What does it mean for an interruption to be relevant? An investigation of relevance as a memory effect

Sandy J. J. Gould, Duncan P. Brumby and Anna L. Cox

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013 57: 149
DOI: 10.1177/1541931213571034

What does it mean for an interruption to be relevant? An investigation of relevance as a memory effect

Sandy J. J. Gould, Duncan P. Brumby, Anna L. Cox
UCL Interaction Centre, University College London
s.gould@cs.ucl.ac.uk, brumby@cs.ucl.ac.uk, anna.cox@ucl.ac.uk

Interruptions cause slower, more error prone performance. Research suggests these disruptive effects are mitigated when interruptions are relevant to the task at hand. However, previous work has usually defined relevance as the degree of similarity between the content of interruptions and tasks. Using a lab-based experiment, we investigated the extent to which memory effects should be considered when assessing the relevance of an interruption. Participants performed a routine data-entry task during which they were interrupted. We found that when participants were interrupted between subtasks, reinforcement and interference effects meant that relevance had a significant effect on interruption disruptiveness. However, this effect was not observed when participants were interrupted within subtasks. These results suggest that interruption relevance is contingent on the contents of working memory during an interruption and that interruption management systems could be improved by modelling potential interfering and reinforcing effects of incoming interruptions.

INTRODUCTION

Being interrupted during the execution of a routine task is disruptive and increases the likelihood that errors are made. Mitigating the negative effects of interruptions has been an area of interest within the human factors and human-computer interaction communities. Several approaches have been taken to reducing disruption, from preventing people being interrupted in the first place (Mark, Voida, & Cardello, 2012), to training people to handle interruptions more effectively (Relihan, O'Brien, O'Hara, & Silke, 2010). Of particular interest to us are interruption management systems that schedule (or defer) routine interruptions until periods when they might be less disruptive to performance. For instance, Iqbal & Bailey's (2008) *Oasis* system holds notifications until a user reaches a natural breakpoint in their activity.

In order for interruption management systems to be able to help users manage interruptions effectively, it is important that they are able to detect 'good' moments for the scheduling of interruptions. That is, when are interruptions least disruptive to performance? This question has received significant attention within the interruptions research literature (see Adamczyk & Bailey, 2004; Czerwinski, Cutrell, & Horvitz, 2000; Iqbal & Bailey, 2008). A critical finding that emerges from this work is that the relevance interrupting tasks affects their disruptiveness. For instance, Iqbal & Bailey (2008) have shown that interruptions relating to users' on-going activities are less distracting than interruptions that have nothing to do with the task at hand. Building on this idea, Arroyo & Selker (2011) developed an interruption management system which assessed the similarity of the contents of tasks and interruptions, letting relevant interruptions through to the user while holding back irrelevant ones.

The degree of similarity between the contents of a task and an interruption is a useful measure of the potential disruptiveness of an interruption. However, it does not capture any of the memory effects that could emerge from the combination of a particular task and interruption. For example, if a task requires memorisation of a set of product codes, then an interruption that requires users to think about a related set of codes has the potential to either boost recall of the original codes or to confuse users about which codes they were working with previously. This is potentially problematic for content similarity measures of relevance because it is well known that memory plays a key role in how people recover from interruptions in routine tasks (Altmann & Trafton, 2002).

In this paper we investigate how interruption relevance and timing affect the disruptiveness of an interruption. We report the results of a lab study built around a routine data-entry task. During task execution participants were interrupted by relevant or irrelevant interruptions that appeared between or within subtasks. In keeping with previous results, we expect that relevant interruptions will be less disruptive than irrelevant interruptions. By using interruptions that focused on participants' knowledge of subtasks and their progress through them, we expect the effect of relevance to be modulated by the timing of interruptions.

Related work

It is well known that interruptions reduce performance and take time to recover from. When interrupted, people must suspend activity on the task that they are working on in order to attend to an interrupting task. After dealing with an interruption it is necessary to recall progress made before the onset of the interruption. The more disruptive the interruption is, the more difficult it is to quickly and accurately recall where to resume. Therefore, the time it takes to resume after an interruption, the *resumption lag* (Altmann & Trafton, 2002), is a good measure of interruption disruptiveness (but see Brumby et al., 2013 for an alternative account). A number of factors that influence the disruptiveness of interruptions have been identified, but two factors have attracted particular attention: relevance and timing.

