for partially-correct solutions.

Interruption Conditions
Three interruption conditions were used in this study. The first of which was an UNINTERRUPTED control condition. The other two conditions presented interrupting tasks. These filled the entire screen, occluding the primary task. In both types of interrupting tasks, an automated randomised sequence of a dozen cartoon images was shown at a rate of 1 image every 1.5s. Total interruption time was roughly 20s After this, the participant was prompted to click in order to dismiss the interruption and return to the interrupted primary task trial. The high-demand ACTIVE interruption (Figure 2f) is a variant of the established `n-back' working-memory task [23], in which participants must vigilantly monitor the series of images and click whenever an image is presented that is the same as the one presented two images prior in the sequence. Hence we used the `2-back' variant of the task (23 of 24 studies surveyed in Owen's meta-analysis [23] studied the *2*-back variant). This task was intentionally chosen to interfere with the WM demands of the primary tasks. In a low-demand PASSIVE interruption (Figure 2e), participants were instructed to watch the
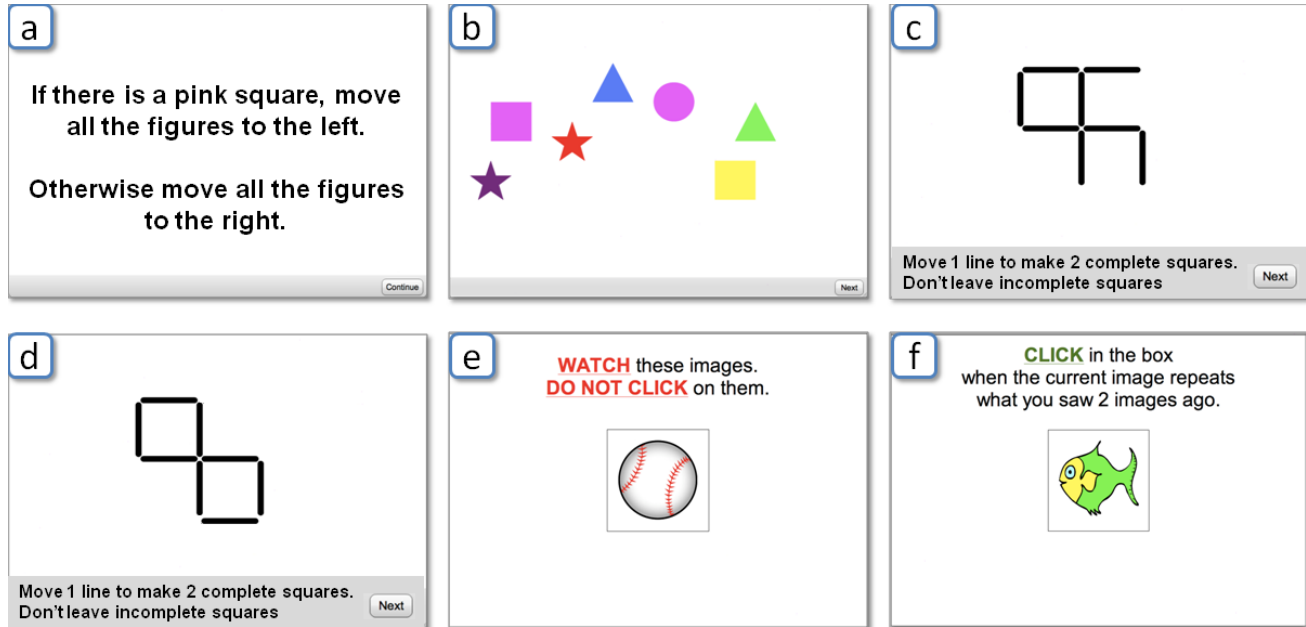
Figure 2. Primary & interrupting tasks used in the study: VERBAL task (a) instruction screen (b) execution screen prior to any user interaction; SPATIAL task (c) initial view (d) same view, completed task; Interruption tasks (e) PASSIVE (f) ACTIVE.

sequence of images passively until prompted to dismiss the interruption. The low-demand PASSIVE condition places no demand on WM. The similarity between interrupting tasks is deliberate; the visual stimulus remains constant, however a response to the stimulus is only required in the high-demand condition. The interrupting tasks are meant to simulate different levels of WM demand posed by interruptions which occur in the home. It is not our intention to map our interrupting tasks to specific household interruptions, but rather to represent a range of possible naturalistic interruptions.

