The Effect of Repeated Cue Exposure on Post-Completion Errors

Maartje G. A. Ament (m.ament@ucl.ac.uk)

Anita Y. T. Lai (anita.y.t.lai@gmail.com)

Anna L. Cox (anna.cox@ucl.ac.uk)
UCL Interaction Centre
University College London
Gower Street
London, WC1E 6BT
United Kingdom

Abstract

Slip errors in routine procedures are notoriously difficult to eliminate. They are not the result of a lack of knowledge, so retraining or increasing motivation is not effective. Recently, a number of studies have used visual cues in an attempt to reduce post-completion errors. These slip errors occur when the last step in a task is forgotten after the main goal has been completed. Visual cues have been shown to be effective in reducing post-completion errors if sufficiently aggressive. However, concerns have been raised over whether repeated exposure to visual cues may lead people to rely on them, and therefore become more prone to errors should the cue inadvertently be unavailable. Alternatively, people may anticipate the appearance of the cue, which may have a positive effect on error rates. We report on an experiment that tests the effect of repeated exposure to a strong visual cue on the post-completion error rate. Participants were repeatedly exposed to an aggressive visual cue just before the post-completion step, and this cue was withdrawn on a number of target trials. Our results show that error rates on these non-cued target trials are similar to those of non-cued control participants. Moreover, eye-movement patterns are similar for target and non-target trials. These results suggest that participants do not become over-reliant on the cue, while they partially support the hypothesis that participants anticipate the appearance of the cue. As such, visual cues are a safe and effective way to reduce slip errors.

Keywords: Human error; routine procedural action; visual cues; post-completion error; eye-tracking; slip errors.

Introduction

Despite having the correct knowledge, people still make errors while executing routine procedures. Often, these slip errors are inconsequential, but in safety-critical domains such as aviation, medicine and nuclear power plant operations they can have severe consequences. While slip errors occur only occasionally, they are nevertheless persistent. They do not occur as a result of a lack of knowledge, and this makes them difficult to eliminate, or even reduce (e.g. Byrne & Davis, 2006).

A growing body of empirical work has studied these cognitive slips in the laboratory. For instance, Byrne and Bovair (1997) argued that the post-completion error (PCE), an error that occurs when the last step in a procedure is forgotten after the main goal has been completed, is a result of a failure of working memory. They found that a higher working memory load could increase the occurrence of the PCE. More recent research has focussed on how to reduce these errors. Back, Cheng, Dann, Curzon, and Blandford (2006) investigated the role of motivation in the occurrence of post-completion errors. They found that motivation to avoid them does not lead to lower error rates on the post-completion (PC) step. In a similar study, Byrne and Davis (2006) investigated the effect of a number of interventions on the PCE, including motivation, retraining and redesign. They found that only redesigning the interface resulted in the complete elimination of PCEs, while reprimanding, praising or retraining participants was not more effective than the control condition with no interventions at all.

Another mitigation strategy that has seen recent investigation is the use of visual cues. Chung and Byrne (2008) studied the effectiveness of a number of visual cues in reducing post-completion errors. They found that a red singleton onset appearing just before the PC step was not effective in reducing the number of PCEs. However, a more aggressive cue with blinking arrows pointing at the PC button achieved the complete elimination of PCEs. Byrne (2008) further investigated the properties a cue must possess in order to be effective. He found that the most important attributes of a successful visual cue are, in order of importance: appearing just-in-time, being specific, and being salient.

However, while visual cues can be effective in reducing slips, they also have their drawbacks. Ratwani, McCurry, and Trafton (2008), for instance, argued that cues can become less effective over time, and repeated exposure to them can be annoying. To address these issues, they developed a real-time system that uses eye movements to predict when an error is likely to occur. This allowed them to provide a cue only when needed, reducing some of the concerns associated with visual cues. Ratwani and Trafton (2009) reported that this system is very effective in reducing error rates.

Byrne (2008) highlighted another potential drawback of the use of visual cues. He argued that providing repeated cues may lead people to become dependent on the cue to remember a particular step. This could be problematic in situations in which the cue fails, such as when loud traffic noise masks the ATM beeping sounds upon returning the card, or when a user uses a different device to perform a task which does not provide cues for the same steps. It is possible that, if people learn to rely on repeated cues, they become particularly prone to errors on the cued step, beyond what is expected of a user who has never received a cue. Conversely, another possibility is that people come to anticipate the strong visual cue. In the absence of the cue, the expectation of seeing it may in itself act as a reminder to reduce error rates. However, to date no studies have addressed these effects. Therefore, the current paper investigates the impact on performance of these failed or absent cues. It considers both the hypothesis that people may become over-reliant on the cue, as well as the hypothesis that anticipation for the cue may act as a reminder.