Intuitively, one might expect that interruptions that are relevant to the primary task will be less disruptive than irrelevant ones because they might contain useful information and maintain context. Support for this intuition has been found in number of studies (e.g., Czerwinski et al., 2000; Iqbal & Bailey, 2008). Assessing the relevance of an interruption has the advantage of allowing for the blocking of irrelevant interruptions and the prioritisation of relevant interruptions. This makes it possible for one of the beneficial aspects of interruptions, timely delivery of information, to be exploited to maximum effect.

Arroyo and Selker's (2011) interruption management system uses relevance to determine whether incoming notifications should be released to the user. The system computes relevance by comparing the content of the current activity and an incoming notification. If there is a good correspondence, the interruption is deemed to be relevant. However, such a system cannot distinguish between an interruption that is relevant to the task at hand, and one that has similar content, but is nonetheless irrelevant.

Indeed, previous investigations into interruption relevance have also treated content similarity and relevance as equivalent. For instance, Czerwinski et al. (2000) defined relevant interruptions as those that gave participants the solution to a question they were trying to answer in the primary task. Irrelevant interruptions gave participants some fact about the environment in which they were working. Iqbal & Bailey (2008) make a similar distinction: relevant interruptions were related to the current task, irrelevant interruptions were not. Although these studies have shown that relevant interruptions are less disruptive than irrelevant interruptions, defining relevance on the basis of content similarity ignores the memory processes that play a significant role in resuming after interruption.

Altmann & Trafton's (2002) Memory for Goals (MfG) theory is an activation-based model of memory that describes how goals representing primary task progress are managed before, during and after interruptions. Their theory suggests that the more that a goal is rehearsed during an interruption, the more likely it will be that recall of a goal on resumption will be quick and accurate. Empirical results corroborate this theoretical assumption (see Monk, Trafton, & Boehm-Davis, 2008; Trafton, Altmann, Brock, & Mintz, 2003). As goals must be maintained in memory during interruptions in order for successful resumption, such goals are open to a variety of memory effects (see Ratwani & Trafton, 2008).

Following the assumptions of the MfG theory, we predict that goals representing task progress should be susceptible to both interference and reinforcement. This means that using content similarity as a proxy for relevance has the potential to cause unanticipated increases to the disruptiveness of interruption. For example, an incoming interruption could have similar content to that of the current activity; an interruption management system might see this as relevant, and prioritise it. However, the interruption might relate to a step in the task that has yet to be started or has been completed already. This could cause proactive interferences with memory of task progress, increasing interruption disruptiveness.

To illustrate this point consider a user working on a particular step of a routine data-entry task. When they are interrupted, it is important that they remember where they were in the primary task. In one scenario, the incoming interruption calls for direct rehearsal of progress by asking participants explicitly about their progress through the primary task. In another scenario, participants are asked random questions about the task environment. While from a content similarity perspective both interruptions are relevant to the primary task, they are likely to have different effects on memory: the former forces rehearsal of where the user needs to resume, whereas the latter might encourage proactive interference by forcing participants to divert their attention to other parts of the task which may have already been completed. Indeed, an interruption with no content similarity might prove to be less disruptive by virtue of it merely preventing rehearsal task progress, rather than causing proactive interference.

Of course, the capacity of an interruption to elicit reinforcement or interference effects depends on the representation of progress that is maintained for the duration of the interruption. These representations change moment-to-moment during task during execution, suggesting that relevance should also be seen as dynamic property, the magnitude of which depends on the timing of an interruption.

So to understand relevance, it is also necessary to understand the effects of interruption timing on disruptiveness. Previous work has shown that interruptions arriving at subtask boundaries are less disruptive than those arriving in the middle of completing a subtask (Cutrell, Czerwinski, & Horvitz, 2000; Monk, Boehm-Davis, Mason, & Trafton, 2004). Iqbal & Bailey (2005) suggest that this is because subtask boundaries represent moments of low workload; it's easier to be interrupted at the end of a sentence than in the middle of writing one. These variations in workload are reflected in the decisions people make about whether or not to deal with a discretionary interruption: that is, given the choice, people wait until subtask boundaries before interrupting themselves (Bogunovich & Salvucci, 2011). This body of work on interruption timing has been translated into an interruption management system, *Oasis* (Iqbal & Bailey, 2010), which holds incoming notifications in a buffer until it detects that a user has reached a subtask boundary.