Coordination of Primary and Interrupting Tasks
The two primary tasks each had three blocks of trials, one block for each of the three interruption conditions, counterbalanced across participants. Three isomorphic sets of trials were randomly allocated per participant to each block. Task instructions were unique between and within each set of trials, and corresponding trials between sets were isomorphic in terms of instruction complexity. In the PASSIVE and ACTIVE conditions, a subset of trials contained interruptions. This subset was selected at random for each participant; a subset of 4 out of 10 trials were interrupted in the VERBAL task while a subset of 3 out of 8 trials were interrupted in the SPATIAL task. For example, a participant interrupted on trials 2, 5, and 7 in the PASSIVE condition would also be interrupted on trials 2, 5, and 7 in the ACTIVE condition. This was necessary as corresponding trials between blocks had isomorphic task instructions. A diagram illustrating the coordination of primary and interrupting tasks can be found in the first author's thesis on p.50 [3]. Interruption onsets were fixed for each VERBAL trial, and would occur between 1s and 3s into the execution step, typically during or before a first move action is attempted. Therefore, the interruption onset was the same

between PASSIVE and ACTIVE conditions for trial *n*. In SPATIAL trials, interruption onsets would occur 500ms after the first or second completed move action, such that one or two outstanding moves were required after the interruption. Different interruption onsets for the two primary tasks was necessary due to differences in task structure and task completion time observed in pilot studies. Following from the observations of related work [1,27], our interrupting tasks were preceded by an interruption lag lasting 2s. During the interruption lag, the primary task is still visible but interaction is disabled; meanwhile a highly-salient interruption notification appears at the top of the screen (Figure 1). As discussed in related work, two seconds provides a sufficient amount of time to encode task goals and form cues for task resumption. Following the interruption lag, the interrupting tasks occupy the entire screen, the intent being to disrupt the primary task to a greater extent than what a partially-occluding or non-occluding interrupting task could accomplish [17,26]. At the end of the interrupting task, the user is prompted to click to dismiss the interruption, returning to the interrupted primary task at the point where it was interrupted.

Quantitative and Qualitative Measures
Three dependent measures were recorded for both primary tasks: Task resumption lag time, completion time, and accuracy. We considered data from the subset of trials in which interruptions occurred. For completion time and accuracy, we also considered the corresponding subset of trials from the UNINTERRUPTED condition. We measured task resumption lag as the time elapsed between returning to the primary task following an interruption and the completion of the first subsequent action (e.g., dragging a shape or line). Beginning at trial onset, we measured completion time as the total *uninterrupted* time elapsed

completing the subset of interrupted trials, which included task resumption time. The total time to complete the corresponding subset of trials was measured in the UNINTERRUPTED condition. We did not include the time spent reading trial instructions in the VERBAL task as part of trial completion time. We measured accuracy, a percentage score, according to a clinical scoring scheme used for C-TOC, based on scoring scheme used in the *Token* test [9]. This scheme accounts for the number of moves and correct relative positioning of shapes or lines, allowing for partially correct responses. Accuracy in the ACTIVE interrupting task was also recorded.

Subjective data concerning task difficulty and demand was collected on a questionnaire following each condition. The questionnaire was adapted from the NASA-TLX [14], a standardised instrument for assessing various dimensions of workload. Six questions were posed regarding mental and physical demand, annoyance, perceived performance, and fatigue. Responses were along a 10-point scale. At the end of the study, we interviewed participants to probe perceptions of task difficulty and task resumption strategies.

Apparatus
A laptop computer running Microsoft Windows XP was used for the experiment. A mouse was also used. The experimental software was written in Adobe Flex 4.0.

Participants
Thirty-six healthy participants were recruited from three age groups (12 each). The justification for these age groups rests on age-related changes in cognition that occur around the ages 55 and 70 [8]. Participants were recruited through advertisements placed throughout the community, and received financial compensation for their participation.

YOUNG *(19-54)*	Range: 19-50, (*M* = 31.0)	8F / 4M
PRE-OLD *(55-69)*	Range: 57-69, (*M* = 63.4)	9F / 3M
OLD *(70+)*	Range: 70-86, (*M* = 74.8)	6F / 6M

Procedure
The experiment was designed to fit into a single 90 minute session, which took place either in an exam room at our clinic or in a private room at a nearby community centre. We first administered the Montreal Cognitive Assessment (MOCA) [20] to help ensure that participants had no existing cognitive impairment. We then administered the North American Adult Reading Test (NAART) [28] to help ensure participants had sufficient English fluency to follow our instructions. Cutoff criteria were used for both tests: participants required a score of 26 or higher (out of 30) on the MOCA and were required to read at least 25% of words used in the NAART correctly. Based on these criteria, we excluded five participants (not included in the 36 above). They were allowed to finish the study, but their data were not included in our analysis. Those who scored less than 26 on the MOCA were later contacted by clinicians to arrange further consultation. For the remaining participants, MOCA and NAART scores were comparable between age groups.

