Tara Radović, Dietrich Manzey

Effects of complexity and similarity of an interruption task on resilience toward interruptions in a procedural task with sequential constraints

Citation details
Radović, T., & Manzey, D. (2022). Effects of complexity and similarity of an interruption task on resilience toward interruptions in a procedural task with sequential constraints. In Journal of Experimental Psychology: Human Perception and Performance (Vol. 48, Issue 2, pp. 159–173). American Psychological Association (APA). https://doi.org/10.1037/xhp0000981.

Effects of Complexity and Similarity of an Interruption Task on Resilience toward Interruptions in a Procedural Task with Sequential Constraints

Tara Radović

Dietrich Manzey

Technische Universitaet Berlin

Word count: 10644 words

*Corresponding author: Tara Radović (contact: tara.radovic@tu-berlin.de)

Dietrich Manzey (contact: dietrich.manzey@tu-berlin.de)

The goal of the present study was to examine effects of complexity and similarity of an interruption task on post-interruption performance in an eight-step procedural task with sequential constraints. In Experiment 1, the primary task was interrupted between different steps with one of four versions of n-back task, which differed in complexity (simple, complex) and similarity in processing codes (verbal, spatial) to the primary task. After the interruption, participants (N = 44) had to resume the primary task as quickly as possible with the next correct step, i.e., the one following the step after which the interruption occurred. Post-interruption performance in terms of resumption times, sequence errors and non-sequence errors was assessed. Results of Experiment 1 revealed longer resumption times and more sequence errors after complex interruptions compared to the simple ones. However, effects of processing-code similarity were less clear. For assessing the effects of similarity in processing codes again in Experiment 2, participants (N = 41) performed the same primary task and were interrupted with a verbal or a spatial classification task. The results revealed no significant effect of processing code on the post-interruption performance. Moreover, a post-hoc analysis revealed that 1-back (sequential) interruption led to longer resumption times compared to the classification (non-sequential) interruption. Overall, our results revealed strong and consistent detrimental effects of interruption complexity on the post-interruption performance and no effect of similarity in processing codes. Finally, we provide preliminary evidence that similarity in sequential structure between the tasks can influence the resilience toward interruptions.

Statement of significance:

This study pinpoints important aspects of an interruption task, which influence post-interruption performance. The study confirmed that increasing interruption complexity has strong detrimental effects on the post-interruption performance, while similarity in terms of processing codes between the tasks does not seem to play a role. These findings strongly suggest involvement of general memory processes in resilience toward interruptions, rather than domain-specific systems proposed by some of the previous research. We provide preliminary evidence that the similarity in terms of sequential structure between the two tasks has an impact on post-interruption resumption times. Finally, our study is the first one to provide empirical evidence for involvement of memory for serial order in execution of procedural tasks and in interruption management.

Introduction

Remembering how to conduct a procedure consisting of several steps that need to be executed in a predefined order is present in everyday life and in professional environments. An example of such a procedure that we encounter daily is preparing espresso using a moka pot that consists of three parts. To prepare the coffee, one needs to disassemble the pot first, then to fill the bottom part with water, to place the middle part on the bottom part of the pot and fill it up with grained coffee, and finally to screw the top part to it. Afterwards, the mocha pot is ready to be put on a stove to heat up until boiling. While alternating the optimal order of steps in this type of everyday task does not have serious consequences, deviating from a prescribed order of steps can impose a serious risk in some other domains. For example, in high-risk domains, such as medicine or aviation, the correct execution of strictly defined procedures often is critical and, thus, typically supported by an implementation of checklists. However, in some cases, also in these domains, procedures must be retrieved and performed based on memory only. The risk of committing a procedural error of skipping or repeating a step can have fatal consequences. One important factor which was shown to significantly elevate the risk of committing such procedural errors are interruptions, i.e., the unanticipated requirement to perform another task for more of less long periods while still completing a certain procedure (Dismukes et al., 1998; Drews, 2007; Latorella, 1996; Loukopoulos et al., 2001, 2003; Scott-Cawiezell et al., 2007; Westbrook et al., 2010).

To examine the effects of interruptions on the performance of procedural tasks in a laboratory environment, several experimental paradigms such as video-cassette-recorder (VCR) programming (e.g., Monk et al., 2002) or the UNRAVEL task (Altmann et al., 2014) are developed, aiming to represent or simulate complex cognitive tasks consisting of several steps. Research using this type of task as a primary (interrupted) task has confirmed detrimental performance consequences of interruptions, specifically in terms of so-called resumption costs, i.e., increased response times and/or error rates upon resumption of the primary task after an interruption compared to an uninterrupted condition (e.g., Altmann et al., 2017; Hodgetts & Jones, 2006; Monk et al., 2008). Most of these effects are typically limited to the step of the primary task immediately following the interruption (Altmann, et al., 2014). One important factor determining the performance consequences of interruptions is the length of an interruption. As it has been shown repeatedly, longer interruptions lead to even more errors and longer response times when resuming the primary task compared to shorter interruptions (e.g., Altmann et al., 2017; Monk et al., 2008; Radović & Manzey, 2019).

These effects are typically interpreted within the *Memory for Goals* (MfG) model proposed by Altmann and Trafton (2002). The model assumes that goals in working memory have different activation levels, with the most active goal governing behavior. This means that it is more likely that the task-relevant goal will be sampled from working memory, if its activation is above interference level, i.e., mean activation level of the most active task-irrelevant goal (distractor). When a procedural task gets interrupted, the goal of the procedural task gets disturbed by the currently relevant and activated goal of the interruption task, and its activation will decay gradually with time. The decay is assumed to be greater, the longer the interruption takes. To resume the primary task after the interruption successfully, the primary task goal needs to be re-activated using internal or external cues, which is enabled through a priming component of the model. It is further assumed that the process of overcoming a decrease in the activation of the primary-task goal takes more time.

A similar and complementary conception of interruption management as a memory problem has been proposed by models of prospective memory (Dodhia & Dismukes, 2009; Einstein et al., 2003). According to this view, interruptions are considered to induce a prospective memory task, as intentions of the primary task must be memorized in order to be executed in the future, i.e., after a time delay caused by the interruption task. Thus, similarly to the MfG, this conceptualization proposes active maintenance of different aspects of the primary task (e.g., interruption position, next step to be executed) during the interruption task by using an internal rehearsal process. Additionally, it introduces a resource-theoretical perspective, as it proposes that the maintenance of delayed intentions is an effortful process that poses a moderate demand on limited cognitive resources (Dodhia & Dismukes, 2009; Einstein et al., 2003). This perspective can explain not only performance consequences in terms of resumption costs, but also possible interference effects between the rehearsal of primary task goals and interruption-task performance. Moreover, it can account for increased error rates when resuming the primary task, e.g., skipping a step of the primary task or repeating a step which already had been performed.

Beside the length of interruption, certain characteristics of the interruption task itself seem to influence the resumption costs after an interruption. These include the complexity of the interruption task and its similarity with the primary task. Effects of the complexity of interruption tasks on the post-interruption performance in a primary task was addressed in several studies (e.g., Cades et al., 2007, 2008; Gillie & Broadbent, 1989; Hodgetts & Jones, 2006; Monk et al., 2008; Zijlstra et al., 1999). For example, Cades and colleagues (2008) investigated how VCR programming would be affected by occasional interruptions with a complex or a simple number categorization task. Their results revealed significantly longer resumption times after the complex, compared to the simple interruption task. Similar results were obtained in the study done by Hodgetts and Jones (2006), who showed that participants needed more time to resume the Tower of London task after being interrupted by a complex arithmetic task compared to a simple arithmetic task. Consistently, the results of these studies show that increasing the complexity of an interruption task also causes greater resumption costs. The explanation for the complexity effect based on the MfG model (Altmann & Trafton, 2002) assumes that more complex interruptions tasks also involve more sub-goals that need to be activated in working memory compared to less complex tasks. This in turn increases the mental clutter and the interference level which needs to be overcome in order to resume the primary task after the interruption. For that reason, re-activation of the primary task goal would be more challenging and time-consuming after a complex interruption than a simple one, resulting in greater resumption costs when resuming the primary task. An even more straightforward explanation of the complexity effect can be derived from the prospective-memory model (Dodhia & Dismukes, 2009; Einstein et al., 2003). Here it is assumed that performing a complex interruption task demands more cognitive resources than performing a simple one. Consequently, resources available for active maintenance of intentions of the primary task during the interruption are reduced more by a complex than a simple interruption task, leading to greater interference and greater resumption costs upon the primary task resumption.

Considerably fewer studies have addressed the similarity between a primary task and an interruption task as a possible determinant of resumption costs. The major body of this research has investigated the similarity between the two tasks in terms of sensory modalities (e.g., visual vs. auditory). Typically, the results revealed increased performance costs when the two tasks posed the same processing demands (e.g., Latorella, 1998; for a review: Lu et al., 2013). More specifically, it has been suggested that this sensory-modality effect results mainly from a cross-modality advantage of maintaining relevant cues for resuming the primary task during the interruption (Ratwani et al., 2008). This would imply that the sensory modality of cues or other strategies (e.g., visual or auditory cues, verbal rehearsal) employed in re-activating the primary task during and after the interruption should be taken into account when assessing the possible performance consequences of interruptions.