Using task boundaries as a proxy for workload is likely to work well as long as moments of low workload align with task boundaries. However, it is unclear whether subtask boundaries are still a useful heuristic for interruption management systems if workload does not align with subtask boundaries. For example, in some data-entry tasks, planning where information needs to be entered might more taxing than act of entering the values itself. If reduced disruptiveness at subtask boundaries were simply a corollary of workload, one would expect that tasks with consistently low workload would confer no advantage on between-subtask interruptions.

In this paper, we investigate whether the relevance and timing of an interrupting task affects resumption performance. We report a study in which participants were interrupted partway through a routine data-entry task. To manipulate relevance, we used interruptions that either asked questions about participants' memory of their progress in the primary task (i.e., relevant to memory of task progress) or asked questions about some feature of the primary task (i.e., similar content to the primary task, but potentially giving proactive interference to memory of task progress). We expect that though both interruption types have similar content, relevant interruptions will increase rehearsal of progress, resulting in less disruption compared to irrelevant interruptions.

A secondary aim of this study is to understand whether subtask boundaries remain the least disruptive moment for an interruption when workload is consistently low throughout a task. To do this we manipulate interruption timing, comparing the effects of interruptions occurring between subtasks with those occurring within subtasks. We expect that our low-workload task will mean that this manipulation has no effect.

METHOD

Participants

Twenty-four participants (14 male) with a mean age of 23 years (SD=5 years) took part in the study. Participants were drawn from a university participant pool and were paid £7 (~$11) for approximately one hour of their time.

Design

The experiment used a within-subjects design with two independent variables; interruption relevance, which had two levels, relevant and irrelevant; and interruption timing, which also had two levels, within-subtask and between-subtask. The primary measure was resumption lag, which is the time between a participant returning to the primary task after interruption and selecting an interface widget.

Materials

Primary task. The primary task was a routine sequential data-entry task based on the Doughnut Machine, a task designed for the investigation of human error and interruption (Li, Cox, Blandford, Cairns, & Abeles, 2006). The task comprised five subtasks completed in a strict order: *Type*, *Shape*, *Colour*, *Packaging* and *Label*. Each subtask was made up of five elements. At the start of a trial, participants were given a list of three orders. Each order was made up of a quantity (e.g. 15, 30, 45) and four attributes (e.g. Tablet, Diamond, Red, Foil) that corresponded with the five subtasks (see Figure 1). Participants had to copy values from the order list into the appropriate parts of the subtasks. Once participants had completed a subtask, the information they had entered was cleared. This was to remove place-keeping cues from the task.

Interrupting task. The interrupting task comprised two short tasks: audit tasks and transcription tasks. Whenever a participant was interrupted, they would complete two audit tasks and two transcription tasks in the order *audit*, *transcription*, *audit*, *transcription*.

Audit tasks questioned participants about some aspect of the primary task. All widgets were removed from the task interface, leaving only the subtask frames visible. In the centre of the screen, the order list was replaced with a question. Participants answered the questions by selecting the subtask that they thought was correct.

The audit tasks were the instrument for manipulating interruption relevance. Therefore, the questions participants had to answer depended on whether the interruption was relevant or irrelevant. In irrelevant audits, participants were asked a question about some aspect of the primary task (e.g. 'Which is the Label subtask?'). The irrelevant interruptions asked questions about random subtasks and required no knowledge of current progress. Conversely, for relevant interruptions, participants were asked about their progress through the trial (e.g. 'What subtask did you just complete?'). To correctly answer the questions, participants required some knowledge about their progress through the trial. Participants received no feedback on their performance on audit tasks. This was to ensure that participants were given no information during the interruption that could have aided place-keeping.

The other kind of task used in the interruptions was a simple transcription task. This acted as a filler task. Participants had to copy a set of four-digit numbers into adjacent text-entry fields. All numbers had to be accurately entered before participants could continue.