Participants were first given examples of the interrupting tasks and asked to practice the ACTIVE interrupting task until they were familiarised with it. Participants then completed 4 blocks of trials for both primary tasks. The first block in each task was a short 4-trial practice block containing UNINTERRUPTED and interrupted trials. Consistent with the C-TOC battery, the remaining three test blocks in the VERBAL task contained 10 trials, while test blocks in the SPATIAL task contained 8 trials. The number of trials in a test block were representative of the number of trials appearing in the corresponding C-TOC tests. Participants were asked to complete each trial as quickly and as accurately as possible. After each block, participants filled out a copy of the questionnaire. Finally, participants were interviewed.

Design
A 3x3x2 mixed design was used; age (YOUNG, PRE-OLD, and OLD) was a between-subjects factor, and level of interruption demand (UNINTERRUPTED, PASSIVE, or ACTIVE) and primary task type (VERBAL, SPATIAL) were within-subject factors. Order of presentation for the within-subjects factors was fully counterbalanced, such that a participant began with all of either the VERBAL or SPATIAL task blocks.

Hypotheses
H1. Age & Interruption Demand.
1. *Overall,* YOUNG *adults will perform better than older (*PRE-OLD, OLD*) adults on the primary tasks.*
2. *Older (*PRE-OLD, OLD*) adults will incur a disproportionately larger* COI *when interruption demand increases.*

H2. Age, Task & Interruption Demand.
1. *Given that the* VERBAL *task places a greater load on memory, increased interruption demand will incur a disproportionately greater* COI *on the* VERBAL *task than on the* SPATIAL *task.*
2. *This difference in* COI *will be greater for older (*PRE-OLD, OLD*) adults.*

Results
Task resumption lag and task completion time results were log-transformed, correcting for positive skews. We performed a 2x3 (level of interruption demand x age) ANOVA on the task resumption lag data and a 3x3 (level of interruption demand x age) ANOVA on the completion time data. The accuracy data for both tasks was negatively skewed, so we performed nonparametric factorial 3x3 ANOVAs using the aligned rank transform [31], a method that can accommodate repeated measures designs and examine interaction effects. Pairwise comparisons were protected against Type I error using a Bonferroni adjustment. We report on measures that were significant (*p* < .05) or represent a possible trend (*p* < .10). Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size, where 0.01 is a small effect size, 0.06 is medium, and 0.14 is large [6]. We report significant findings on data from 36 participants (Figure 3).
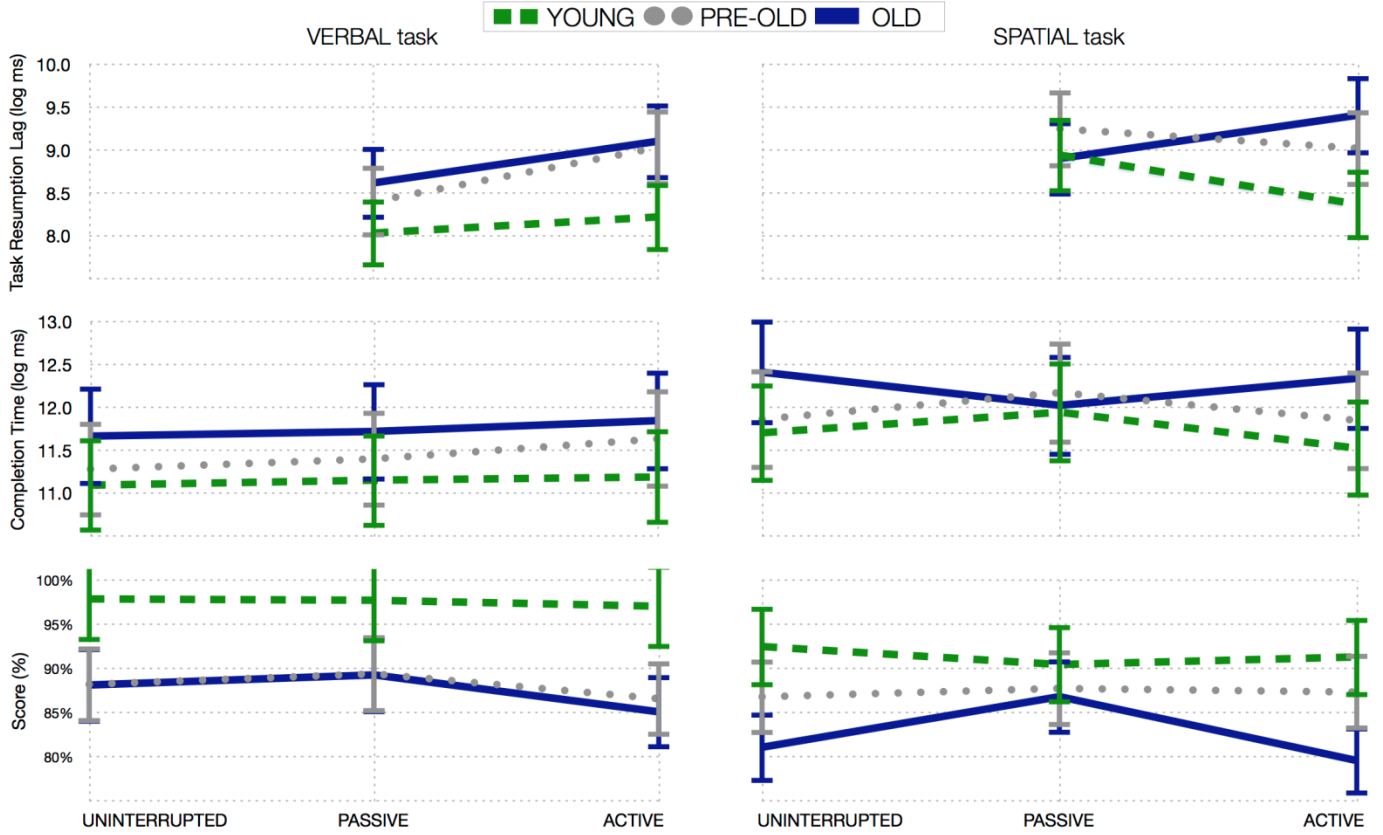
Figure 3 Quantitative results from the Verbal task (left) and the Spatial task (right). Top to bottom: log task resumption lag times, log completion times, and accuracy scores. Error bars show 95% CI (N = 36).