Besides the similarity with respect to sensory modalities, only few studies have examined other aspects of similarity between a primary task and an interruption task in relation to resumption costs (Gillie & Broadbent, 1989; Lee & Duffy, 2015; Ratwani, 2004). The results suggest that similarity of processing codes (verbal vs. spatial) between a primary task and an interruption task might be relevant as well. For example, in a series of experiments done by Ratwani (2004) participants were interrupted with either a spatial (mental rotation) or a verbal (arithmetic) interruption task while conducting different types of primary tasks that pose a high spatial demand. The results revealed significantly longer resumption times after the spatial than after the verbal interruptions. A similar explanation could account for the findings obtained by Monk and colleagues (2008, Exp. 3). They found a verbal n-back interruption task to be more disruptive for resuming a (verbal) VCR programming task than a perceptual-motor interruption task (tracking), not involving any verbal demands. Even though the authors themselves relate this effect to the difference in general memory demands, different degrees of similarity in terms of processing codes also could have played a role here.

The obtained effects of similarity could not be accommodated easily by the major theories in interruption research. Namely, neither the MfG model nor the prospective memory model make explicit assumptions regarding the performance consequences of similarity in terms of processing codes of the tasks. While this could potentially imply that goals and intentions of the tasks are retained in a modality-free form, processing status and performance consequences of different environmental cues and verbal rehearsal needed for maintaining task goals during an interruption remain unclear within these theories. However, the theoretical view that maintaining task goals is an effortful process requiring cognitive resources might be compatible with the obtained effects, if assumed that separate cognitive resources are involved in processing of verbal and spatial material. Such a distinction is suggested by theories proposing different modality-specific working-memory domains, e.g., a phonological loop and a visual scratch-pad introduced by Baddeley and Hitch (1974), or related theories more generally distinguishing between different resources for verbal and spatial processing (e.g. the Multiple Resources Theory, MRT, Wickens, 2002; 2008). According to these models, and given that all other task aspects are equal, two tasks demanding the same cognitive processing codes (e.g., both verbal) should interfere more with each other when performed concurrently, compared to the situation when they use different codes (e.g., verbal vs. spatial). Transferred to the case of interruptions, this should result in greater interference between the primary and interruption tasks and higher resumption costs after the interruption, if both tasks use the same in contrast to different processing codes. Furthermore, both theories mentioned above would predict an interaction between the similarity of processing codes used by two tasks and a variation in processing complexity. For example, according to the multi-component working-memory model, a variation of the complexity of the memory demands of the interruption task would be expected to affect the performance in the primary task only if both tasks demand the same working-memory domains. In a similar way, the MRT would predict an increased interference between the tasks only if the variation in complexity of the interruption task would affect the demands on processing codes shared by both tasks. In the primary task performance, the increased interference effect should be reflected in larger resumption costs when returning to the primary task after a complex compared to a simple interruption task. However, no such differentiated effects of a complexity variation in the interruption task should emerge if the processing codes affected by the complexity variation are different than those used by the primary task (difficulty insensitivity, Wickens, 1984).

Based on these considerations, the current research includes two experiments, addressing the role of processing-code similarity between a primary task and an interruption task as reflected in the performance consequences of interruptions. The primary task used in this research was the WORTKLAU task (Radović & Manzey, 2019). This task represents a German adaptation of the UNRAVEL task, which has been introduced by Altmann and colleagues (2014). It can be considered as an abstract simulation of a procedural task with sequential constraints, similar to the tasks in aviation or medicine that require performing an exactly prescribed procedure from memory (e.g., Au, 2005). In each trial of this task, participants are presented with a complex visual stimulus, which has to be responded to by a sequence of steps. At each step a binary decision about a certain property of the stimulus had to be made and indicated by providing an input at a standard keyboard. Following the same logic of the UNRAVEL task, the term WORTKLAU represents a mnemonic acronym that supports learning the steps and their prescribed order. However, compared to the UNRAVEL tasks the length of the sequence has been extended from seven to eight steps. With respect to its basic cognitive demands, the WORTKLAU task can be considered as a memory-based task which relies heavily on verbal rehearsal. As reported by a vast majority of participants in our previous research (Radović & Manzey, 2019), verbal rehearsal is employed in the learning phase for memorizing the task, during the uninterrupted execution, and, most importantly, also during the interruption task for memorizing the critical step in the primary task where they left off.

During the experiments, participants performed multiple trials of this task and were repeatedly interrupted between single steps by an interruption task that also demanded working memory. The aim of the first experiment was to investigate the effects of complexity and processing codes (verbal vs. spatial) of interruption tasks, as well as possible interaction effects between these factors on resumption costs, reflected in time losses or mistakes when resuming the task after the interruption. While this experiment confirmed the effects of complexity, the results were not completely conclusive with respect to the effect of processing codes. Thus, a second experiment was conducted. Here, we investigated the possibility that the effects of similarity in processing codes between the primary and interruption task might have been masked by the fact that both, the primary task and the interruption task, required a sort of serial memory, i.e., memorization of order of items involved in the tasks. In the second experiment two classification tasks were used as interruption tasks, differing in processing code demands. However, again no effects of processing code were found on resumption performance in the primary task after an interruption.

Experiment 1

In the first experiment participants performed the WORTKLAU task and were repeatedly interrupted between single steps with one of four possible memory tasks. The interruption tasks differed in terms of complexity (complex vs. simple) and in terms of processing codes (verbal vs. spatial). Based on the previous research and the theoretical considerations presented above, we, first, expected that complex interruptions lead to longer resumption times and more errors at the first step of the primary task after the interruption. Second, we expected that interruption tasks using the same processing code as the primary task (i.e., verbal), would lead to longer resumption times and more errors at the first step after the interruption in comparison to a dissimilar (i.e., spatial) interruption task. This hypothesis was based on the assumption that verbal and spatial tasks occupy separate cognitive resources, as proposed by the MRT (Wickens, 2002) and working memory models (e.g., Baddeley, 1992). Thus, it would be expected that the verbal rehearsal of goals of the primary task would suffer greater interference from a verbal interruption than from a spatial interruption, as they both pose a demand on the same cognitive resources, which are limited. Finally, an interaction between complexity and similarity factors was also expected. That is, the effect of raising the complexity of memory demands of the interruption task on resumption time and post-interruption sequence errors was expected to emerge for the verbal n-back interruption task, but not for the spatial one.

Participants

44 university students (26 female; *M* = 24.64 years, *SE* = 0.27) took part in the study for monetary compensation or a course credit. A total sample size of 21 participants was calculated using G-power sample size calculator (Faul et al., 2007; 2009) for within-subject ANOVA with four measurements, $p = .05$, power of .95, and partial eta-square (η^2) .10 for an interaction between the factors. As the previous studies did not examine an interaction between complexity and processing code factors, the particular value of $\eta^2 = .10$ was chosen as a relatively small effect size compared to the ones reported in the previous interruption research (e.g., Monk et al., 2008), which however remains theoretically relevant to examine. Thus, the chosen sample size should be sufficient even if single participants have to be excluded based on their performance or other data issues.

Tasks

Primary Task. The WORTKLAU task (Radović & Manzey, 2019) was used as the primary task. A stimulus in this task represents a specific visual pattern consisting of a dot, a number (1, 2, 8, or 9), a letter (A, B, U, or X), and a box. An example is shown in Figure 1. The stimulus can vary in eight different features: the color of the dot (white or black), the font style of the number (underlined or not), the color of the number/letter (red or blue), the position of the number/letter (below or above the box), the sound of the letter (consonant or vowel), the font style of the box (dotted or lined), the position of the letter in alphabet (near the beginning or the end), and the parity of the number (odd or even). In each trial, participants are required to make a set of eight binary decisions in response to this stimulus, each regarding one feature of a complex visual stimulus. These decisions have to be made in a prescribed order and manual responses provided by pressing a key on a standard keyboard. As a mnemonic acronym the term WORTKLAU (literal translation from German: Word-theft) is provided, in order to support learning and correct execution of the primary task. This acronym is derived from the first letters of one possible set of decision alternatives. The choice rules, the corresponding responses, and the association with the acronym are shown in Table 1.

Figure 1. Two examples of task stimuli of the primary WORTKLAU task

Table 1. List of steps, choice rules and possible answers in the WORTKLAU task translated from German to English. Possible answers that form the acronym are provided in both German (direct link to the acronym by first letter of one of the alternatives) and English.

Step	Choice rules	Possible responses to be typed in the keyboard	
1	Dot is white or black	**W** (*w*eiss/White)	S (*s*chwarz / black)
2	Letter is without or with line	**O** (*o*hne/without)	M (*m*it / with)
3	Letter/number is red or blue	**R** (*r*ot / red)	B (*b*lau / blue)
4	Letter/number is below or above the box	**T** (*t*iefer / below)	H (*h*oher / above)
5	Letter is consonant or vowel	**K** (*K*onsonant / consonant)	V (*V*okal / vowel)
6	Box consists of lines or dots	**L** (*L*inien / lines)	P (*P*unkten / dots)
7	Letter is at the beginning or end of alphabet	**A** (*A*nfang / beginning)	E (*E*nde / end)
8	Number is odd or even	**U** (*u*ngerade / odd)	G (*g*erade / even)

Interruption tasks. Four different n-back tasks (Kirchner, 1958) were included in the research as interruption tasks. The task demands varied in two dimensions. The first dimension was complexity, operationally defined by 1-back and 2-back tasks. In the 1-back task, participants need to respond when a presented stimulus is the same as the one presented at one place before, i.e., preceded the current stimulus. In the 2-back task, participants need to respond when the presented stimulus is the same as the one presented two places before. The second dimension represented the processing code of the task (verbal vs. spatial), varied through different types of stimuli included in the n-back tasks. While the verbal n-back tasks involved numbers as stimuli (3, 4, 5, 6, and 7), the spatial stimuli included five abstract spatial patterns which are presented in Figure 2. In all n-back tasks, participants are presented with a series of stimuli presented visually, one at the time, in the middle of the screen.

