Topics in Cognitive Science 2 (2010) 154–167

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01070.x

A Generalized Model for Predicting Postcompletion Errors

Raj M. Ratwani,[a,b] J. Gregory Trafton[a]

[a]*Naval Research Laboratory, Fairfax, Virginia*
[b]*Department of Psychology, George Mason University*

Received 1 October 2009; accepted 14 October 2009

Abstract

A postcompletion error is a type of procedural error that occurs after the main goal of a task has been accomplished. There is a strong theoretical foundation accounting for postcompletion errors (Altmann & Trafton, 2002; Byrne & Bovair, 1997). This theoretical foundation has been leveraged to develop a logistic regression model of postcompletion errors based on reaction time and eye movement measures (Ratwani, McCurry, & Trafton, 2008). This study further develops and extends this predictive model by (a) validating the model and the general set of predictors on a new task to test the robustness of the model, and (b) determining which specific theoretical components are most important to postcompletion error prediction.

Keywords: Human error; Procedural error; Eye tracking; Cognitive models; Predictive models; Interruptions

1. Introduction

Even while performing a routine procedural task that has been performed hundreds of times in the past, occasional errors still occur (Reason, 1990). These procedural errors have been termed skill-based errors (Rasmussen & Jensen, 1974) and occurred despite having the correct knowledge of how to perform a particular task. A common type of procedural error is the postcompletion error; this error is associated with forgetting a final step, which occurs after the main goal of a task has been completed (Byrne & Bovair, 1997; Byrne & Davis, 2006; Chung & Byrne, 2008). There are several examples of postcompletion errors, such as leaving an original document on the glass of a copy machine after making a copy or leaving the gas cap on top of the car after filling up the car with fuel.

Correspondence should be sent to Raj M. Ratwani, 4400 University Drive, MS3F5, Fairfax, VA 22030.
E-mail: rratwani@gmu.edu

The holy grail of error research is to be able to predict when an error is going to occur before the error actually occurs (Reason, 1990). To be able to make advances toward error prediction, strong theoretical accounts of the cognitive mechanisms underlying procedural errors are required (e.g., Botvinick & Plaut, 2004; Cooper & Shallice, 2000). In the case of postcompletion errors, these theoretical accounts do exist; Byrne and Bovair (1997) have put forward a theory specific to postcompletion errors, and Altmann and Trafton (2002) explain postcompletion errors using a general theory of goal memory, called memory for goals. Both theories are activation-based memory accounts and there is substantial overlap between the theories.

Byrne and Bovair (1997) suggest that postcompletion errors are due to goal forgetting and inattention to the postcompletion step. Specifically, postcompletion errors occur because the postcompletion step of a task is not maintained in working memory and, thus, is not executed as part of the task. The main goal of a task and the subsequent subgoals are stored in working memory and must remain active to be executed. The main goal provides activation to the subgoals. When the main goal of a task is satisfied, the goal no longer provides activation to the subgoals; consequently, the remaining subgoals may fall below threshold and may not be executed.

The memory for goals theory (Altmann & Trafton, 2002, 2007) accounts for goal-directed behavior with the constructs of activation and associative priming. The theory suggests that behavior is directed by the current most active goal and that the activation level of goals decay over time. In order for a goal to direct behavior, the goal must have enough activation to overcome interference from previous goals; thus, the goal must reach a certain threshold to actually direct behavior.

Goal activation is determined by two main constraints. The strengthening constraint suggests that the history of a goal (i.e., how frequently and recently the goal was retrieved) will impact goal activation. The priming constraint suggests that a pending goal will be retrieved and will direct behavior if the goal is primed from an associated cue. These cues can either be in the mental or environmental context.

Leveraging these theoretical accounts, Ratwani, McCurry, and Trafton (2008) developed a logistic regression model predicting when a postcompletion error will occur on a computer-based procedural task. A logistic regression analysis was used because the outcome variable (occurrence of an error) was a dichotomous variable, which violates many of the assumptions of standard linear regression (Tabachnick & Fidell, 2001). A simple description of logistic regression is that it is a multiple linear regression model with a dichotomous variable as an outcome variable; a more detailed description can be found in Peng, Lee, and Ingersoll (2002).