Verbal Task

Task resumption lag increases with age and interruption demand. With ACTIVE interruptions, OLD adults appear to be disproportionally slower than YOUNG adults to resume the task. The main effect of age was significant ($F_{2,33} = 11.89$, $p < .001$, $\eta^2 = .420$). Pairwise comparisons showed that OLD YOUNG adults ($p < .001$, $p = .002$, respectively).

The main effect of interruption demand is also significant ($F_{1,33} = 35.99$, $p < .001$, $\eta^2 = .522$); participants were slower to resume the task following ACTIVE interruptions than PASSIVE ones ($p < .001$). A trend suggests, as hypothesised, an interaction between age and interruption demand ($F_{2,33} = 3.22$, $p < .053$, $\eta^2 = .163$). Pairwise comparison on the interaction effect showed that PRE-OLD and OLD adults were slower to resume the task following ACTIVE interruptions than PASSIVE ones ($p < .001$, $p = .005$, respectively).

Completion time increases with age and interruption demand. Closely mirroring the resumption lag results, there was a main effect of age on completion time ($F_{2,33} = 12.00$, $p < .001$, $\eta^2 = .422$). Pairwise comparisons showed that OLD adults were slower than YOUNG adults (p < .001). There was also a significant main effect of interruption demand ($F_{2,66} = 12.47$, $p < .001$, $\eta^2 = .274$). Completion times were longer in the ACTIVE condition than in the PASSIVE and UNINTERRUPTED conditions ($p < .001$, $p = .006$,

respectively). Unlike the resumption lag results, however, there was no interaction of age and interruption demand.

OLD are less accurate than YOUNG. The main effect of age was significant ($F_{2,33} = 10.46$, $p < .001$, $\eta^2 = .388$), where OLD were less accurate than YOUNG adults ($p = .001$). The latter performed at ceiling levels.

Spatial Task

Task resumption lag increases with age, however YOUNG resume the task faster with ACTIVE interruptions, whereas OLD do not. The main effect of age was significant ($F_{2,33} = 3.40$, $p = .046$, $\eta^2 = .171$). The effect of interruption demand was not significant, however the interaction between age and interruption demand was significant ($F_{2,33} = 5.60$, $p = .008$, $\eta^2 = .253$). YOUNG adults resumed the task faster after ACTIVE interruptions than after PASSIVE ones ($p = .019$). A trend suggested that OLD adults were slower to resume the task after an ACTIVE interruption than after a PASSIVE one.

Completion time increases with age, however, no significant difference between OLD and YOUNG with PASSIVE interruptions. There was a main effect of age ($F_{2,33} = 4.09$, $p = .026$, $\eta^2 = .199$). Pairwise comparisons showed that OLD adults were slower than YOUNG adults ($p = .022$). There was no significant effect of interruption demand. However, different levels of interruption demand affected the age groups differently (interaction effect: $F_{4,66} = 3.28$, $p = .017$,

η^2 = .166). Pairwise comparison on the interaction effect showed that OLD adults are slower than YOUNG adults in the UNINTERRUPTED condition (p = .010). Surprisingly, there were no significant differences between groups in the PASSIVE condition. In the ACTIVE condition, OLD adults are again slower than YOUNG adults (p = .004).