In order to understand how cues work and why people may learn to rely on them, we turn to a theoretical account of goal-directed action. The Memory-for-Goals (M-f-G) model (Altmann & Trafton, 2002, 2007) has previously been used to explain the effectiveness of cues in preventing post-completion errors (e.g. Byrne, 2008; Chung & Byrne, 2008). The M-f-G theory argues that goals in memory have an associated activation level, and the goal with the highest activation level is the one that directs behaviour. This means that a goal's activation level must be higher than the interference level, which is equal to the activation level of the next most active goal in memory. The activation of a goal decays over time, and therefore a mechanism is needed to regulate goals' activation levels. As such, goals must be strengthened (through frequent and recent use) and primed (from mental and contextual cues) to gain enough activation. This priming constraint makes a number of predictions about visual cues:

- Cues should be just-in-time, to ensure the cued goal will not be masked by other goals and lose its activation before being executed.
- Cues should be specific, so that the cue and its target goal are associatively linked.
- Cues should be salient, to ensure that the cue will be processed by the cognitive system.

These predictions are in line with Byrne's (2008) findings on the essential properties of visual cues. Chung and Byrne (2008) further argue that cues are effective in reducing PCEs because they provide a low-effort strategy for remembering the post-completion step. In line with the law of least mental effort (e.g. Kool, McGuire, Rosen, & Botvinick, 2010), people will exploit available cues in a task interface to reduce cognitive effort.

It is not immediately clear what the M-f-G model predicts about the effect of repeated cues on PCEs. In the absence of cues, the post-completion error is normally avoided because the previous step in the procedure provides associative activation for the PC subgoal. These associative links form between steps in a routine procedure, with each step providing an internal cue for the next one. One way in which repeated cueing may lead to higher error rates if the cue fails may be through the loss of these associative links to the PC step. If people receive a cue for the PC step each time they execute it, it is likely that they will develop an associative link between the cue and the PC step. However, since it is this cue that provides the necessary boost in activation for the PC step to be executed, such a link between the previous step and the PC step may be substantially weaker. If the visual cue is then unavailable, the remaining source of activation for the PC step may not be sufficient for the step to reach the required activation level. An error is thus more likely to occur.

Conversely, because the completion of the pre-PC step and the appearance of the visual cue happen at the same time, an associative link may form between them. While the appearance of the cue is not a goal per se, it may become associated with the response of focussing attention on the visual cue and therefore the PC step button. If the goal of moving attention to the visual cue is sufficiently activated through this, it may come to be associatively primed by the pre-PC step, even in the absence of the visual cue. This would make it more likely that the PC step is remembered. Moreover, the repeated reminder for the PC step may reinforce this step in the task representation. This would also make it more likely that the step is remembered, even when the cue is absent.

The current paper presents an experiment that addresses the effect of consistent exposure to a strong visual cue in a well-known post-completion task. Participants are habituated to the cue over a period of time, after which it is withdrawn on three target trials. Performance on these target trials is compared to baseline performance of a control group, who have received no cues at all. Two hypotheses are tested: if participants become over-reliant on the cue, it is expected that error rates would increase beyond the baseline error rate on cue-absent trials after repeated exposed to the cue. Alternatively, if participants anticipate the cue, fewer errors would be made on these cue-absent trials. Eye movements are used to further investigate cue anticipation, which can be revealed by prompt fixation to the cue location even when the cue is not present.

Method

Participants

Forty-five students (25 female) from University College London took part in the experiment. They were aged between 18 and 36 years, and their mean age was 23.7 years. Each received £6 for their time and effort.