To build their logistic regression model, Ratwani et al. (2008) recorded and developed eye movement and reaction time measures as the behavioral indicators of the cognitive constructs outlined by the Byrne and Bovair (1997) and Altmann and Trafton (2002) theories. Specifically, three predictors were used in the logistic regression model: time between the postcompletion action and the prior action, total number of fixations between the postcompletion action and the prior action, and fixation on the postcompletion action button. The logistic regression model was as follows:

$$\text{Predicted logit of error} = .12 + (\text{time} \times -.001) + (\text{total fixations} \times .63) + (\text{postcompletion fixation} \times -5.7)$$

The motivation for these predictors followed directly from cognitive theory. The time predictor represents goal decay. The total number of fixations represents decay, but it may also capture individual differences in decay rates and differences in visual and cognitive processing demands (Just & Carpenter, 1976; Rayner, 1998). Finally, the fixation on the postcompletion action represents inattention/attention to this action and associative priming provided by the postcompletion action button on the task interface.

Using this model, over 90% of the postcompletion error and correct actions were correctly classified on the dataset from which the model was based. The successful classification rate with this model provides some confidence that the predictors, and, more importantly, the underlying theoretical constructs, which the predictors represent, are able to account for the cognitive mechanisms that contribute to postcompletion errors.

Despite the successful classification rate of this model, several important issues remain. First, the model was developed and tested on a single task. Consequently, it is difficult to determine how robust the specific model and the general set of predictors are. Second, the relative importance of each of the predictors, and the underlying cognitive constructs is unknown. For example, is decay a more important theoretical and predictive component than associative activation? How do we differentiate between the predictors?

We sought to address these issues in two ways. To determine the robustness of the logistic regression model and the set of predictors, the logistic regression model from Ratwani et al. (2008) was applied to a new task to see how well the model could account for postcompletion errors on that task. Performance of the model was compared with a task-specific logistic regression model using receiver-operating characteristic (ROC) curves and by examining classification success rates. To differentiate between predictors, discriminant function analysis (DFA) was utilized to linearly separate the predictors. This analysis provides insight into which specific predictors, and underlying cognitive constructs, are contributing the most to predictability.

2. Experiment

Reaction time and eye movement data were collected on a computer-based procedural task, called the financial management task. Like the sea vessel task used by Ratwani et al. (2008), this task has a postcompletion step. Aside from content of the tasks, the financial management task differs from the sea vessel task in two major ways. First, in the financial management task the order of subgoals to be completed is spatially congruent with the interface layout. Second, once a subgoal was completed on the financial management task, the information remained on the interface providing cues for the user to track progress on the task; these cues provide a method for global place keeping (Gray, 2000). The interface of

the sea vessel task did not provide these cues, thus increasing memory load of the user. The interface differences between the two tasks made the financial management task considerably more intuitive than the sea vessel task. Similar to the methodology used by Ratwani et al., participants were interrupted while performing the financial management task to increase the rate of postcompletion errors (Li, Blandford, Cairns, & Young, 2008).

To determine the robustness of the original logistic regression model (called the sea vessel model), this model was used to predict the occurrence of a postcompletion error on the financial management task. A task-specific logistic regression model (called the financial model) was also created, and the models were compared. We examined which predictors loaded significantly and the weights of the predictors. Although some differences are expected between the two models because they are based on different tasks, if the underlying theoretical constructs account for the cognitive mechanisms contributing to postcompletion errors the models should share the same general set of predictors and these models should have similar predictive power. Similarities in regression weights between the two models would suggest not only that the variables themselves are important but also that their relative contributions are similar.