With regards to accuracy, the main age effect was at a trend level (p = .078). Future inquiry is needed here.

Between-Tasks Analysis

Differences in performance between the two primary tasks were expected. Results of omnibus 3 (age) x 3 (interruption demand) x 2 (primary task) ANOVAs were thus not surprising: less time was taken to resume and complete the VERBAL task than the SPATIAL task (both p < .001). As expected, interactions between age, task, and interruption demand were also found, as shown in Figure 3. Accuracy did not differ as a factor of task or interruption demand.

Active Interruption Task

Participants attended to the ACTIVE interrupting task. YOUNG adults performed at ceiling, better than OLD adults. Mean scores on the ACTIVE `n-back' task were 9.5 (YOUNG), 8.8 (PRE-OLD), and 8.2 (OLD) out of 10, indicating that participants attended to the task, as instructed. There was a significant main effect of age on score ($F_{2,33}$ = 12.08, p < .001, η^2 = .437). Pairwise comparisons showed that OLD adults were less accurate than YOUNG adults (p < .001).

Subjective Findings: Questionnaire Responses

Responses to questionnaire questions revealed that OLD adults reported higher levels of mental demand than YOUNG adults (p = .050) and PRE-OLD adults (p = .030). Not surprisingly, the UNINTERRUPTED condition was reported to be less annoying than both the interrupted conditions (PASSIVE: p = .003; ACTIVE: p < .001). Surprisingly, the PASSIVE condition was not seen as less annoying than the ACTIVE condition. Consistent with our quantitative results, OLD adults reported lower performance than YOUNG adults (p = .012). Furthermore, in the SPATIAL task, OLD adults reported highest performance in the PASSIVE condition (p = .061), mirroring what is shown in Figure 3 (right column).

Subjective Findings: Interview Comments

The quantitative results include several divergent and trend levels results. This suggests that interruptions are experienced differently by the three age groups on the two primary tasks. We reviewed participants' interview comments regarding these differences.

When asked about the relative difficulty of the two primary tasks, regardless of interruption condition, the majority of participants (7 YOUNG, 8 PRE-OLD, 8 OLD) said that the SPATIAL task was more difficult. However, a majority (9 YOUNG, 6 PRE-OLD, 12 OLD) said that the VERBAL task was disrupted to a greater extent by interruptions.

When asked about task resumption strategies, there were no clear dominant strategy across age groups. All participants agreed that the ACTIVE interruption was too demanding to allow continued thinking of the primary task. Many participants, particularly those in the OLD group, did not form any task resumption strategy for the SPATIAL task. They claimed that interruptions had little or no effect on their performance, which is at odds with empirical findings.

Altogether, participants' comments do not decisively explain why ACTIVE and PASSIVE interruptions give rise to different performance effects for YOUNG and OLD participants, particularly in the SPATIAL task. However, we noted different reported behaviour among participants who did not rehearse primary task cues during the PASSIVE interruption: some YOUNG participants reported attending to the PASSIVE interruption while others let their mind wander. OLD participants reported ignoring it and feeling impatient.

Summary

Task	Resumption lag	Completion time	Accuracy
VERBAL	OLD disproportionally slower than YOUNG in ACTIVE	all groups slower in ACTIVE	
SPATIAL	YOUNG faster in ACTIVE than PASSIVE	no age differences in PASSIVE *	age effect not sig. *

Table 1. Summary of quantitative findings from the experiment. The effect of age was found everywhere except in those cells with an *.

We summarise our results according to our hypotheses:

H1. Age & Interruption Demand.
1. *Overall,* YOUNG *adults will perform better than older (*PRE-OLD*,* OLD*) adults on the primary tasks*. Supported. OLD adults took longer to resume and complete a primary task and were less accurate than YOUNG adults. However, there was no difference in accuracy in the SPATIAL task.
2. *Older (*PRE-OLD*,* OLD*) adults will incur a disproportionately larger COI when interruption demand increases*. Partially supported. OLD adults take disproportionally longer than YOUNG adults to resume the VERBAL task in the Active condition, when compared to the PASSIVE condition. This is not supported by completion time or accuracy results, nor by SPATIAL task results.

H2. Age, Task & Interruption Demand.
1. *Given that the* VERBAL *task places a greater load on memory, increased interruption demand will incur a disproportionately greater COI on the* VERBAL *task than on the* SPATIAL *task*. Partially supported. The effect of increased interruption demand was significant for the VERBAL task but not for the SPATIAL task, in terms of completion time and resumption lag time, but not accuracy.
2. *This difference in COI will be greater for older (*PRE-OLD*,* OLD*) adults*. Partially supported. OLD adults have disproportionately longer task resumption lags than YOUNG adults following an Active interruption in the VERBAL task, however completion times were not disproportionately longer. In the SPATIAL task, OLD and

Young experience the PASSIVE and ACTIVE interruptions differently. Subjective responses fail to explain this finding. Despite this, there is no significant COI as a factor of interruption demand.