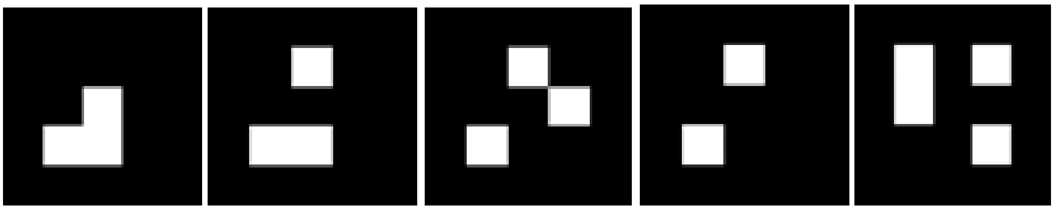

Figure 2. Spatial stimuli presented in n-back tasks

Apparatus and Stimuli

The experimental stimuli were presented on an Acer LCD screen (1920 x 1200 px, sampling with 60 Hz). In the primary WORTKLAU task, stimuli were presented in the center of the screen on white background (RGB (244, 244, 244)). Size of the complex stimulus was 144 x 190 px in total, where a letter, a number (36 px in size), a box, and a dot were presented on white background (RGB (255, 255, 255)). Participants responded by pressing a key on a standard keyboard. The set of possible answers in the primary task is presented in Table 1. In the interruption task, white stimuli were presented on black background (RGB (0, 0, 0)) one at the time in the center of the screen. The size of the stimuli in interruption tasks was 179 x 179 px, and they were presented in a way that the view onto the stimulus of the primary task was fully blocked. Participants responded by simultaneously pressing two shift-keys on the keyboard using index fingers. Stimulus presentation and response recording were controlled by a custom-made JAVA software running on an Intel Pentium (2.9 GHz, 8 GB RAM; Windows 7 Pro).

Procedure

Ethical approval for this study was obtained from the ethics committee, Intitut für Psychologie und Arbeitswissenschaften (IPA), Technische Universitaet Berlin. Participants were tested individually in two sessions (approx. 90 minutes each), distributed over two consecutive days in the *Human Performance Laboratory* of the Chair of Work, Engineering and Organizational Psychology at Technische Universitaet Berlin. On the first day, participants first signed an informed consent and a demographic questionnaire (age, gender, and self-assessment of their typing skills). Then they were introduced to the primary WORTKLAU task. After familiarization with the list of choice rules of the task, their order and possible responses, a short practice of three WORTKLAU trials followed with an immediate feedback on accuracy after each step. As defined here, a WORTKLAU trial corresponds to a complete pass through the eight-step WORTKLAU sequence, with each step requiring a binary decision about a feature of the given stimulus. Afterwards, participants were introduced to the verbal and spatial n-back tasks. Half of participants were introduced to the simple (1-back) tasks at the first session (on the first testing day) and to the complex (2-back) ones on the second session, while it was vice versa for the other half of participants. This was done to avoid potential forgetting or mixing up the instructions of 1- and 2-back tasks on the same day. After this introduction of the interruption tasks, three practice trials had to be performed with each of the two versions of the tasks (verbal, spatial). Upon completing this introduction and familiarization part, a knowledge test addressing the procedure and choice rules of the WORTKLAU task was administered. If participants answered any of the questions wrong, they would receive negative feedback and get additional time to familiarize with the procedure and choice rules, before doing another knowledge test. Upon passing the knowledge test, a short practice phase consisting of 10 trials followed. During this practice, four trials of the primary task got interrupted by each version (verbal, spatial) of the interruption task, while the remaining two trials were uninterrupted. After this final practice phase and a short break (2 min), the data collection started, which consisted of two experimental blocks separated by a short break. Each of these experimental blocks were equally structured in three parts. The first part included a first baseline (Baseline 1) of that n-back task, used as the interruption task in the following second part of the given block. This baseline data collection consisted of three n-back trials, each lasting 30s. During one trial, a total of 20 items were presented for 500 ms each, with an inter-stimulus interval of 1000 ms. The second part of the experimental block included six uninterrupted and 20 interrupted WORTKLAU trials, presented in a random order. Each interruption lasted for 30s and could occur with equal probability at five different positions within the WORTKLAU trial (before steps R, T, K, L, A). When interrupted, the stimuli of the interruption task replaced the stimulus of the primary task immediately and completely. After 30s, the stimulus of the primary task reappeared right away, and participants were instructed to resume the primary task as quickly as possible by answering the step immediately following the last step performed before the interruption and to continue from there until the last step was completed. After the last step of the WORTKLAU trial was completed, the stimulus disappeared, and the stimulus of the next trial was presented after a brief inter-stimulus interval of 300 ms. Finally, the third part of each experimental block included another 1.5 min baseline performance of the given n-back task (Baseline 2), corresponding to the first baseline assessment. While both experimental blocks were equally structured, they differed with respect to the type of n-back task used as the interruption task. Namely, the n-back task was always of the same complexity (either 1-back or 2-back task), but different in terms of processing code (spatial vs. verbal). The order of the two experimental blocks within each session was counterbalanced between participants.

On the next day, the second experimental session took place. It started with a reminder of the list of choice rules of the WORTKLAU task, and three refresher trials of this task with an immediate feedback after each step. Then, participants were introduced to the spatial and verbal versions of the interruption task used in this session (e.g., 2-back if 1-back tasks were used in the first session), practice phase and two experimental blocks had the same structure as in the first session. A complete timeline of the experimental procedure for both experimental sessions is presented in Figure 3.

At the end, a structured interview with the participants was conducted. In each of these interviews, the participants were asked in the form of an open question what specific strategies they had used to complete the different tasks. In addition, they were provided with a list of possible strategies to memorize the interrupted position of a WORTKLAU sequence during the interruption phase and asked to mark the ones they actually had employed (e.g., rehearsal of a letter; rehearsal of a word; numbering; visualization; using fingers; other – what?). The main purpose of these interviews was to identify participants, who had employed other strategies than the presumed ones for performing the task during the interruption phase. Such “unwanted” strategies typically involved verbalizations used for the spatial interruption tasks (e.g., naming of stimuli) and strategies to remember a position in the primary task other than internal verbal rehearsal (e.g., keeping fingers on the respective keys or using a sort of mental imagery). Data of participants who reported to have used such strategies consistently were not included in the data analysis.

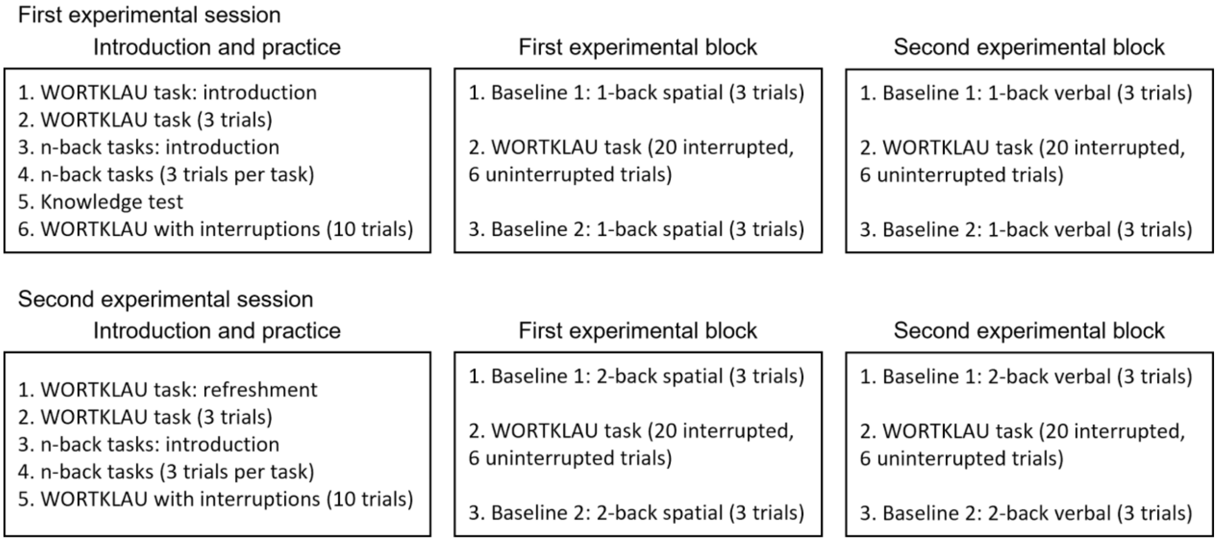

Figure 3. Timeline of the experimental procedure.

Experimental design

The design used in the present experiment for examining the effects of interruptions on the WORTKLAU task corresponded to a 2 (Complexity: 1-back vs. 2-back) x 2 (Processing code: verbal vs. spatial) factorial design with repeated measures for both factors. For investigating performance in the n-back interruption tasks, a 2 (Complexity: 1-back vs. 2-back) x 2 (Processing code: verbal vs. spatial) x 2 (Context: baseline vs. interruption) factorial design with repeated measures on for all three factors was used.

Dependent Variables

In total, four performance measures were registered, of which three performance measures were used to assess the effects of interruptions in the primary task.

Resumption time was defined as the additional time needed to return to a certain WORTKLAU step after an interruption. It was calculated by subtracting the mean response time obtained in the uninterrupted trials for a certain step from the post-interruption response time for that step. Response time for the first (“W”) step was always defined as the time passed after the appearance of the primary task stimulus until the first response. For all other steps it was defined as the length of inter-response intervals between two consecutive responses when no interruptions occurred. The post-interruption response times for steps immediately following an interruption was defined as the time that passed from the reappearance of the primary task stimulus after the interruption to the first answer provided. Mean response times were based on the correctly answered steps only.