To determine the relative contribution of the predictors in the logistic regression models, DFA was used. This statistical technique was used because logistic regression (unlike multiple regression) does not provide an indication of the importance of predictor variables (Tabachnick & Fidell, 2001). DFA is a linear technique that will provide an indicator of the importance of each predictor. DFA was used on the dataset from Ratwani et al. (2008) and on the data from this experiment to compare the relative importance of each the predictors across tasks.

2.1. Method

2.1.1. Participants

Twenty George Mason University undergraduate students participated for course credit.

2.1.2. Materials

The primary task was a complex financial management task. The goal of the task was to successfully fill clients' orders for different types of stocks. The orders were to either buy or sell and were presented four at a time at the top of the screen (see Fig. 1). The current prices of the stocks associated with the orders were presented in the center of the screen in the Stock Information column. The actual stock price fluctuated every 45 s.

To complete an order, participants first had to determine which of the client orders was valid by comparing the client's requested price with the actual market price of the stock from the Stock Information column. Once an order was determined to be valid, the participant clicked the Start Order button for the respective stock. To actually fill the order, the participant had to enter details from the order itself and the Stock Information column into eight different modules on the screen. Participants had to follow a specific procedure to complete the order; the specific sequence was as follows: Quantity, Cost, Order Info, Margin, Stock Exchanges, Transaction, Stock Info, and Review. The spatial layout of the

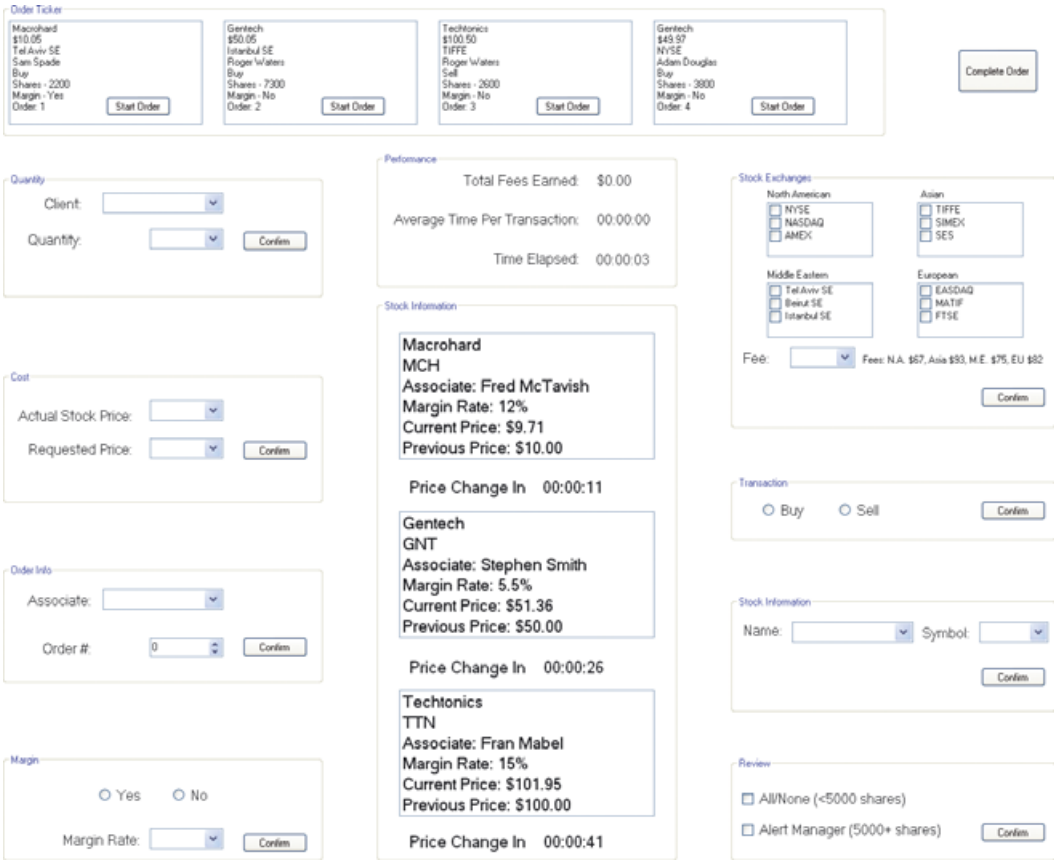

Fig. 1. Screenshot of the financial management task.

interface is quite intuitive (working down the left column and then the right column of Fig. 1), unlike the sea vessel task used by Ratwani et al. (2008).