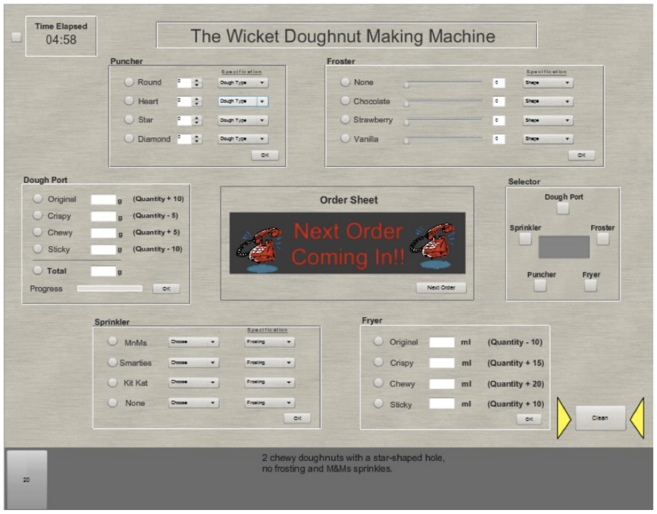
Stimuli

Ament, Cox, Blandford, and Brumby's (2010) variant of the Doughnut task (originally developed by Li, Cox, Blandford, Cairns, and Abeles (2006)) was used. It consists of a routine procedural task that requires participants to make virtual doughnuts, and a monitoring task, as shown in Figure 1.

A trial starts with receiving and displaying a new doughnut order, which has to be entered into the machine. The main doughnut task consists of five compartments, or widgets, which need to be operated in the order:

Dough Port → Puncher → Froster → Sprinkler → Fryer.

Before data can be entered, a widget needs to be activated by clicking the appropriate selector button on the selector panel



2 chewy doughnuts with a star-shaped hole,
no frosting and M&Ms sprinkles.

Figure 1: The Doughnut Task interface. The banner at the bottom is the Doughnut Live Feed. The visual cue is shown in the lower right-hand corner, on both sides of the PC step button.

on the right-hand side. Clicking the ‘OK’ button then confirms the entry for that widget. Once all widgets have been completed, the order needs to be processed by clicking the ‘Process/Clean’ button. A pop-up screen then indicates the completion of the trial, and displays the number of doughnuts made. This serves as the false completion signal, indicating that the main goal of the task has been completed. After dismissing the pop-up screen, the machine must be cleaned by clicking the ‘Process/Clean’ button again. This serves as the post-completion step. A competing signal is also present: a flashing telephone indicates that the next order is waiting.

Like Byrne and Bovair (1997), the current experiment used a secondary monitoring task to increase working memory load. This task, called the Doughnut Live Feed, ensured that the error rates were at a measurable level. It requires participants to monitor the number of doughnuts sold in the shops. The Live feed was shown at the bottom of the screen, and displayed a description of a doughnut at irregular intervals. Participants had to monitor and keep track of all doughnuts with ‘chewy’ dough. Once they had counted twenty chewy doughnuts, they had to click the ‘20 sold!’ button on the left of the Live Feed and start counting from zero again. This allowed the experimenter to assess whether a participant was successfully monitoring the live feed.

To ensure effective monitoring, new items on the live feed did not capture visual attention. This was achieved by using a background that changed from grey to white and back in continuous cycles. Each doughnut description faded in on top of that from white to black, and faded out again after a random number of cycles. Figure 2 shows the progression through one cycle. Each cycle took three seconds, and items remained visible for between 2 and 4 cycles. This randomness made it impossible for participants to predict when a new doughnut description would be shown. The monitoring task and primary tasks were carried out simultaneously.

Visual cues consisted of two yellow-and-red flashing arrows, pointed at the post-completion button (see figure 1, bottom right). On cued trials, the cue appeared upon the completion of the previous step (dismissing the pop-up button), and disappeared again when the post-completion button was successfully clicked.

Design

A between-subjects design was used, with the level of visual cues as the independent variable. This variable had two conditions. In the cued condition, participants were habituated to the visual cue during training, and subsequently received the cue just before the post-completion step on all trials, with the exception of trials 4, 8 and 11. On these three target trials, the cue was withdrawn. In the control condition, on the other hand, participants received no visual cues at all.

The primary dependent measure was the number of post-completion errors made on the target trials. An error was defined as any action that deviates from the required action. Only one error could be counted on the PC step, regardless of the number of incorrect mouse-clicks made between the pre-PC and PC steps. Another dependent measure of interest was the eye-movements between the pre-PC and the PC step. Fixations on the PC step button and the cue could reveal if participants anticipated the appearance of the cue even when it was withdrawn. Eye movements were recorded using a Tobii 1750 eye tracker.

Procedure

Participants carried out the task individually. They first read the paper-based instructions, after which they observed the researcher executing one trial of the task, and were reminded of the PC step and the importance of the monitoring task. Participants in the experimental group were briefed about the cue and its significance.