Post-interruption sequence errors were defined as the proportion of all responses where participants deviated from the predefined order of steps in the primary task by either skipping (e.g., directly after R step answering K step) or repeating one or more steps.

Post-interruption non-sequence errors were defined as the proportion of all responses at the step where participants answered the correct step in the sequence, i.e., evaluated the correct feature of the stimulus (e.g., a color of the dot), but provided a wrong answer (e.g., white instead of black).

Performance in the interruption (n-back) tasks was assessed by the mean percentage of correct responses. Correct responses included both hits (correct key presses in response to targets), as well as correct rejections (refrain from key press in response to non-targets).

Finally, a structured debriefing interview was used to identify different strategies participants had used for conducting the primary tasks and the interruption tasks in a respective experimental block. With respect to the interruption tasks, the main purpose of this interview was to identify participants who had used some sort of verbalization strategies in the spatial interruption tasks in order to exclude them from further analyses. With respect to the primary task, the main purpose of the interview was to examine strategies used for refreshing goals of the primary task during the interruption and for resuming the task after the interruption.

Results

On an individual basis, all response times for a single WORTKLAU step which were faster than 500ms or slower than 3 SD above the individual mean of the respective condition were considered as outliers and excluded from further analyses. In addition to that, all trials in an interruption in which performance just corresponded to chance level (accuracy < 61 %) were excluded from further analyses, including all response times, sequence errors and non-sequence errors at the step which followed directly that interruption. Moreover, the entire data set of five participants who reported to have used some sort of verbalization strategy for performing the spatial n-back tasks, i.e., naming or rehearsal of spatial stimuli, were excluded from the further analyses. In addition, another participant was excluded for using fingers during the interruption task to mark where to continue the primary task. Thus, the final statistical analyses were based on 38 participants. An observed power of 0.91 was obtained using PANGEA software (Westfall, 2015) for 20 observations per factor level, 38 participants, who were included as a random factor, Complexity and Processing Codes as two fixed factors, and a partial eta-square (η^2) of .10 for their interaction.

Post-interruption performance in the primary task

Mean resumption times for correctly answered steps after an interruption, dependent on the complexity and the processing code of the interruption task are shown in Figure 4. The 2 (Complexity) x 2 (Processing code) ANOVA for repeated measures revealed significant main effects of Complexity, $F(1, 38) = 16.43$, $p < .001$, $\eta^2 = .307$, and Processing code, $F(1, 38) = 5.20$, $p = .028$, $\eta^2 = .123$. As it becomes apparent from Figure 4, complex interruptions led to longer resumption times than simple ones (*M* = 3730 ms, *SE* = 326 vs. *M* = 2411 ms, *SE* = 217), and verbal interruption tasks led to longer resumption times than spatial ones (*M* = 3250 ms, *SE* = 251 vs. *M* = 2890 ms, *SE* = 223). No significant effect emerged for the interaction between the two factors, $F(1, 38) = .51$, $p = .48$, $\eta^2 = .014$.

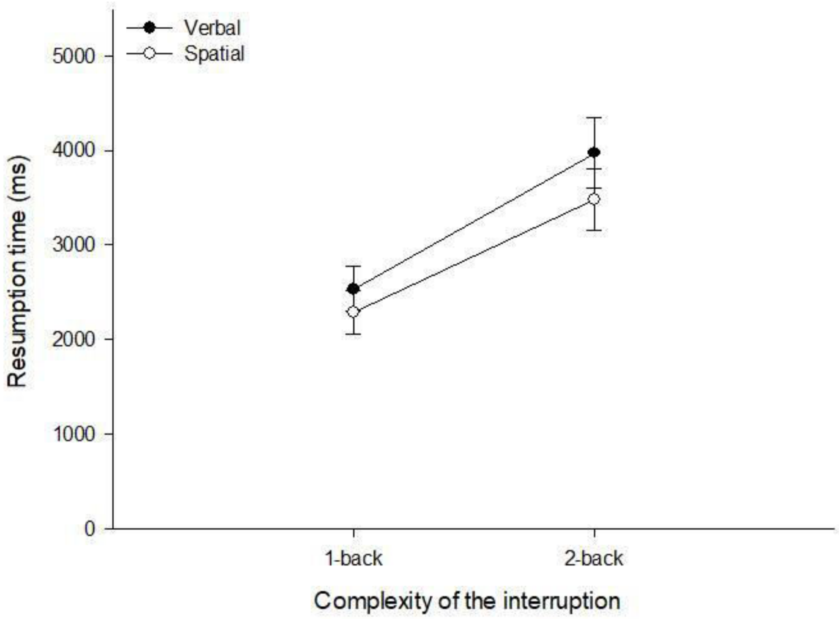

Figure 4. Mean resumption times and standard errors depending on complexity and processing code of interruption task

The corresponding effects for mean proportion of post-interruption sequence errors are presented in Figure 5. As expected, complex interruptions lead to more sequence errors at the first post-interruption step compared to the simple ones (*M* = .184, *SE* = .025 vs. *M* = .073, *SE* = .014), $F(1, 38) = 21.11$, $p < .001$, $\eta^2 = 0.36$. In contrast, neither the main effect of Processing code, $F(1, 38) = 1.53$, $p = .223$, $\eta^2 = 0.04$, nor the interaction between the factors became significant, $F(1, 38) = 2.32$, $p = .14$, $\eta^2 = 0.06$.

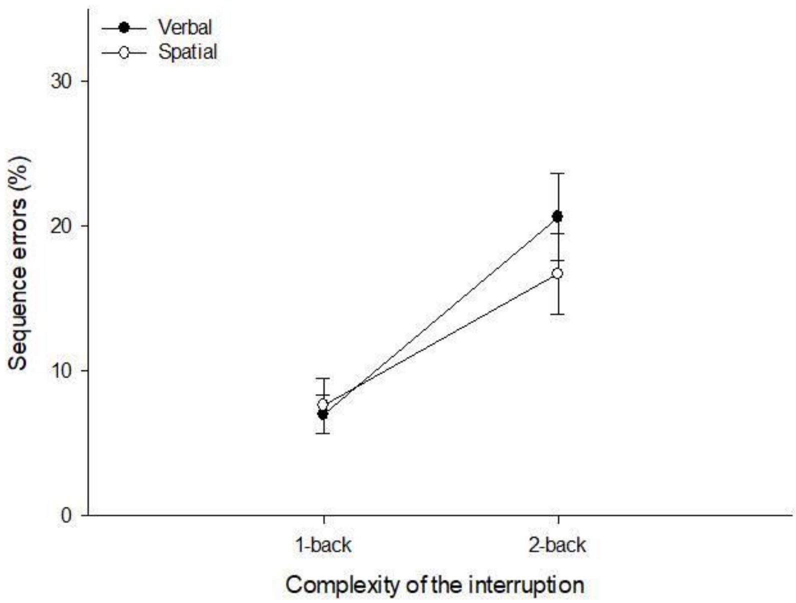

Figure 5. Mean proportion of sequence errors and standard errors depending on complexity and processing code of interruption task

When the post-interruption non-sequence errors were subjected to the same analysis, neither the main effect of Complexity, $F(1, 38) = 2.12$, $p = .153$, $\eta^2 = 0.054$ nor the main effect of Processing code, $F(1, 38) = .57$, $p = .45$, $\eta^2 = 0.015$, were found. Also, the interaction between the factors did not reach significance, $F(1, 38) = .34$, $p = .56$, $\eta^2 = 0.009$.

Performance in n-back tasks

Mean percentage of correct responses in the n-back tasks during baseline and interruption trials for the two levels of complexity and the two processing codes are shown in Figure 6. These effects were analyzed by a 2 (Complexity: 1-back, 2-back) x 2 (Processing code: verbal, spatial) x 2 (Context: baseline, interruption) ANOVA for repeated measures. This analysis revealed a significant main effect of Complexity, $F(1, 38) = 139.58$, $p < .001$, $\eta^2 = .79$, and Processing code, $F(1, 38) = 16.58$, $p < .001$, $\eta^2 = .31$, while a main effect of Context did not reach significance, $F(1, 38) = 1.83$, $p = .185$, $\eta^2 = 0.047$. In addition, a complex pattern of interactions effects emerged, including significant two-way interactions between Complexity and Processing codes, $F(1, 38) = 10.90$, $p = .002$, $\eta^2 = .228$, Complexity and Context, $F(1, 38) = 4.20$, $p = .048$, $\eta^2 = .102$, and Processing codes and Context, $F(1, 38) = 15.03$, $p < .001$, $\eta^2 = .289$, as well as a significant three-way interaction between all factors, $F(2, 38) = 6.24$, $p = .017$, $\eta^2 = .14$.

As it becomes apparent from Figure 6, these interaction effects were mainly due to the different pattern of effects emerging in the baseline context and the interruption context. Thus, the analysis was broken down in two 2 (Complexity) x 2 (Processing code) ANOVAs addressing the effects for baseline and interruption trials separately. For the baseline trials, this analysis revealed only a significant effect of Complexity, $F (1, 38) = 109.80$, $p < .001$, $\eta^2 = .75$. As becomes evident from Figure 6 (left), a significantly higher accuracy was observed in the verbal 1-back task and the spatial 1-back task (M = 98.2%, SE = 0.3 and M = 98.1%, SE = 0.4, respectively) than in the verbal 2-back task and the spatial 2-back task (M = 88.1%, SE = 1.2 and M = 88.2%, SE = 1.2, respectively), whereas a main effect of Processing code, $F(1, 38) = .003$, $p = .96$, $\eta^2 = 0$, and an interaction were not found, $F(1, 38) = .02$, $p = .90$, $\eta^2 = 0$. This suggests that the manipulation in complexity of the interruption tasks was successful.