DISCUSSION

In this section, we reflect on the key quantitative findings listed in Table 1. Qualitative findings from questionnaires and interviews are also considered.

Old adults compensate for slower task resumption. We reported a COI on a verbal memory task, where increased interruption demand incurred disproportionally longer task resumption times among OLD adults. This finding is also supported by cognitive ageing literature [30], where task switching response times have been shown to be greater for older adults. Unexpectedly, however, our OLD participants were not disproportionately slower to complete the task. Therefore, after being initially slow to resume the task, OLD adults compensated by increasing their rate of completion, relative to their rate of completion in the UNINTERRUPTED condition. We speculate that this behaviour is the result of an age-specific *Zeigarnik effect* [32] for older adults: a motivated effort to work with heightened efficiency after being interrupted, making up for lost time incurred by the initial COI. Alternatively, longer resumption lags may have allowed for the formulation of more efficient strategies for completing the primary task. However, this was not confirmed by interview responses.

Primary task accuracy was not affected by interruptions, regardless of age. This indicates that task goals, such as instructions in trials of the verbal memory task, were successfully encoded into WM before an interruption took place [22]. Thus, task goals were largely resistant to interference caused by demanding interruptions.

Methodological Implications

Low-demand interruptions affect age groups differently. While there was no COI in the SPATIAL task, we observed that different age groups experienced low- (PASSIVE) and high-demand (ACTIVE) interruptions differently. In particular, YOUNG adults were faster than OLD adults to resume and complete the task in the ACTIVE condition, but the groups did not differ in the PASSIVE condition. In the PASSIVE condition, mind-wandering may have caused young adults' performance to slip, as suggested by some interview comments. This mind-wandering afforded by the low-demand interruption may have actually had a greater negative effect on primary task performance than the high demand interruption. It is also possible that older adults were more conscientious than younger adults, and resisted mind-wandering. Cognitive testing is a sensitive topic for older adults [10], and thus a fear of poor performance may have resulted in increased conscientiousness. We deliberately designed our low-demand interruption task to require no action, while maintaining a high visual similarity to the high-demand interruption task. We expect a low-demand interruption task that requires a simple action (i.e., clicking on every image) would reduce mind-wandering.

Regardless of age, low-demand PASSIVE interruptions give participants a choice: they can allow their mind to wander, or they can use the opportunity to rehearse primary task cues. Interviews with participants revealed no age-specific trends with regards to their cognitive processes occurring during PASSIVE interruptions. An aim of future research will be to identify any age-related differences in strategy.

High-demand interruptions may not have been difficult enough for our YOUNG *adults*. Among our YOUNG participants, there were no significant differences in task performance between the interruption conditions, with the exception that young participants' task resumption lag times in the SPATIAL task were significantly shorter with high-demand ACTIVE interruptions than with low-demand PASSIVE interruptions. By contrast, Monk [19] showed that increased interruption demand incurs a COI in terms of longer task resumption lag times. This suggests that our ACTIVE interruption task may not have been sufficiently demanding for our young participants. We also observed a ceiling effect for young participants in terms of score on the ACTIVE interruption task; our OLD participants performed worse, but were still far from floor levels, and thus could have endured a more challenging task. While the 2-back variant of the *n*-back task is the most studied in the literature [23], future work could increase *n* in the `*n*-back' task, beyond we used, until ceiling effects are avoided.

An alternate explanation is that *the combination* of primary and interrupting tasks was not sufficiently difficult for our young adults. This would not be altogether surprising, as the C-TOC tasks are intended for older adults, and should be considerably easier for young adults.
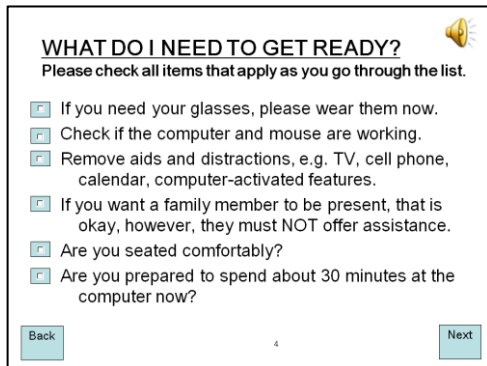
Design Implications

Our primary tasks were adapted from C-TOC, a self-administered computerised cognitive assessment. As a reminder, C-TOC need not be as accurate as a full clinical assessment; it is only needed to triage patients. In light of our results, we address implications for preventing, detecting, and mitigating interruptions in terms of UI design for the purposes of preserving test validity.