Once participants had worked through the two pairs of audit and transcription tasks, they were returned to the primary task. Participants were instructed to continue what they were doing before they were interrupted. This was made difficult by the absence of place-keeping cues in the task. Participants had to remember what they were doing when resuming regardless of whether they were interrupted between or within subtasks.

Interruption timing had two levels, *within* and *between*. Within-subtask interruptions occurred while participants were in the middle of entering values into a subtask. Participants had to enter values into three elements for each subtask. Within-subtask interruptions occurred when participants had entered some but not all of the values into a subtask. This meant that participants were guaranteed to be in the middle of working on the subtask when they were interrupted.

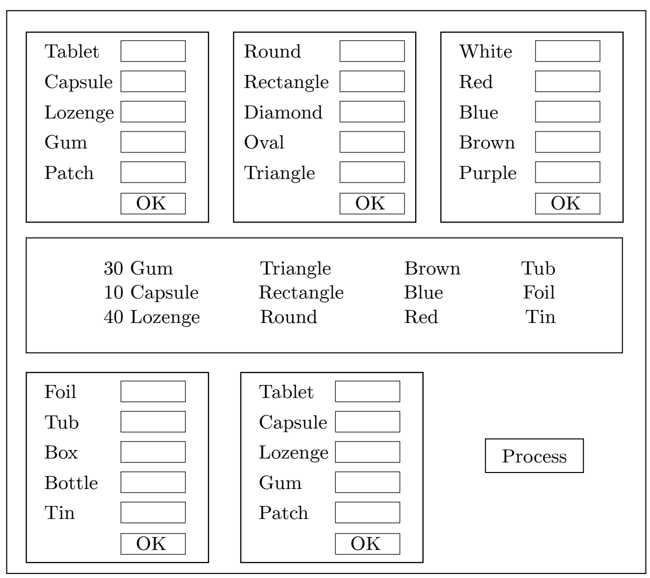

Figure 1: An illustration of the primary task, comprising five subtasks and order list. Tasks are completed left to right and top to bottom.

Between-subtask interruptions occurred after participants completed a subtask, but before they started the next. The interruptions appeared the moment that participants clicked the 'OK' button of a subtask, signifying they had completed work on that subtask.

Procedure

After reading an introduction and giving informed consent, participants watched a training video that provided a walkthrough of the task and an explanation of its features. Participants then completed four training trials. The first training trial was uninterrupted and was followed by a relevant, between-subtask interruption. The final pair of training trials both had within-subtask interruptions: one relevant and one irrelevant. Once familiar with the task, participants completed 12 experimental trials. Conditions were randomly allocated in three blocks of four (within-subtask, relevant; within-subtask, irrelevant; between-subtask, relevant; between-subtask, irrelevant). Participants had the option of a two-minute rest after completing six trials. At the end of the experiment participants completed a six-item questionnaire and gave verbal responses to four free-response questions. They were then debriefed and paid.

RESULTS

Errors and outliers. Each participant was interrupted 24 times during the course of the experimental trials, for a total of 576 interruptions across all participants and conditions. For the purposes of calculating resumption lag, were particularly interested in accurate resumptions, so incorrect resumptions were removed from the set of resumption lag data.

The first kind of inaccurate resumptions to be removed from the dataset were subtask-level sequence errors. These occurred when participants attempted to work on the wrong subtask. For example, if a participant was supposed to work on the 'Label' subtask, but instead worked on 'Type' subtask, this would be counted as a between-subtask level sequence error. They could occur after either between- or within-subtask interruptions. Eighty-six inaccurate resumptions of this kind were removed from the dataset. Of these 86 erroneous resumptions, 63 occurred when participants were resuming after between-subtask interruptions, meaning a further 23 occurred after within-subtask interruptions.

The second kind of inaccurate resumptions to be removed from the dataset were within-subtask level resumptions. These occurred when participants, on resuming after a within-subtask interruption, selected a subtask element that was different to the one they were about to work on at the point of interruption. There were 57 inaccurate resumptions of this kind, and these were removed from the dataset.