Furthermore, it proves that the manipulation of processing codes was independent of the complexity manipulation, i.e., that the spatial and verbal versions of the n-back tasks were equivalent with respect to their general cognitive demands.

However, when the n-back tasks were used as interruption tasks, not only a significant main effect of Complexity emerged, $F(1, 38) = 156.06$, $p < .001$, $\eta^2 = .81$, but also a main effect of Processing code, $F(1, 38) = 45.53$, $p < .001$, $\eta^2 = .55$, as well as a Complexity x Processing code interaction, $F(1, 38) = 16.76$, $p < .001$, $\eta^2 = .31$. As becomes evident from Figure 6 (right), the latter effects reflect that the n-back-task performance in the interruption context was slightly better in the verbal ($M = 97.6\%$, $SE = 0.3$) than the spatial ($M = 97.0\%$, $SE = 0.5$) version of the task, and that this effect was more pronounced for the more complex version of the task ($M = 90.4\%$, $SE = 0.9$ vs. $M = 86.3\%$, $SE = 1.1$, respectively).

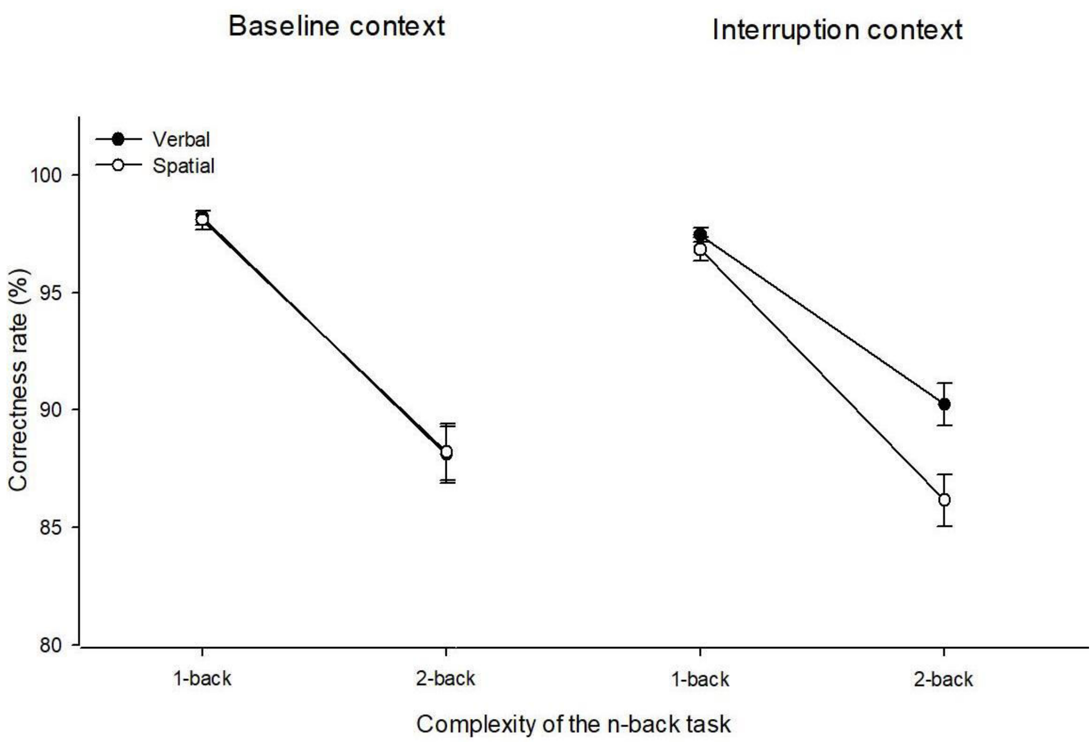

Figure 6. Mean performance accuracy (%) in n-back tasks depending on the context in which they were conducted

Discussion

The aim of the present experiment was to investigate the effects of complexity and similarity in terms of processing codes between the primary task and the interruption task, as well as the interaction between the factors on the post-interruption performance in a primary task. The primary task used was a complex verbal task with sequential constraints, comparable to working through a verbal checklist from memory. The interruption task was an n-back task varying in complexity (1-back vs. 2-back) and similarity of processing-codes (similar: verbal vs. dissimilar: spatial).

The first hypothesis regarded the effects of complexity. As expected, performing a 2-back task during the interruption phase led to higher costs in terms of response times and sequence errors when resuming the primary task compared to performing a 1-back task. The results are in line with previous research (e.g., Cades et al., 2008) and theoretical predictions derived from relevant theoretical models, i.e., the MfG model (Altmann & Trafton, 2002) and the prospective memory model (Dodhia & Dismukes, 2009; Einstein et al., 2003).

Our second main hypothesis concerned the effects of similarity in terms of processing code between a primary and an interruption task. Namely, as verbal rehearsal was repeatedly shown to be the main strategy used to resume the WORTKLAU task after an interruption (Altmann et al., 2014; Radović & Manzey, 2019), we expected that a verbal n-back interruption task would lead to longer resumption times and more sequence errors upon resumption of the WORTKLAU task than the spatial one. In addition, we expected an interaction effect in a way that the complexity of the n-back task should make a difference with respect to post-interruption costs only if it was a verbal one. However, our results are less straightforward in this respect. At first glance and just by looking at the resumption costs in the primary WORTKLAU task, it seems that the similarity in processing code indeed led to different interruption effects in the predicted direction. As expected, resumption times were significantly slower after verbal n-back interruptions compared to the spatial ones. However, no comparable effect was obtained for post-interruption sequence errors, and neither of the two performance measures indicated the expected interaction effect, except for a discernible trend in the expected direction for the sequence errors. Moreover, a straightforward interpretation of the effects of processing code of the interruption task on resumption times as indicator of a similarity effect gets challenged when looking at the performances in the n-back tasks during the interruption phases. These effects directly mirror the significant main effect of processing code on resumption times and the descriptive effect on sequence errors. More specifically, the accuracy in the verbal n-back interruption tasks was significantly higher than in the spatial n-back interruption tasks, particularly when their complexity was high. Such a difference between the n-back tasks was not found in the single-task baseline condition but emerged only when the n-back tasks had to be performed during the interruption phase. Thus, it seems that these effects do not simply reflect differences in task difficulty dependent on the different processing codes but a difference in effort invested in the tasks during the interruption phase. In other words, it seems that the participants put a higher emphasis on verbal n-back interruption tasks during the interruption phase, compared to the spatial ones, particularly when the complexity was high. This further suggests that the longer resumption times observed after verbal interruptions in the present study simply resulted from the fact that more cognitive resources were drawn from maintaining the task goals of the WORTKLAU task during the interruption phase when participants performed the verbal interruption tasks compared to the spatial ones.

A possible reason for such difference in emphasis put on the two types of interruption tasks might be found in the perceived difficulty of these tasks. Namely, most of our participants (60%) reported that they found the verbal tasks easier to perform than the spatial ones. This could have resulted in participants unconsciously being more motivated and willing to invest more effort and cognitive resources in the verbal n-back tasks during the interruptions than in the presumably more difficult - spatial ones, specifically when the complexity was high. Taken together, the pattern of results of this first experiment challenge the assumption that the similarity in processing codes of the primary task and the interruption task might make a difference for post-interruption performance. This contrasts previous research, which at least has provided some evidence suggesting that the resumption of a primary task after an interruption might also be affected by the similarity of processing codes between the primary task and the interruption task (e.g., Ratwani, 2004).

However, before we fully abandon the idea of similarity in processing codes between the primary task and the interruption task playing a role in the magnitude of resumption costs, we need to account for possible alternative explanations of the results found in this first experiment.

First, it remains possible that the WORTKLAU task did pose some sort of spatial demand which was difficult to assess just by introspection in the post-experimental interview. This could also explain why both interruption tasks led to more or less the same interference effects reflected in post-interruption performance of the primary task independent of the processing codes involved. However, such assumption would also imply that the spatial and verbal demands involved in the primary task were comparable. Given the high significance of verbal strategies involved in the WORTKLAU task that were consciously reported by all participants in our experiment, this latter assumption seems to be highly unlikely. Thus, even if some sort of spatial processing was involved in the WORTKLAU task we would still expect the verbal demands to be higher and the interference effects with the verbal interruption tasks compared to the spatial ones to be stronger, given that similarity of processing codes matters.

Second, and more important, we have to acknowledge the fact that the n-back tasks used as interruption tasks in our experiment require the memorization of the serial order of presented items. Independent of the processing codes involved, this marks another basic similarity with our primary WORTKLAU task, which also puts considerable demand on serial memory. Thus, one might argue that the specific combination of the sequential primary WORTKLAU task and sequential n-back interruption tasks might have posed a specific interference between mechanisms involved in memory for order and serial recall during the interruption phase (Burgess & Hitch, 1999; Brown et al., 2000; Botvinick & Plaut, 2006). This interference, in turn, might have masked any effects due to the similarity of processing codes. Such possibility is suggested by previous studies showing that memory for serial order indeed might represent a specific domain which operates largely independent of the specific processing codes involved in processing the stimuli (Depoorter & Vandierendonck, 2009; Hachmann et al., 2014; Jones et al., 1995). Moreover, it has been argued that this memory function is not only involved in the typical, verbal tasks used to address serial order memory, but generally all tasks which require learning and recognition of order of items, or reproduction of order of movements (Hurlstone et al., 2014; Logan, 2021), which is also the case with performing any procedural tasks with sequential constraints. Admittedly, our assumption that n-back tasks could also rely on the short-term memory for serial order is somewhat speculative, as it was not examined in this context so far and was not considered as such in any previous theoretical considerations. However, to exclude the possibility that this might have masked the effects of processing-code similarity, we run a second experiment. In this experiment, we used the same primary task. However, the interruption tasks used differed only with respect to their processing code (verbal vs. spatial) but did not involve any serial-memory demand.