After entering information in each module, the participant clicked the Confirm button and could then move on to the next module. After clicking Confirm on the final module (the Review module), a pop-up window appeared confirming the details of the order. The participant then had to acknowledge the window by clicking Ok. Finally, to complete the order the participant clicked the Complete Order button (upper right corner). This final action was the postcompletion step and the pop-up window is a false completion signal that is generally associated with postcompletion errors (Reason, 1990).

All of the information required to complete the task is directly available on the task interface. After completing a particular module and clicking the Confirm button, the information remained in the module providing a mechanism for global place keeping (Gray, 2000). If a participant attempts to work on a module or clicks a button that deviates from the strict procedure, the computer emits a beep signaling that an error has been made. The participant must then continue working on the task until the correct action is completed.

The interrupting task consisted of multiple choice addition problems. Each problem contained five single-digit addends and five possible solutions (4 incorrect, 1 correct). A single addition problem and solution set was presented at one time; participants completed as many problems as possible during the interruption.

2.1.3. Design

Control and interruption trials were manipulated in a within-participants design. The completion of one order on the financial management task constituted a trial. Participants completed 12 trials; six were control and six were interruption trials. The order of control and interruption trials was randomized. Each interruption trial contained two interruptions. There were eight possible interruption points. These points occurred after clicking the Confirm button following the first seven modules and after acknowledging the false completion signal, just prior to the postcompletion action. The location of the interruptions was randomized with the constraint that exactly two interruptions occurred just prior to the postcompletion step and at least one interruption occurred at each of the other seven possible locations. The interruption itself lasted for 15 s.

2.1.4. Procedure

Participants were seated approximately 47 cm from the computer monitor. After the experimenter explained the financial management and interrupting tasks to the participant, the participant completed two training trials (one with and one without interruptions). To begin the experiment, participants had to complete two consecutive error-free trials to ensure the task was well learned.

Each participant was instructed to work at his/her own pace. When performing the interrupting task, participants were instructed to answer the addition problems as soon as the solution was known. Upon resumption of the financial management task, the previous information that was entered prior to the interruption remained on the interface.

2.1.5. Measures

Keystroke and mouse data were collected for every participant. Eye track data were collected using a Tobii 1750 eye tracker (Tobii Technology Sweden) operating at 60 Hz. A fixation was defined as a minimum of five eye samples within 30 pixels (approximately 2° of visual angle) of each other, calculated in Euclidian distance. The Complete Order button was defined as an area of interest and subtended an area greater than 1.5°. This button was separated from its nearest neighbor by at least 2°.

A postcompletion error was defined as skipping the step of clicking the Complete Contract button and making an action that is related to a new order on the financial management task (e.g., erroneously attempting to click the Start Order button or attempting to work on the first module). The percent of postcompletion errors in the interruption trials was compared with the percent of postcompletion errors in control trials. The percent of errors was a ratio of the actual number of postcompletion errors to the opportunity for a postcompletion error.

The three predictors of interest (time, number of fixations, and fixation on postcompletion step) were calculated for every postcompletion action. In the cases where there was no

interruption prior to the postcompletion step, the period of measurement was from the completion of acknowledging the pop-up window to the next action (i.e., correctly making the postcompletion action or making an error). In the cases where an interruption occurred just prior to the postcompletion action, the period of measurement was from the onset of the financial management task immediately following the interruption to the next action. The time predictor was measured in milliseconds. The total number of fixations predictor was a count of the number of fixations. Fixation on the postcompletion action button was a binary variable that was coded 0 if the participant did not fixate on the postcompletion button and 1 if the participant fixated on the postcompletion button.