After removing the erroneous resumptions from the dataset, mean resumption lags were calculated for each participant and outlying resumptions were removed if a particular resumption was ±1.96 standard deviations (i.e. 95%) from the participant's mean. Very fast resumptions suggest guessing, and very slow ones suggest that the participant was doing something other than simply resuming the task. Under this criterion, a further 23 resumptions were removed from the dataset, leaving a total of 410 resumptions for analysis.

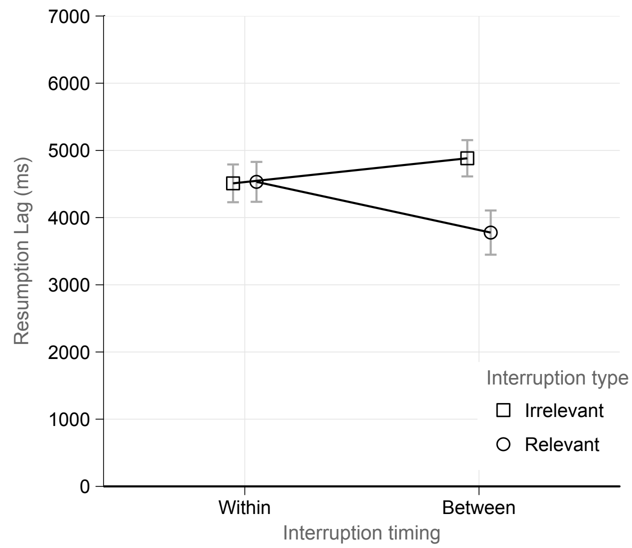

Figure 2: Plot showing time to resume after interruption for each of the four conditions.

Post-interruption resumption. Resumption lag, the time between an interruption finishing and a participant restarting work on the primary task, was the primary measure of performance. A factorial repeated measures ANOVA was used to determine whether the manipulations of interruption relevance and timing had any effect on resumption lag.

Figure 2 shows the mean resumption lag for each condition. There was a significant main effect of interruption type, such that participants had shorter resumption lags when recovering from relevant interruptions than irrelevant ones, $F(1,23)=12.94$, $p<.01$, $\eta_p^2=.36$. As can be seen in the figure, the effect of interruption relevance was clearly moderated by the timing of the interruption; that is, there was a significant interaction between interruption relevance and timing, $F(1,23) = 5.04$, $p<.05$, $\eta_p^2=.18$. A follow-up simple effects test showed that participants had faster resumption lags after relevant interruptions, but only when they arrived in between subtasks, $F(1,23) = 294$, $p<.001$. Overall, there was no significant main effect of interruption timing, $p=.49$.

DISCUSSION

The primary aim of this experiment was to determine the extent to which relevance can be thought of as the combined effects of interference and reinforcement in memory. The interaction effect we observed suggests that the relevance of a particular interrupting task varied significantly over the course of task execution. When participants were interrupted within a subtask, relevance had no effect on disruptiveness. When participants were interrupted between subtasks, relevant interruptions reduced disruptiveness while irrelevant interruptions increased it. As all interruptions had content related to the primary task, this supports our main hypothesis that content similarity is not a sufficient criterion for judging the potential disruptiveness of interruptions.

The secondary aim of this study was to investigate whether subtask boundaries remained the least disruptive moment for interruption even when primary and interruption task workloads are low. Our results suggest timing only affected disruptiveness by mediating the effect of relevance. In this study, interruption timing acted as a mechanism for manipulating relevance. As a result, our manipulation of relevance had no effect on the disruptiveness of interruptions when participants were interrupted within a subtask.

In the context of previous studies on the role of memory in interrupted work, it seems likely that the interaction effect we observed arose from differences in the encoding of place-keeping information during task execution. For relevant, between-subtask interruptions, asking questions about progress through the task required participants to think about what they were doing when they were interrupted, thus reinforcing this place-keeping information. Conversely, irrelevant interruptions had the opposite effect. Asking participants to recall information that was irrelevant but superficially related to the task at hand interfered with the correct memory of progress, increasing disruptiveness.