1 crispy doughnut with a star-shaped hole,
no frosting and KitKat sprinkles.

1 crispy doughnut with a star-shaped hole,
no frosting and KitKat sprinkles.

1 crispy doughnut with a star-shaped hole,
no frosting and KitKat sprinkles.

1 crispy doughnut with a star-shaped hole,
no frosting and KitKat sprinkles.

Figure 2: The Doughnut Live Feed. (a) A cycle starts out completely white. (b) The background then quickly fades to grey, while the item fades from white to black. (c) Halfway through the cycle, the background and the item are at its darkest, and the item is clearly visible. (d) At the end of the cycle, the background fades to white again while the item may either stay visible or fade as well.

Participants then completed two practice trials in the presence of the researcher. The experimental group received visual cues before each PC step; the control group did not. Any errors were pointed out using the default Windows XP error sound, and had to be corrected before the participant was allowed to carry on. At the end of each trial, participants were asked how many chewy doughnuts they had counted, to encourage effective monitoring of the Live Feed. Once the two initial practice trials were completed, participants then completed a further 4 practice trials to habituate them to the cue. Participants in both conditions did this, but as before, only the experimental participants received cues.

To prevent self-monitoring, participants were not told that errors were being studied. Moreover, experimental participants were not told that the cue would occasionally be absent. Participants were instructed to work as quickly and as accurately as possible. All participants completed 11 experimental trials. The total duration of the experiment was approximately one hour.

Results

Data from 45 participants was recorded. One had to be excluded due to a lost data file, leaving 44 participants for analysis, 22 in each condition.

Error Rates

We were primarily interested in the post-completion error rates on target and non-target trials. Error rates were calculated for each participant. Only one PCE was possible on each trial, giving a total of 44 x 11 = 484 PCE opportunities. Forty-one post-completion errors were made across all participants, giving an overall PCE rate of 8.47%. Table 1 shows the numbers of errors observed on the different trials and conditions.

	Trial	
Condition	**Target**	**Non-Target**
Experimental	10 (66)	4 (176)
Control	9 (66)	18 (176)

Table 1: Number of errors observed on target and non-target trials, for both experimental and control conditions. The number of opportunities for error is given in brackets.

Non-Target Trials To test whether the cue worked effectively, error rates on non-target trials were investigated. Figure 3 shows the average error rates on non-target trials for experimental and control groups. Since the data was not normally distributed, a Mann-Whitney U test was used to compare the groups. This showed a significant difference between experimental (cued) and control (non-cued) groups, $Z = 3.18, p = 0.001$. This indicates that the cue was effective in reducing error rates.

Target Trials Figure 4 shows the average error rates on target trials for experimental and control groups. A Mann- Whitney U test showed no significant difference between experimental and control groups, $Z = 0.29, p = 0.77$. This indicates that repeated cues did not lead to higher or lower error rates on cue-absent trials.

Eye Movements

Eye movement data was only collected for a subset of the participants due to issues with the equipment. This resulted in 15 usable data files. Eye movement data was isolated for the time period from the pre-PC step (dismissing the pop-up screen) until the PC-button was clicked. Only correct cases were included, for two reasons. First, eye movements on erroneous trials tend to be more erratic and may introduce outliers. Sec-

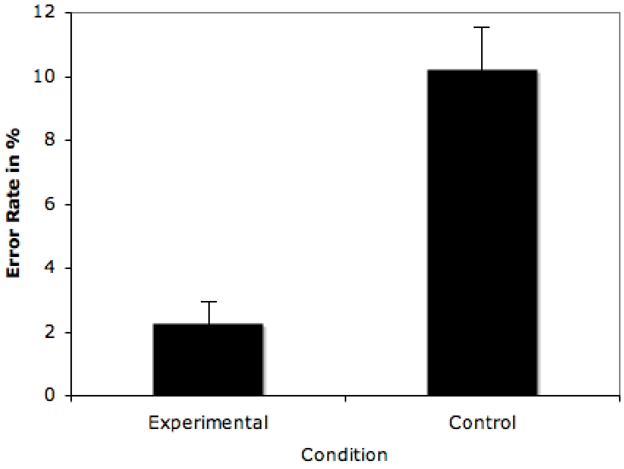


Figure 3: Error rates on non-target trials for experimental and control groups.