Experiment 2

In this experiment participants again performed the WORTKLAU task and were repeatedly interrupted between single steps. However, now the interruption task was one of two versions of a classification task. Whereas the verbal version required to classify pairs of figures as being smaller or bigger than five, the spatial one involved a comparison of three-dimensional shapes as being identical or mirror images of each other. Thus, these tasks differed with respect to the processing codes involved, but neither of them requested any sort of serial-memory function. Given that the similarity of processing codes involved in maintaining primary task goals during an interruption phase and the interruption task would make a difference, we assumed that the differences between verbal and spatial conditions now would be able to emerge. That is, due to higher interference of verbal interruptions with verbal rehearsal of cues of the primary task during the interruption, it is expected that the verbal classification task would cause higher resumption costs that the spatial one.

Method

Participants

47 University students (28 female; M = 24.06 years, SD = 2.29) took part in the study for monetary compensation or a course credit. A total sample size of 32 participants was calculated using G-power sample size calculator (Faul et al., 2007; 2009) for within-subject ANOVA with two measurements, p = .05, power of .95, and η^2 = .15 for the main effect of

Processing codes. The value of η^2 was chosen based on the results of Experiment 1, as we now expect to find an effect of that size or greater.

Tasks

Primary task. The same WORTKLAU task as in Experiment 1 was used as the primary task.

Interruption tasks. Two classification tasks were included as interruption tasks, differing in their processing code. In the verbal classification task, participants were presented with a series of pairs of figures (from the set: 3, 4, 6, 7; see examples of stimuli in Figure 7). Participants were required to respond to a pair whenever both figures were smaller or bigger than five (e.g, Figure 7, first row) but refrain from responding if only one of the numbers was smaller than five (e.g., Figure 7, second row). In the spatial classification task, a series of pairs of two differently rotated 3-D figures were presented. Subjects were required to respond if the figures were identical (see Figure 7, third row) and to refrain from responding if the figures were mirrored versions of each other (Figure 7, fourth row).

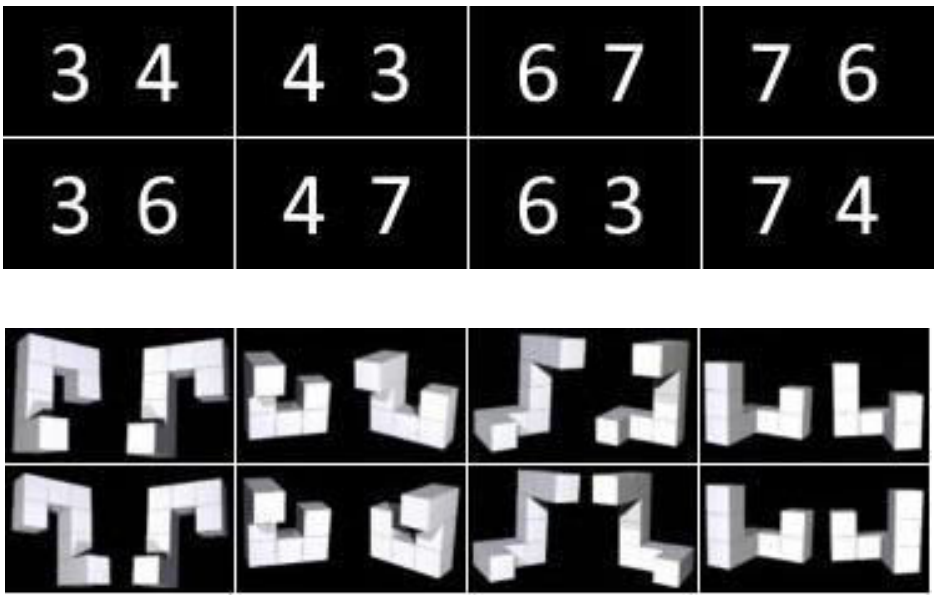

Figure 7. Stimulus sets used in the Experiment 2.

Apparatus and Stimuli

Stimulus presentation and response recording were controlled by a custom-made JAVA software running on an Intel Pentium (2.9 GHz, 8 GB RAM; Windows 7 Pro). Presentation of the stimuli and recording of the responses of the WORTKLAU task were the same as in the first experiment. For the classification tasks, white stimuli (RGB (255, 255, 255)) were presented on black background (RGB (0, 0, 0)) one at a time in the center of the screen. Size of the stimuli in interruption tasks was 92 x 63 px. Participants responded by simultaneously pressing two shift-keys on the keyboard using index fingers. During the presentation of stimuli of the interruption tasks, the view on the stimulus display of the primary WORTKLAU task was not available.

Procedure

Ethical approval for this study was obtained from the ethics committee, Intitut für Psychologie und Arbeitswissenschaften (IPA), Technische Universitaet Berlin. Participants were tested individually in one session (approx. 90 minutes) in the Human Performance Laboratory of the Chair of Work, Engineering and Organizational Psychology at Technische Universitaet Berlin. All procedural details concerning the length and structure of familiarization, practice trials as well as experimental blocks corresponded completely to the first testing day of Experiment 1. Only three details concerning the use of tasks, the presentation of task stimuli, and the way of collecting information from the participants after the experiment (debriefing) were different. The first difference was the type of interruption task, i.e., the verbal n-back task was replaced with the verbal classification task, and the spatial n-back with the spatial classification task. The second difference concerned the presentation time of stimuli and the length of the interstimulus interval (ISI) in the classification tasks. The presentation time of each pair of figures or 3D patterns was 750 ms long, followed by a 750 ms ISI. This resulted in the presentation rate of 20 stimuli over a 30s period (one trial length), just as in the n-back tasks in the first experiment. Finally, the third difference involved using a short debriefing survey instead of an interview at the end of the experiment addressing the strategies applied for performing the tasks in the experimental blocks. The survey was an adaptation of the structured debriefing interview of Experiment 1 and was used instead in order to reduce contact with participants during the COVID-19 pandemic.

Experimental design

A one-factor (Processing code: verbal vs. spatial) within-subjects design was used. The design for investigating performance in the interruption tasks involved a 2 (Processing code: verbal vs. spatial) x 2 (Context: baseline vs. interruption) within-subjects design.

Dependent Variables

In total, four performance measures were registered. Three measures were calculated to assess the resumption performance in the primary task after an interruption (resumption time, post-interruption sequence errors, post-interruption non-sequence errors) and were defined in the same way as in the first experiment.

Performances in interruption (classification) tasks were expressed by the percentage of correct responses in the classification tasks. The percentage of correct responses was calculated using both, hits (correct key presses in response to stimuli that fulfill the classification criterion), as well as correct rejections (refrain from key press in response to stimuli that do not fulfill the classification criterion).

Results

On an individual basis, all response time faster than 500ms slower than 3 SD above the individual mean of the respective condition were considered as outliers and excluded from further analyses. In addition to that, all interruption trials in which low performance was measured (accuracy < 61 %) were excluded from further analyses, including all response times, sequence, and non-sequence errors in the interrupted WORTKLAU task which followed that interruption task. In addition, the full data sets of four participants who reported to have used fingers as a strategy to cope with interruptions were excluded from further analyses. Finally, the data sets of two other participants had to be excluded because they did not understand the instructions of the experiment, resulting in a considerable number of missing data either in the main task or in the classification tasks. As in the previous experiment, all our participants reported using verbal rehearsal of either the last step executed before the interruption took place or the one that should take place after the interruption in order to refresh the primary task goal during the interruption task. Thus, the statistical analyses were based on 41 participants. An observed power of 0.87 was obtained using PANGEA software (Westfall, 2015) for 20 observations per factor level, 41 participants, who were included as a random factor, Processing Codes as a fixed factor, and for $\eta^2 = .15$ for the main effect.

Post-interruption performance in the primary task

A t-test for paired samples did not reveal significant effect of Processing code on the resumption time, $t(41) = .83$, $p = .43$, $\eta^2 = .017$, the post-interruption sequence error rates, $t(41) = .23$, $p = .82$, $\eta^2 = .001$, or the non-sequence error rates, $t(41) = .98$, $p = .33$, $\eta^2 = .024$.

For examining this null effect further, a Bayesian paired sample t-test with a default Cauchy prior was conducted for resumption times. A two-sided analysis revealed a Bayes factor (BF01) suggesting the data were 4.27 times more likely under the null hypothesis than the alternative hypothesis median effect size of -0.12. The same analysis for the post-interruption sequence errors revealed Bayes factor (BF01) suggesting the data were 5.73 times more likely under the null hypothesis than the alternative hypothesis median effect size of 0.03. For the post-interruption non-sequence errors similar results were obtained, that is, the data were 3.7 (BF01) times more likely under the null hypothesis than the alternative hypothesis.