2.2. Results and discussion

2.2.1. Error rates

Participants made a total of 16 postcompletion errors; 12 participants made at least one postcompletion error. Participants made significantly more postcompletion errors during interruption trials (M = 11.7%) than control trials (M = .83%), $F(1, 19) = 12.8$, MSE = 91.7, $p < .01$. The low error rate in the control trials suggests that participants understood the procedures to complete the task. The large percent of postcompletion errors in the interruption trials suggests that even with an interface that has cues, which encourage global place keeping, participants still make postcompletion errors. Thus, simply providing a mechanism for keeping track of task progress is not enough to eliminate postcompletion errors.

2.2.2. Logistic regression models

The data from the financial management task (all postcompletion error and nonerror actions, regardless of trial type) were formulated into the three predictors of interest, as described in the measures section of the method. These predictors were used to create a task-specific logistic regression model (the financial model). The task-specific regression model was significant χ^2 (3) = 76.1, $p < .001$ and the equation was as follows:

$$\begin{aligned}\text{Predicted logit of error} = {} & -3.8 + (\text{time} \times .0008) + (\text{total fixations} \times .35) \\ & + (\text{postcompletion fixation} \times -5.7)\end{aligned}$$

Table 1
Logistic regression results

Predictor	β Sea Vessel	β Financial	SE β	Wald's χ2	df
Intercept	.12 (n.s.)	−3.8 ($p < .001$)	1.1	−3.8	1
Time	−.001 (n.s)	.0007 ($p = .2$)	.0005	1.4	1
Total fixations	.63 ($p < .001$)	.35 ($p < .05$)	.16	2.2	1
Postcompletion fixation	−5.7 ($p < .001$)	−5.7 ($p < .001$)	1.6	−3.5	1

Table 1 shows the logistic regression results for this model; the beta weights are under the column labeled ''β Financial.'' To the right of this column are the associated *SE*, Wald's test, and degrees of freedom values. As indicated in Table 1, time did not load significantly in the financial model.

Table 1 also displays the beta weights from the original sea vessel model. Comparing the weights between the models illustrates several interesting things about the predictors. The time predictor did not load significantly in either model. The total number of fixations predictor is significant in both models, suggesting that this is a fairly robust predictor; however, the weights of the predictors are quite different. While the likelihood of a postcompletion error increases with each additional fixation in both models, the increase in probability is not as drastic for the financial task model. The difference in the value of the weights may reflect differences in the perceptual processing required in each of the tasks.

The postcompletion fixation predictor loaded significantly in both models and the values of the weights are exactly the same. As participants fixate on the postcompletion action button the probability of a postcompletion error decreases by the same amount in each model. This comparison suggests the inattention and associative activation constructs represented by the predictor are prevalent across tasks and are important in accounting for errors.

Note that one of the assumptions of logistic regression is independence of data, which we are clearly violating in these models (e.g., a participant contributes multiple data points to the dataset). There are several reasons to believe that these models have some validity, even while violating this assumption. First, the similarity of the models given different datasets and different people suggests that in this case the violation is not critical. Second, we have cross-validated the original model on a different dataset and then used that model to prevent errors in real time (R. M. Ratwani & J. G. Trafton, unpublished data). Thus, the original model has been successfully used on at least four different datasets. These results suggest that the data independence violation is not critical to the success of the models.

Although there are some differences in the beta weights themselves, the predictors that loaded significantly in each model were the same. However, simply comparing the value of the weights and the significance of the weights does not provide insight as to how well the models can actually predict when a postcompletion error will occur. Specifically, can the sea vessel logistic regression model account for the occurrence of postcompletion errors on the financial management task?