Prevent interruptions with prompts tailored to each test. A prompt to prevent external interruptions and distractions currently appears in the preparation screen displayed once at the beginning of C-TOC (Figure 5). Since we found divergent effects of interrupting tasks on different C-TOC tasks, it may be necessary to repeat this prompt at the outset of tasks that are particularly sensitive to the effects of interruptions, such as the VERBAL task. Performance is currently weighted differently between completion time and accuracy for each test. Weights could be made explicit to test-takers, to increase awareness of how each test is scored and how an interruption may affect their score.

Detect interruptions by requiring user response. Mitigate interruptions with trial replacement and test restarts. Periods of inactivity and unusual variation in test performance cannot always be assumed to be caused by an interruption: an older individual may be challenged by the test and a lack of activity may represent genuine performance. A threshold amount of inactive time should be determined for each C-TOC test. Once this threshold is reached, a highly salient prompt should appear, querying the test-taker to determine if the current inactivity is due to an interruption; mouse movement or a key press would dismiss the prompt, allowing immediate continuation of the primary task trial. If the prompt is not quickly dismissed, C-TOC could resolve that an interruption has occurred. In this case, the interrupted primary task trial should be discarded and replaced with an isomorphic trial upon task resumption. Finally, in cases of prolonged interruptions, a global inactive time threshold should also be determined; once passed, the test-taker would be required to restart the current test, or, if need be, the entire C-TOC battery.

Figure 5. Preparation screen displayed before beginning C-TOC, which includes a prompt to prevent external interruptions and distractions.

Detect interruptions by examining variation in rate of activity. If the task is one in which older adults are known to compensate following task resumption (such as our VERBAL task), the rate of activity, or average time between valid actions in a task (i.e., moving objects), before and after the period of inactivity should always be compared, once the trial is completed. When it appears likely that an interruption occurred, the performance could be classified as invalid and the user could be required to complete an isomorphic replacement trial.

The user was interruped. Is their performance invalid? Effects of interruptions on primary task performance may not always incur a cost to performance, as was observed in the SPATIAL task. In these cases, trial completion time (minus inactive time) can be retained for assessment purposes. In cases such as the VERBAL task, completion time results may no longer be valid, however accuracy results will still be reliable. The decision to retain performance data despite the occurrence of interruptions will vary from test to test and will depend greatlty on how

the test is scored. Alternative scoring schemes may need to be developed for interrupted tests.

Accuracy ultimately remains the most important performance criteria, in C-TOC and in existing clinical testing. Completion time is a secondary measure of performance. Given our result that accuracy remains unaffected by interruptions, C-TOC test results remain largely valid even if the user was interrupted.

In general, segment tasks and determine inactivity thresholds. Our findings are relevant to the design of all applications used by older adults in contexts where interruptions and distractions might occur and have potentially detrimental effects, such as online banking or booking a travel itinerary. Segmenting longer tasks into smaller sub-tasks and setting inactive time thresholds based on the task structure can limit the effects of interruptions.

CONCLUSION AND FUTURE WORK

Our contributions include the significant finding of divergent effects of increased interruption demand on older adults' primary task performance. Increased interruption demand can incur a cost to performance for older adults; however, these effects are dependent on the cognitive processes required by the primary task. We also contribute design implications for C-TOC, a self-administered computerised cognitive assessment test. Many of these implications are promising for the design of other applications used by older adults.

Increased interruption demand did not affect task accuracy. Highly demanding interruptions caused longer task completion times in a verbal memory task, but this was true of both YOUNG and OLD participants. Overall, the results suggest that healthy older individuals are fairly robust to interruptions, even when they are demanding. This is a reassuring result with respect to the viability of C-TOC in the home. As individuals with memory impairments are expected to struggle following interruptions, lower task accuracy will help identify impaired individuals.

In parallel to the study reported in this paper, we have been investigating the behaviour of older adults as they complete C-TOC in home settings, while documenting the types and effects of naturalistic interruptions and distractions. We are also currently replicating this study's methodology with different clinical groups (i.e., those diagnosed as having Mild Cognitive Impairment) and with different levels of low and high interruption demand. This investigation will also examine interactions of age and interruption demand on other C-TOC tests, engaging cognitive processes other than verbal memory and spatial problem solving. We hope to attain a deeper qualitative understanding of age differences in strategy in response to different primary tasks and levels of interruption demand, as well as the impact of varying levels of computer literacy. Finally, we plan to evaluate methods for preventing, detecting, and mitigating

interruptions, as well as the effects of these methods on test validity.