Performance in the classification tasks

Performance in the classification task was analyzed by a 2 (Processing code: verbal vs. spatial) x 2 (Context: baseline vs. interruption) ANOVA for repeated measures. This analysis revealed a significant main effect of Context, $F(1, 41) = 19.60$, $p < .001$, $\eta^2 = .33$, that is, mean performances were somewhat better in the baseline context than in the interruption context ($M = 95.3\%$, $SE = .60$ vs. $M = 93.6\%$, $SE = .60$). However, this effect emerged independently of processing codes involved in the classification tasks. Neither the main effect of Processing code, $F(1, 41) = 3.54$, $p = .067$, $\eta^2 = .08$, nor the interaction between the factors, $F(1, 41) = .006$, $p = .94$, $\eta^2 = 0$, reached the level of significance.

Exploratory analysis

In order to test our initial assumption that the similarity between the primary task and the interruption task in terms of sequential constraint had an impact on the post-interruption performance, the data of the two experiments were formally compared in an exploratory analysis. As the performance in 1-back tasks in Experiment 1 was comparable with the performance in classification tasks in Experiment 2 (*M* = 97.5%, *SE* = .7 vs. *M* = 94.4%, *SE* = .5, respectively), only the data from these experimental blocks was included in the analysis. For controlling the differences in experience with the tasks, a subgroup of 21 participants who performed the 1-back tasks on the first testing day in Experiment 1 was selected, while all 41 participants from Experiment 2 were included in the analyses. In total, data from 62 participants were reanalyzed for the purpose of this exploratory study.

Mean resumption times for the first step after the interruption for the sequential and non-sequential interruption tasks are shown in Figure 8. Resumption times were analyzed using an independent sample t-test with Task type (sequential vs. non-sequential) as a factor. The analysis revealed significant main effect of Task type, $t(62) = 2.52$, $p = .014$, $\eta^2 = .096$. As it becomes obvious from Figure 8, longer post-interruption resumption times emerged after the sequential interruption tasks compared to the non-sequential interruption tasks (*M* = 2832 ms, *SE* = 266 vs. *M* = 2006 ms, *SE* = 190). However, when the post-interruption sequence error rates were subjected to the same analysis, neither significant main effects nor interactions between the factors emerged, all $p > .73$. The same results were obtained in the analysis of the non-sequence error rates, all $p > .47$.

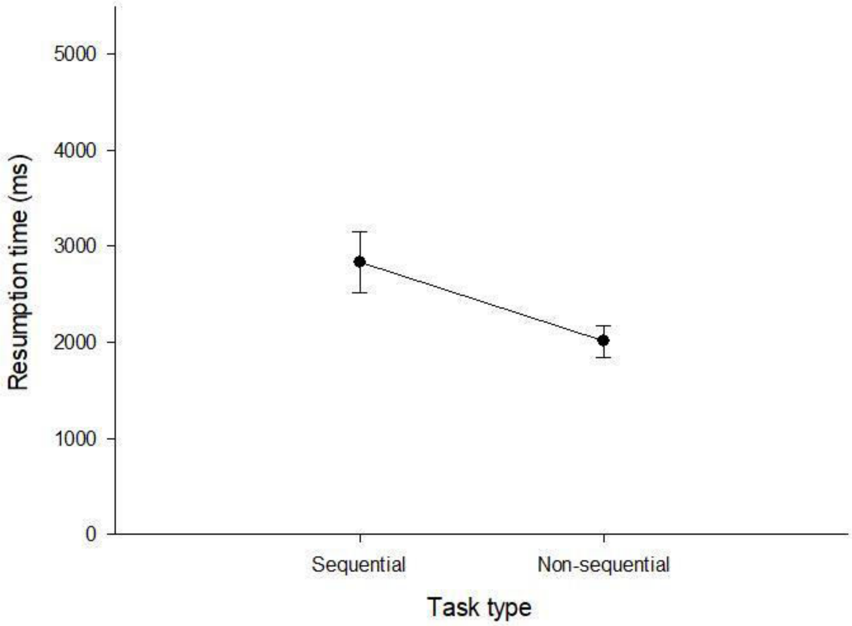

Figure 8. Mean resumption times and standard errors depending on the type of interruption task.

Discussion

In this experiment, we accounted for the possibility that the effects of verbal and spatial processing codes involved in the interruption task were not able to emerge unambiguously in Experiment 1, due to the structural similarity between the primary task and the interruption task in terms of sequential constraints. For that reason, in this experiment we administered classification tasks with simultaneous presentation of two either verbal or spatial stimuli as interruption tasks. However, the data again did not support our hypothesis. Despite all our participants reported again engaging in verbal rehearsal during the interruption task to maintain the cue needed for the resumption of the primary task, the verbal classification task did not impair the post-interruption performance more than the spatial one. This was not supported just by the finding of a non-significant effect, but also by our additional Bayesian analysis, which revealed a null effect as much more likely than an effect of processing codes, given our data. Thus, overall, the results converge with the data from the first experiment and provide no evidence that the similarity in terms of processing codes between the primary task and the interruption task plays a role when resuming the primary task after the interruption. However, the exploratory analysis contrasting the resumption costs after the interruptions in Experiment 1 with the ones in this second experiment showed that participants needed more time to resume the primary task after an interruption, which also involved a sort of sequential constraint, i.e., the 1-back task, compared to a non-sequential interruption, i.e., classification task. This result confirmed our speculation about the significance of a serial order memory or a placekeeping mechanism shared by the primary task and interruption task and will be further discussed in the General discussion.

General Discussion

The main goal of the present study was to investigate the effects of complexity of the interruption task, the effects of similarity between the primary and the interruption task in terms of processing codes, and the interaction between these factors. In addition, an exploratory analysis comparing the two experiments hinted that an aspect of structural similarity between the primary task and the interruption task could potentially play a role in magnitude of resumption costs. To the best of our knowledge, this is the first study that has addressed these questions directly and in a systematic way.

Let us first consider the effects of complexity of the interruption task on resumption times and sequence errors at the first post-interruption step of the primary task. It was expected that the complex interruption task will lead to longer resumption times and more sequence errors at the first step after the interruption compared to the simple interruption. In accordance with this hypothesis, our results revealed that complex interruptions led to greater interference when resuming the primary task compared to the simple ones in both resumption times and sequence errors. This finding is in line with the previous research (e.g., Cades et al., 2008; Hodgetts & Jones, 2006), which also have shown that resumption costs increase with higher complexity of an interruption task. The results can be interpreted within the MfG model (Altmann & Trafton, 2002) or the resource-theoretical view, which conceptualizes interruptions in terms of a prospective memory model (Dodhia & Dismukes, 2009; Einstein et al., 2003). Both theoretical perspectives predict greater resumption costs in a primary task as the complexity of an interruption task increases. The MfG model predicts higher costs in terms of additional time needed to re-activate goals of the primary task after more complex interruptions, while the prospective memory model predicts increased error rates. Both predictions were indeed confirmed in our research with 2-back interruptions compared to 1-back interruption tasks.

Taken together, the results of the present study suggest that the effects of interruption complexity can also be extended to the procedural tasks with sequential constraints as simulated by the WORTKLAU task. The finding that only post-interruption resumption times and sequence errors were affected suggests that the interruptions specifically interfered with the maintenance of goals of the primary task and did not cause a general impairment of the performance. The latter would be reflected in increased rates of post-interruption non-sequence errors, which was not the case in our study. Such specific effects of interruptions were reported previously in studies on interruption length with the UNRAVEL (Altmann et al., 2014) and the WORTKLAU tasks (Radović & Manzey, 2019).

When it comes to the effects of similarity between the primary and interruption tasks, we expected that a predominantly verbal primary task would be more negatively affected by an interruption task involving verbal stimuli, leading to longer resumption times and more sequence errors upon resumption compared to a spatial interruption task. Taken together, the results of the two experiments did not provide support for this hypothesis. This finding stands in odds with predictions of the MRT (Wickens, 2002) and working memory models (e.g., Baddeley, 1992), which would propose greater interference between verbal interruption task and verbal rehearsal necessary to re-activate the primary task. This suggests that maintaining the goals of a procedural primary task as simulated by the WORTKLAU tasks during an interruption depends on general memory processes, which operate largely independent of specific processing codes or specific memory subsystems. Possible mechanisms involved might be a specific placekeeping mechanism, which monitors the progress within a task by keeping track of of completed and incoming steps (Burgoyne et al., 2020; Carlson & Cassenti, 2004; Trafton et al., 2011; Hambrick & Altmann, 2015) or a modality-independent serial order memory, as proposed by Depoorter and Vandierendonck (2009), and Jones and colleagues (1995). Such an assumption could also explain our finding of higher resumption costs when the WORTKLAU task was interrupted by an n-back task, which also involves keeping track of item sequences, compared to the interruption by a classification task without such demands. This suggests that the similarity between a primary and an interruption tasks in terms of such demands are more important than just a similarity of processing codes. Cooper et al. (2018) have already raised a similar suggestion concerning the similarity of a placekeeping process but, as to the best of our knowledge, the current study is one of the first to provide empirical support for such an effect.

Having provided evidence for interruption effects being independent of the similarity of processing codes involved in the primary task and interruptions task, the question remains to what extent this effect is generalizable also to non-verbal primary tasks. Although somewhat unlikely, recent results presented by Helton and Russell (2011, 2015, 2017) suggest that effects of similarity of processing codes on interruptions tasks might be different for verbal versus spatial primary tasks[1]. Specifically, they investigated how verbal and spatial tasks interfered with a primary vigilance task involving verbal (Helton & Russell, 2011) and spatial processes (Helton & Russell, 2015, 2017), respectively. The paradigms used at least resembled interruption paradigms, even though it simulated more a sort of additional load and of breaks offered as a countermeasure for the monotony of the vigilance task. In line with our results, the authors did not find any different effects of the processing codes involved in the “interruption” tasks on a *verbal* vigilance task (Helton & Russell, 2011). However, the performance in a *spatial* vigilance task was somewhat more disrupted by the spatial compared to the verbal interruption tasks (Helton & Russell, 2017). Ratwani (2004) has also reported stronger disruptive effects of spatial interruption tasks compared to verbal ones on a spatial primary task. Nevertheless, the current evidence of these different effects is weak and additional load or breaks during vigilance tasks might be a different story than typical interruptions. In any case, more research would be needed to substantiate such a possible difference between verbal and spatial primary tasks.