2.2.3. Receiver-operating characteristic analysis

To begin to examine how well each model predicts the occurrence of a postcompletion error on the data from the financial management task, a ROC analysis was conducted. For each participant on the financial management task, his or her data from each postcompletion step was entered into each of the logistic regression models. Each model produced a predicted probability of postcompletion error. These predicted probabilities were then compared with the actual occurrence of an error to determine how accurate each model was.

Because the logistic regression models result in predicted probabilities, a threshold value must be established to categorize cases as errors and nonerrors. For example, Ratwani et al. (2008) suggested a threshold value of 75% for their model on the task on which it was

developed. This value means when the logistic regression model is applied to a particular postcompletion case, if the predicted probability is greater than or equal to 75%, the case should be classified as an error. If the predicted probability is less than 75%, the case should be classified as a nonerror.

A ROC analysis provides a method for visualizing the performance of the logistic regression models at different threshold values (Fawcett, 2006). To develop the ROC curves, these threshold values were systematically varied from 0% to 100% in each model. The predicted errors and nonerrors at each threshold value were compared with the actual data to generate the true positive and false positive rates. Each of these pairs of values was then used to generate the ROC curves seen in Fig. 2.

The ROC curves in Fig. 2 are plotted in ROC space. Points that fall in the upper left-hand corner represent perfect prediction; the points result in a high true positive rate and a low false positive rate. By visually examining Fig. 2, one can see that the ROC curves for the two models are nearly identical, suggesting that the sea vessel equation is robust enough to account for the data on the financial management task.

To quantitatively determine how robust the logistic regression models are at predicting postcompletion errors on this dataset, the area under the ROC curve can be examined. The area under the curve represents the probability that the logistic regression model will rank a

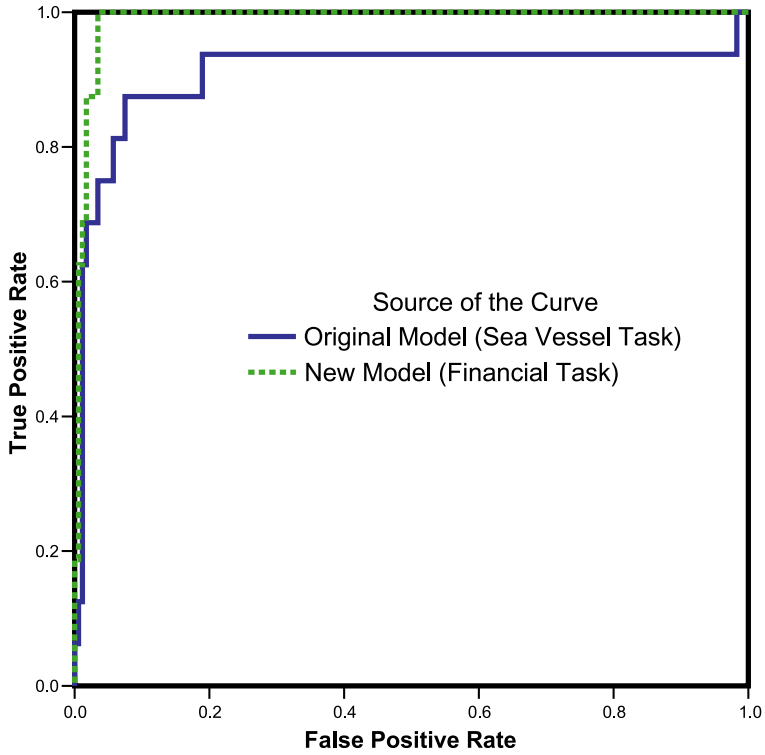

Fig. 2. ROC curves for the logistic models.

randomly chosen positive instance (i.e., an error) higher than a randomly chosen negative instance (i.e., nonerror) (Fawcett, 2006; Macmillan & Creelman, 2005). The area under the ROC curve for the sea vessel model is .91. The area under the curve for the financial model is .99. These values are considered excellent and suggest that the logistic regression models are correctly ranking nearly every case.