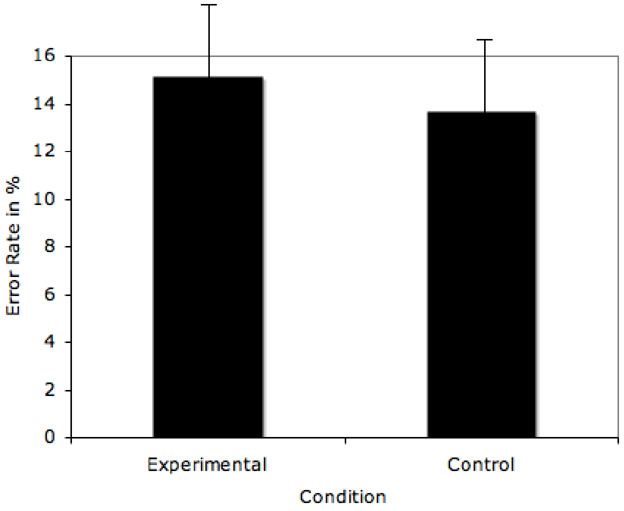


Figure 4: Error rates on target trials for experimental and control groups.

Condition	**Trial**	**Number of Fixations Prior to Fixating on PC Step**
Experimental	Target	2.5 (1.4)
	Non-Target	2.9 (0.9)
Control	Target	3.8 (1.2)
	Non-Target	3.9 (0.9)

Table 2: Summary of eye movement data on target and non-target trials, for both experimental and control conditions.

ond, it can be argued that eye movements on erroneous trials are not an accurate reflection of successful routine procedural action.

Table 2 shows an overview of the eye movement measures. The main measure used was the number of fixations prior to fixating on the PC step. It was argued that the visually salient cue would capture participants' attention and therefore reduce the number of fixations prior to looking at the PC button. If participants anticipated the appearance of the cue even when it was not present, it was expected that a similar number of fixations would be made before fixating on the PC step button on target and non-target trials.

An independent samples t-test showed that the experimental participants made significantly fewer fixations (M = 2.86, SD = 0.88) than control participants (M = 3.98, SD = 0.71) before fixating on the PC step, $t(14) = 2.72, p < 0.05$. Moreover, there was no significant difference in the number of fixations before fixating on the PC step button between target (M = 2.95, SD = 0.91) and non-target trials (M = 2.52, SD = 1.43) in the experimental condition, $t(7) = 1.13, p = 0.30$.

Discussion

The current experiment investigated the effect of repeated exposure to a strong visual cue. Error rates were compared for target (non-cued) trials when participants had previously been cued (experimental) or received no cues at all (control). Eye movements were also recorded to further investigate any possible effects of anticipation.

The results of this study support previous findings that cues are an effective way to reduce post-completion errors. On cued trials, participants made significantly fewer errors than on non-cued trials, although a complete elimination, as found by Chung and Byrne (2008), was not achieved. Nevertheless, this finding was important for the validity of the current experiment, as an ineffective cue would have undermined the ability of the experiment to detect cue reliance.

The results show that cue-trained experimental participants performed similarly to non-cued controls on trials in which no cue was present. This has several implications. First, it indicates that a cue is needed on every trial in order to be effective. This is in line with previous findings by Chung and Byrne (2008) and Byrne (2008), who showed that cues must be just-in-time to be effective.

Second, this finding indicates that cue-trained participants did not become dependent on the cue to remember the post-completion step. When the cue was absent, they were able to remember the PC step as often as the participants who never received a cue. If they had become overly reliant on the cue to remember the step, it would have been expected that participants would not have been able to remember it on their own, without a cue.

These results are in line with the M-f-G model (Altmann & Trafton, 2002, 2007). This model accounts for the correct execution of PC steps using associative links: the pre-PC step provides associative activation for the PC step. Priming cues can also come from the external environment: sufficiently aggressive visual cues are very effective in priming the PC step (Chung & Byrne, 2008). The current results suggest that both internal primes through associative links and external primes through visual cues can independently provide activation for the PC step.

However, what is not immediately clear is how these internal and external cues interact. The current findings suggest that if an external cue is available, it is likely to be used. This is signified by the lower error rates on cued trials in experimental participants, compared to non-cued trials in both experimental and control participants. When this external cue is unavailable, the internal associative links are still in place and can provide the required activation instead. Indeed, the current results support this by showing that performance on non-cued trials is no worse when strong visual cues have been available than when no such external primes were provided.