REFERENCES

1. Altmann, E. M. & Trafton, J. G. Task interruption: Resumption lag & the role of cues. In *Proc. Cognitive Science Society*, 43-48, 2104.

2. Bailey, B. P., Konstan, J. A., & Carlis, J. V. Measuring the effects of interruptions on task performance in the user interface. In *Systems, Man & Cybernetics*, 757-762, 2000.

3. Brehmer, M. M. *Usability and the effects of interruption in C-TOC: Self-Administered Cognitive Testing on a Computer*. M.Sc thesis. University of British Columbia, 2011.

4. Cabeza, R. Task-independent & Task-specific Age Effects on Brain Activity during Working Memory, Visual Attention and Episodic Retrieval. *Cerebral Cortex, 14*(4):364-375, 2004.

5. Clapp, W. C. & Gazzaley, A.. Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiology of Aging*: 1-15, 2010, In Press.

6. Cohen, J. Eta-squared and partial eta-squared in communication science. *Human Communication Research, 28*:473-490, 1973.

7. Craik, F .I. M. & Byrd, M. Aging and cognitive deficits: The role of attentional resources, 19--222. In Craik, F. I. M. & Trehub, S. E. (eds.), *Aging & Cognitive Processes*. Plenum Press, 1982.

8. Craik, F. I. M. & Salthouse, T. A., Eds. *The Handbook of Human Aging & Cognition*. Erlbaum, Hillsdale, NJ, 2nd edition, 1992.

9. De Renzi, E. & Vignolo, L. A. The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain: A Journal of Neurology 85*: 665-678, 1962.

10. Dickinson, A., Arnott, J., & Prior. S. Methods for human-computer interaction research with older people. *Behaviour & Information Technology 26*(4):343-352, 2007.

11. Farrimond, S., Knight, R. G., & Titov, N. The effects of aging on remembering intentions: performance on a simulated shopping task. *Applied Cognitive Psychology, 20*(4):533-555, 2006.

12. Folstein, M. F., Folstein, S. E., & McHugh, P. R. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12:189-198, 1975.

13. Gillie, T. & Broadbent, D. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research, 50*(4):243-250, 1989.

14. Hart, S.G. & Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, 139-183. In Hancock, P. & Meshkati, N. (Eds.). *Advances in Human Psychology: Human Mental Workload*. Elsevier Science, 1988.

15. Hasher, L. & Zacks, R. T. Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning & Motivation, 22*:193-225, 1988.

16. Iqbal, S. T. & Bailey, B. P. Leveraging characteristics of task structure to predict the cost of interruption. In *Proc. CHI*, 741-750, 2006.

17. Iqbal, S. T. & Horvitz, E. Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proc. CHI*, 677-686, 2007.

18. Jacova, C., Lee H. S., Le Huray, S., McGrenere, J., Beattie, B. L., Feldman, H., and Hsiung G-Y. R. Cognitive Testing on Computer (C-TOC): Designing a computerized test battery for evaluation of cognitive impairment with user and community health professional input. *Alzheimer's & Dementia, 6*(4):S317-S318*, 2010.

19. Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied, 14*(4):299-313, 2008.

20. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4):695-699, 2005.

21. O'Conaill, B. & Frohlich, D. Timespace in the workplace: Dealing with interruptions. In *Proc. CHI*, 262-263, 1995.

22. Oulasvirta, A. & Saariluoma, P. Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies, 64*(10):941-961, 2006.

23. Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1):46-59, 2005.

24. Salthouse, T. A. The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3):403-428, 1996.

25. Speier, C., Vessey, I., & Valacich, J. S. The effects of interruptions, task complexity, and information presentation on computer-supported decision-making Performance. *Decision Sciences, 34*(4):771-797, 2003.

26. Storch, N. A. *Does the user interface make interruptions disruptive? A study of interface style and form of interruption.* Technical report, Lawrence Livermore National Laboratory, Springfield, 1992.

27. Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58*(5):583-603, 2003.

28. Uttl, B. North American Adult Reading Test: age norms, reliability, and validity. *Clinical & Experimental Neuropsychology, 24*(8):1123-1137, 2002.

29. Uttl, B. Transparent meta-analysis of prospective memory and aging. *PloS ONE, 3*(2):e1568 1--31, 2008.

30. Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. Aging and task switching: a meta-analysis. *Psychology & Aging, 26*(1):15-20, 2011.

31. Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proc. CHI*, 143-146, 2011.

32. Zeigarnik, B. Das behalten erledigter und unerledigter handlungen. *Psychologische Forschung, 9*:1-85, 1927.

33. Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. *Temporal factors in mental* work: Effects of interrupted activities. *Journal of Occupational & Organizational Psychology,72*(2):163-185,1999.