To illustrate the robustness of the sea vessel model, a confusion matrix was created by applying the sea vessel model to the financial dataset. A 75% threshold was used to classify nonerrors and errors. The 75% threshold was determined to maximize true positives and minimize false positives for this particular logistic regression model on the sea vessel task (R. M. Ratwani & J. G. Trafton, unpublished data). As can be seen in Table 2, the sea vessel model predicts more than 85% of the postcompletion errors. This result is based on taking the equation ''as is.''

A confusion matrix was also generated using the financial logistic regression model. This model was applied to the financial management dataset using a threshold value of 18.2%. Based on the ROC analysis, this threshold value maximizes true positives and minimizes false positives. Table 3 shows the results. Comparing the tables, both models performed well, but it is obvious that financial model outperforms the sea vessel model with a true positive rate of 100% and a true negative rate of nearly 97%.

The ROC analysis and the confusion matrix provide strong evidence for the robustness of the sea vessel model and for the set of predictors in the model. The sea vessel model predicted over 85% of the errors on the new financial task with the 75% threshold, suggesting that the underlying theoretical constructs are accounting for the cognitive mechanisms contributing to postcompletion errors, regardless of task.

From an applied aspect, note that both models accurately identify over 85% of actual errors and result in less than 10% false positives, despite the differences in threshold values. The robustness and accuracy of the predictors strongly suggests that these models could be used in an applied context to prevent errors before they occur. The fact that the model does not make many false alarms in either direction suggests that any system that relies on this information would be usable as well as accurate.

Table 2
Confusion matrix based on the sea vessel model

Predicted Value	Actual Value
True positive 14 (87.5%)	False positive 13 (7.5%)
False negative 2 (12.5%)	True negative 161 (92.5%)

Table 3
Confusion matrix based on financial model

Predicted Value	Actual Value
True positive 16 (100%)	False positive 6 (3.4%)
False negative 0 (0%)	True negative 168 (96.6%)

2.2.4. Discriminant function analysis

Logistic regression is an excellent statistical technique for maximizing predictive value, but it is difficult to draw conclusions about the strength of the relationship of the predictors and the outcome variable. Specifically, our interest was in determining which predictor has the strongest relationship to the occurrence of an error and to examine whether these relationships hold true across models and tasks. To answer these questions, DFA was used. DFA is a linear analysis technique to predict group membership from a set of predictors. Critically, DFA provides information on the strength of the relationship between each predictor and the outcome variable and the importance of these predictors. It is an aspect of DFA that we are most interested in.

Ratwani et al. (2008) did not perform a DFA on the sea vessel dataset. Therefore, we have taken that dataset, as well as the data from the financial management task, and conducted a DFA on each set of data. As expected, there was a strong association between the set of predictors and the occurrence of an error in Ratwani et al.'s (2008) dataset, χ^2 (3) = 350.9, $p < .001$. There was also a strong association between the set of predictors and the occurrence of an error, $\chi 2$ (3) = 130.7, $p < .001$, in the financial management dataset. These findings confirm the findings from logistic regression analyses.

To determine the relationship of each predictor to the classification function, we examined the canonical coefficients for each model. These coefficients indicate the unique contribution of each predictor to the classification function (outcome variables). This value is analogous to determining the unique variance accounted for by each predictor in multiple regression. Table 4 shows the canonical coefficients for each of the models on their respective datasets. Whether the coefficient is positive or negative it is irrelevant for determining the strength of association to the classification function.

In both models, the postcompletion fixation predictor has the strongest relationship to the classification function. This finding suggests that cue association and inattention to the postcompletion action are the most important theoretical components accounting for postcompletion errors. One explanation for this strong association is that a fixation on a to-be-completed action button is generally necessary before the physical clicking of the button in computer-based tasks. There were, however, several instances where participants fixated on the postcompletion action button and failed to complete the postcompletion step. In the original dataset that Ratwani et al. (2008) used, participants fixated on the postcompletion error button and still made a postcompletion error in 21% of the error cases; this occurred 13% of the time in the financial task dataset.