The second hypothesis that was investigated is whether participants anticipated the appearance of the cue. This could have provided a (weak) reminder for the PC step. This hypothesis was not supported by the error data, as it showed that experimental participants did not perform better than controls on non-cued target trials.

We further investigated this cue anticipation using the eye movement data. On average, control participants made more fixations prior to fixating on the PC step than those in the experimental condition. This further supports the finding that the cue successfully captured participants' attention. If anticipation occurred when the cue was absent, we would expect to see participants focus on the cue location within the same number of fixations for cued and non-cued trials. The eye movement data supported this, as it showed no significant difference in the number of fixations before a fixation on the PC step took place. Interestingly, this is not consistent with the corresponding finding in the error data, which showed that a similar number of PCEs were made on uncued experimental trials and control trials. As such, it is possible that *some* anticipation for the cue is taking place, but that this is not strong enough to remind participants to execute the PC step. As such, the current study is inconclusive about whether or not participants anticipate the appearance of the cue. Future studies must address this further.

A limitation of the current study is the relatively short duration of the experiment, and the somewhat limited training participants received. While previous studies have obtained good results with similar or less training (e.g. Li, Blandford, Cairns, & Young, 2008; Ament et al., 2010), it can be argued that the more experience participants have with doing the task, the more their performance represents true routine procedural performance. It is not known how long it takes before associative links fully form, or how long it takes for participants to learn to rely on visual cues. As such, it is possible that it takes much longer than the timescale of the current study for effects to become noticeable. Therefore, future studies can train participants for much longer on the cued and un-cued conditions of the task, and test the effects of unavailable cues after a larger number of habituation trials.

A related limitation of the current study is the small number of target trials each participant contributes. Due to experimental constraints and fatigue effects, it was not possible to include more target trials in the experimental phase, as it was essential that the disappearance of the cue remained unexpected for experimental participants.

The current study has important practical implications for the use of visual cues. The findings suggest that despite repeated exposure to a strong visual cue, people do not make more errors than un-cued controls when the cue is suddenly unavailable. This means that visual cues are an effective and safe way to reduce post-completion errors. Nevertheless, the current experiment also supports previous findings that the effects of visual cues are short-lived. When the cue is absent, for whatever reason, people revert to 'baseline' performance, and make as many errors as those who have never received visual cues. Therefore, in safety critical situations where errors could have severe consequences, additional measures should be in place in case the cue inadvertently fails.

Conclusion

In short, the current study showed that repeated exposure to strong visual cues has no reliable negative or positive effects on performance when the cue is absent. People do not become overly reliant on the cue. Instead, they revert to 'baseline' performance. Moreover, there is some evidence that people learn to anticipate the appearance of the cue, but this may not be strong enough to mitigate errors.

Acknowledgements

We thank Simon Li for the use of his Doughnut task.

References

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

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review*, *14*(6), 1079-1084.

Ament, M. G. A., Cox, A. L., Blandford, A., & Brumby, D. (2010). Working memory load affects device-specific but not task-specific error rates. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the cognitive science conference.*

Back, J., Cheng, W. L., Dann, R., Curzon, P., & Blandford, A. (2006). Does being motivated to avoid procedural errors influence their systematicity? In *Proceedings of hci 2006* (Vol. 1, p. 2-9).

Byrne, M. D. (2008). Preventing postcompletion errors: How much cue is enough? In *Proceedings of the thirtieth annual conference of the cognitive science society.* Austin, TX.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science*, *21*(1), 31-61.

Byrne, M. D., & Davis, E. M. (2006). Task structure and postcompletion error in the execution of a routine procedure. *Human Factors*, *48*(4), 627-638.

Chung, P. H., & Byrne, M. D. (2008). Cue effectiveness in mitigating postcompletion errors in a routine procedural task. *International Journal of Human-Computer Studies*, *66*, 217-232.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665-682.

Li, S. Y.-W., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: An account based on the activation-based goal memory model. *Journal of Experimental Psychology: Applied*, *14*(4), 314-328.

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. In *Proceedings of cogsci '06.* Vancouver, Canada.

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2008). Predicting postcompletion errors using eye movements. In *Proceedings of chi '08.*

Ratwani, R. M., & Trafton, J. G. (2009). Developing a predictive model of postcompletion errors. In *Proceedings of cogsci '09.*