Although the results suggest that participants were encoding within-subtask place-keeping goals in a way that meant there were no marginal reinforcement or interference effects during interruptions, it is not clear how participants were encoding within-subtask place-keeping goals. Feedback from post-experiment questioning suggests that participants used a variety of phonological encoding strategies for remembering where they were in a subtask. For example, they would rehearse a particular number or word associated with their progress. Given that between-subtask interruptions in tasks like the one used in this study have been shown to encourage spatial encoding of progress information (Ratwani & Trafton, 2008), we hypothesise that encoding differences were responsible for the divergence in disruptiveness observed between the two interruption timing conditions: when participants were interrupted within a subtask they used a phonological encoding strategy, when they were interrupted between subtasks they used a spatial encoding strategy.

These results suggest that interruption management systems (e.g., Arroyo & Selker, 2011; Iqbal & Bailey, 2010) cannot assume an interruption is relevant simply because its content is similar to that of the current activity. Indeed, our results suggest that this approach could result in *worse* performance by prioritising potentially disruptive interruptions over more benign ones. This poses a problem: how can we model relevance if we need to know about the mental states of users at the point of interruption? One way to get around the problem of predicting memory states is to focus on understanding how people adapt their representations of progress during task execution. These representations are vital to resuming after interruptions, so understanding such adaptations will improve the prediction of interference and reinforcement effects. In turn, this could improve identification of the potential disruptiveness of an interruption.

CONCLUSION

Through an experimental investigation of interruption timing and relevance, we have demonstrated that content similarity is not always the same as relevance and that using the concepts interchangeably in interruption management systems has the potential to *increase* the disruptiveness of interruptions. This suggests that researchers should be mindful of theories of memory when thinking about interruption relevance in the context of management systems. Keeping track of progress requires working memory, even when supported by environmental artefacts. Memory is often prone to interference and reinforcement effects, and this is no less the case when dealing with interruptions. Future work should examine whether memory effects still play a significant role in relevance in more complex tasks where people might rely more heavily on artefacts to guide their resumption. This is a necessary step before the findings of this study can be utilised in interruption management systems.

ACKNOWLEDGEMENTS

The authors are supported by the CHI+MED project, which is funded by UK EPSRC grant EP/G059063/1.

REFERENCES

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: the effects of interruption at different moments within task execution. *CHI '04* (pp. 271–278).

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cognitive Science*, *26*(1), 39–83.

Arroyo, E., & Selker, T. (2011). Attention and Intention Goals Can Mediate Disruption in Human-Computer Interaction. *Interact '11* (Vol. 6947, pp. 454–470).

Bogunovich, P., & Salvucci, D. D. (2011). The effects of time constraints on user behavior for deferrable interruptions. *CHI '11* (pp. 3123–3126).

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. J. (2013). Recovering from an interruption: Investigating speed-accuracy tradeoffs in task resumption strategy. *Journal of Experimental Psychology: Applied, 19*(2), 95-107.

Cutrell, E. B., Czerwinski, M., & Horvitz, E. (2000). Effects of instant messaging interruptions on computing tasks. In *CHI '00* (p. 99), ACM.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant Messaging: Effects of Relevance and Timing. *BCS HCI '00* (Vol. 2, pp. 71–76).

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *CHI '05* (pp. 1489–1492).

Iqbal, S. T., & Bailey, B. P. (2008). Effects of intelligent notification management on users and their tasks. *CHI '08* (pp. 93–102).

Iqbal, S. T., & Bailey, B. P. (2010). Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM ToCHI*, *17*(4), 15:1–15:28.

Li, S. Y. W., Cox, A. L., Blandford, A., Cairns, P., & Abeles, A. (2006). Further investigations into post-completion error: the effects of interruption position and duration. *CogSci '06* (pp. 471–476).

Mark, G., Voida, S., & Cardello, A. (2012). 'A pace not dictated by electrons': an empirical study of work without email. *CHI '12* (pp. 555–564).

Monk, C. A., Boehm-Davis, D. A., Mason, G., & Trafton, J. G. (2004). Recovering From Interruptions: Implications for Driver Distraction Research. *Hum. Factors*, *46*(4), 650 –663.

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.

Ratwani, R. M., & Trafton, J. G. (2008). Spatial memory guides task resumption. *Visual Cognition*, *16*(8), 1001.

Relihan, E., O'Brien, V., O'Hara, S., & Silke, B. (2010). The impact of a set of interventions to reduce interruptions and distractions to nurses during medication administration. *Qual Saf Health Care*, *19*(5), 1 –6.

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