Table 4
Canonical coefficients for each model

	Standardized Canonical Discriminant Function Coefficients	
Predictor	Sea Vessel Model	Financial Model
Time	.35	.61
Total fixations	−.74	.29
Postcompletion fixation	.88	−.72

In the sea vessel model, the total fixations predictor has a stronger relationship to the classification function than time. However, in the financial model, time has a stronger relationship to the classification function than total fixations despite the fact that the time predictor did not load significantly in the model. While both of these predictors represent decay, total fixations may also represent individual differences in visual processing. The visual processing demands of the two tasks may be different and this may account for the differences in coefficients. The weights from the logistic regression models and the discriminant function coefficients suggest that the total number of fixations predictor and the time predictor are more variable across tasks than the postcompletion fixation predictor.

3. General discussion

The logistic regression and DFA analyses suggest that the postcompletion action fixation was the most important predictor. Most striking was the nearly similar beta weight for the postcompletion fixation in both models. We argue that these analyses provide strong evidence that inattention and cue association are critical theoretical components accounting for postcompletion errors, regardless of the task.

The time and total number of fixations predictors are also important components of the logistic regression model. However, several questions remain about the differences in the weights of the predictors and the differences in the DFA correlations between the predictors and the classification function. Total number of fixations may be accounting for individual differences in visual processing and possibly individual differences in decay rates. The variability in the total number of fixations predictor may be due to the different visual processing demands of the two tasks. It is unclear why time, a clear measure of goal decay, was not a significant predictor in either logistic model. Yet the DFA showed that time had a strong association to the occurrence of a postcompletion error in the financial model. It is clear that neither Byrne and Bovair's (1997) nor Altmann and Trafton's (2002) theories can adequately account for these subtleties in their current form. The relationship between these two predictors needs to be examined further.

Regardless of these differences, these results show the robustness of the predictive power of the original logistic regression model and the general set of predictors. This model accounted for over 85% of the postcompletion errors on a new task, even when the new task allowed for global place keeping, which alleviates memory load. These results suggest that, given a completely different task with a postcompletion step, it would not be necessary to run a new experiment to determine the variables and weights to use to predict when a postcompletion error would occur. The overall predictive power of this logistic regression model on two different tasks is encouraging and suggests that an understanding of the cognitive mechanisms underlying procedural errors can lead to prediction.

Some researchers have suggested that using complex statistical tools can result in data overfitting and extremely fragile models (Gigerenzer & Brighton, 2009; Roberts & Pashler, 2000). By using two different datasets and two different tasks, we have shown that our model

is robust and not overfit; in other work, we have also shown that our model can be used for real-time prediction (R. M. Ratwani & J. G. Trafton, unpublished data). We attribute our success of the current logistic models to our use of strong process models (Altmann & Trafton, 2002; Byrne & Bovair, 1997): The logistic model is a reflection of the process models.

Acknowledgments

This work was supported by grant N0001405WX20011 from the Office of Naval Research to Greg Trafton and by a fellowship from the National Research Council to Raj Ratwani. The authors thank Kristina O'Connell and Jennifer Chen for conducting the experiment.

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 and Review*, *14* (6), 1079–1084.

Botvinick, M. M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential behavior action. *Psychological Review*, *111*, 395–429.

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.

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*, 297–338.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27* (8), 861–874.

Gigerenzer, G., & Brighton, H. (2009). Home heuristics: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gray, W. D. (2000). The nature and processing of errors in interactive behavior. *Cognitive Science*, *24* (2), 205–248.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.

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.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theories: A user's guide*. Mahwah, NJ: Erlbaum.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*, 3–14.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics*, *17*, 293–307.

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2008). Predicting postcompletion errors using eye movements. In M. Czerwinski, A. M. Lund, & D. S. Tan (Eds.), *Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008* (pp. 539–542): ACM.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *85*, 618–660.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.