Limitations of the current study include the typical limitations of laboratory studies. Our participants were university students, who might be an already highly selected population

[1] We thank one of our reviewers for making us aware of this research

with respect to the level of their cognitive capabilities. However, this could imply that the effects of interruption complexity and structural similarity between the tasks would be even larger in the general population. In addition, the WORTKLAU task used in our research to simulate a procedural task with sequential constraints certainly is an abstract laboratory task, thus we just assume that the cognitive demands of this task closely resemble the ones needed in many procedural tasks in everyday environments and applied settings. Moreover, the consequences of committing errors in task execution were not quite comparable to typical tasks outside the laboratory. Thus, further research should show whether the effects found in this research can be replicated with more representative samples and more realistic tasks in relevant field settings.

Conclusion

To the best of our knowledge, the current study is one of the first attempts to examine both the effects of complexity of an interruption task and the effects of similarity in processing codes between a procedural primary task with sequential constraints and the interruption task on the post-interruption resumption costs. The results provide evidence for a large and consistent effect of complexity on the resumption costs through increased interference with goals of the primary task in working memory. Furthermore, the data do not provide support for a role of similarity between the primary and the interruption in terms of the processing code in resumption costs in a complex primary task with sequential constraints after an interruption. At the same time, the study provides preliminary evidence that the similarity between the tasks in terms of their sequential structure amplifies post-interruption resumption times, which is a novel finding in the domain of interruption research that should be consolidated in future research.

Authors note:

The data are made available under the Creative Commons CC0 public domain dedication at OpenScienceFramework at https://osf.io/7b3z4/?view_only=efb167c0f5674f8ebadac7d0faf287f9

Literature

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, *26*(1), 39-83. https://doi.org/10.1207/s15516709cog2601_2

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General, 143*(1), 215–226. https://doi.org/10.1037/a0030986

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2017). Effects of interruption length on procedural errors. *Journal of Experimental Psychology: Applied, 23*(2), 216–229. https://doi.org/10.1037/xap0000117

Au, H. (2005). Line pilot performance of memory items. In *2005 International Symposium on Aviation Psychology*, 29-35. https://corescholar.libraries.wright.edu/isap_2005/140

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556-559. DOI:10.1126/science.1736359

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47-89. https://doi.org/10.1016/S0079-7421(08)60452-1

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*(2), 201-233. https://doi.org/10.1037/0033-295X.113.2.201

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181. https://doi.org/10.1037/0033-295X.107.1.127

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581. https://doi.org/10.1037/0033-295X.106.3.551

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2021). Incremental validity of placekeeping as a predictor of multitasking. *Psychological Research*, *85*(4), 1515-1528. https://doi.org/10.1007/s00426-020-01348-7

Cades, D. M., Davis, D. A. B., Trafton, J. G., & Monk, C. A. (2007). Does the difficulty of an interruption affect our ability to resume? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(4), 234-238. https://doi.org/10.1177/154193120705100419

Cades, D. M., Werner, N., Boehm-Davis, D. A., Trafton, J. G., & Monk, C. A. (2008). Dealing with interruptions can be complex, but does interruption complexity matter: A mental resources approach to quantifying disruptions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(4), 398-402. https://doi.org/10.1177/154193120805200442

Carlson, R. A., & Cassenti, D. N. (2004). Intentional control of event counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1235-1251. https://doi.org/10.1037/0278-7393.30.6.1235

Cooper, R. P., Byde, C., de Cecilio, R., Fulks, C., & Morais, D. S. (2018). Set-shifting and place-keeping as separable control processes. *Cognitive Psychology*, *105*, 53-80. https://doi.org/10.1016/j.cogpsych.2018.07.001

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *The Quarterly Journal of Experimental Psychology*, *62*(3), 531-549. https://doi.org/10.1080/17470210801995002

Dismukes, R. K., Young, G. E., Sumwalt III, R. L., & Null, C. H. (1998, January). Cockpit interruptions and distractions: Effective management requires a careful balancing act. In

CRM Industry Conference, *10*, 4-9. https://human-factors.arc.nasa.gov/flightcognition/Publications/Distractions.pdf

Dodhia, R. M., & Dismukes, R. K. (2009). Interruptions create prospective memory tasks. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(1), 73-89. https://doi.org/10.1002/acp.1441

Drews, F. A. (2007, October). The frequency and impact of task interruptions in the ICU. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(11), 683-686. https://doi.org/10.1177/154193120705101117

Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied, 9*(3), 147. https://doi.org/10.1037/1076-898X.9.3.147

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, *50*(4), 243-250. https://doi.org/10.1007/BF00309260

Hachmann, W. M., Bogaerts, L., Szmalec, A., Woumans, E., Duyck, W., & Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of Dyslexia*, *64*(2), 121-136. DOI 10.1007/s11881-013-0089-5

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin & Review*, *22*(4), 1104-1110. https://doi.org/10.3758/s13423-014-0764-5

Helton, W. S., & Russell, P. N. (2011). Working memory load and the vigilance decrement. *Experimental Brain Research*, *212*(3), 429-437. https://doi.org/10.1007/s00221-011-2749-1

Helton, W. S., & Russell, P. N. (2015). Rest is best: The role of rest and task interruptions on vigilance. *Cognition*, *134*, 165-173. https://doi.org/10.1016/j.cognition.2014.10.001

Helton, W. S., & Russell, P. N. (2017). Rest is still best: The role of the qualitative and quantitative load of interruptions on vigilance. *Human Factors*, *59*(1), 91-100. https://doi.org/10.1177/0018720816683509

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the Tower of London task: support for a goal-activation approach. *Journal of Experimental Psychology: General*, *135*(1), 103-115. https://doi.org/10.1037/0096-3445.135.1.103

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339-373. https://doi.org/10.1037/a0034221

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *21*(4), 1008-1018. https://doi.org/10.1037/0278-7393.21.4.1008

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352-358. https://doi.org/10.1037/h0043688

Latorella, K. A. (1996). Investigating interruptions: An example from the flightdeck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(4), 249-253. https://doi.org/10.1177/154193129604000423

Latorella, K. A. (1998). Effects of modality on interrupted flight deck performance: Implications for data link. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(1), 87-91. https://doi.org/10.1177/154193129804200120

Lee, B. C., & Duffy, V. G. (2015). The effects of task interruption on human performance: A study of the systematic classification of human behavior and interruption frequency. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *25*(2), 137-152. https://doi.org/10.1002/hfm.20603

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1-44. http://dx.doi.org/10.1037/rev0000253

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2001). Cockpit interruptions and distractions: A line observation study. *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1-6). Ohio State University Press Columbus, OH.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2003). Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations. *Proceedings of the 12th*

International Symposium on Aviation Psychology (pp. 737-742). The Wright State University Dayton, OH.

Lu, S. A., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors*, *55*(4), 697-724. https://doi.org/10.1177/0018720813476298

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(22), 1824-1828. https://doi.org/10.1177/154193120204602210

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied, 14*(4), 299-313. https://doi.org/10.1037/a0014402

Scott-Cawiezell, J., Pepper, G. A., Madsen, R. W., Petroski, G., Vogelsmeier, A., & Zellmer, D. (2007). Nursing home error and level of staff credentials. *Clinical Nursing Research, 16*(1), 72-78. https://doi.org/10.1177/1054773806295241

Trafton, J. G., Altmann, E. M., & Ratwani, R. M. (2011). A memory for goals model of sequence errors. *Cognitive Systems Research, 12*(2), 134-143. https://doi.org/10.1016/j.cogsys.2010.07.010

Ratwani, R. M. (2004). *A spatial memory mechanism for guiding primary task resumption* (Order No. 3313849). [Doctoral dissertation, George Mason University]. ProQuest Dissertations and Theses Global.

Ratwani, R. M., Andrews, A. E., Sousk, J. D., & Trafton, J. G. (2008). The effect of interruption modality on primary task resumption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(4), 393-397. https://doi.org/10.1177/154193120805200441

Radović, T., & Manzey, D. (2019). The impact of a mnemonic acronym on learning and performing a procedural task and its resilience towards interruptions. *Frontiers in Psychology*, *10* (2522), 1-17. https://doi.org/10.3389/fpsyg.2019.02522

Radovic, T. (2021, May 5). Effects of interruption complexity and similarity on procedural task with sequential constraints. https://doi.org/10.17605/OSF.IO/7B3Z4

Westbrook, J. I., Woods, A., Rob, M. I., Dunsmuir, W. T., & Day, R. O. (2010). Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine, 170*(8), 683-690. doi:10.1001/archinternmed.2010.65

Westfall, J. (2015). PANGEA: Power analysis for general ANOVA designs. Unpublished manuscript. http://jakewestfall. org/publications/pangea. pdf.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (eds.), *Variaties of attention* (pp. 63-102). Orlando: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455. https://doi.org/10.1518/001872008X288394

Zijlstra, F. R., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology, 72*(2), 163-185. https://doi.org/10.1348/096317999166581