Title
Cognitive Strategies in HCI and Their Implications on User Error

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Halbrügge, Marc
Quade, Michael
Engelbrecht, Klaus-Peter

Publication Date
2016

Peer reviewed

Cognitive Strategies in HCI and Their Implications on User Error

Marc Halbrügge (marc.halbruegge@tu-berlin.de)
Quality and Usability Lab, Telekom Innovation Laboratories
Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin

Michael Quade (michael.quade@dai-labor.de)
DAI-Labor, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin

Klaus-Peter Engelbrecht (klaus-peter.engelbrecht@alumni.tu-berlin.de)
Quality and Usability Lab, Telekom Innovation Laboratories
Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin

Abstract

Human error while performing well-learned tasks on a computer is an infrequent, but pervasive problem. Such errors are often attributed to memory deficits, such as loss of activation or interference with other tasks (Altmann & Trafton, 2002). We are arguing that this view neglects the role of the environment. As embodied beings, humans make extensive use of external cues during the planning and execution of tasks. In this paper, we study how the visual interaction with a computer interface is linked to user errors. Gaze recordings confirm our hypothesis that the use of the environment increases when memory becomes weak. An existing cognitive model of sequential action and procedural error (Halbrügge, Quade, & Engelbrecht, 2015) is extended to account for the observed gaze behavior.

Keywords: Human Error; Memory for Goals; Eye-Tracking; ACT-R; Cognitive Modeling

Introduction and Related Work

Our daily life is dominated by routine sequential behavior like making coffee, washing clothes, or commuting to work. While we have practiced these activities hundreds of times, they are still subject to sporadic errors (Reason, 1990), a well-known example being postcompletion errors (e.g., forgetting a bank card in a vending machine after having completed a purchase, Byrne & Davis, 2006).

This paper presents recent advances of a cognitive modeling project on sequential behavior and error (Halbrügge & Engelbrecht, 2014; Halbrügge et al., 2015). The model is based on the memory for goals (MFG; Altmann & Trafton, 2002) framework which formulates how sequential behavior is controlled using declarative memory. We call this the *knowledge-in-the-head* strategy (Norman, 2002). By adding a second cognitive strategy to the model that relies more on external cues than on internal memory (the *knowledge-in-the-world* strategy; Norman, 2002), we can reproduce omission rates for different sub-tasks, comprising but not limited to postcompletion errors. Furthermore, the model can account for intrusions, an error type that has not been captured by MFG-based cognitive models before.

While the original model is targeting only error rates, the addition of the in-the-world strategy predicts behavior changes in other areas as well that allow empirical validation. In the visual domain, using the bottom-up in-the-world strategy means continously searching the UI for any suitable element, while the top-down in-the-head strategy just looks for the specific element that is needed to perform the current task. We therefore designed an eye-tracking experiment that should reproduce our previous findings and should confirm the predictions of the knowledge-in-the-world assumption in the visual domain.

The rest of this paper is organized as follows. We first give an overview on the types of errors we are aiming to address alongside the theoretical underpinnings of our model. We will then introduce an experiment that shows the connection of visual behavior to procedural error and compare the findings to the predictions of our cognitive model. We conclude with a discussion of the strengths and limitations of the model and a summary of our contributions.

Procedural Error

Human error in its most general sense is commonly refered to as "those occasions in which a planned sequence of mental or physical activities fail to achieve its intended outcome, [and] when these failures cannot be attributed to the intervention of some chance agency" (Reason, 1990). It can be further decomposed based on the level of action control on which an error occurs (Rasmussen, 1983). Knowledge-based behavior on the highest level of control is characterized by explicit planning. Skill-based behavior on the opposite, lowest level consists mainly of sensory-motor actions without conscious control. Interaction with computer systems is mainly located on the intermediate rule-based level of action control. On this level, behavior is generated using stored rules and procedures that have been formed during training or earlier encounters. Errors on the rule-based level are not very frequent (below 5%), but pervasive and cannot be eliminated through training (Reason, 1990). While Norman (2002) subsumes these errors within the 'slips' category, Reason (1990) refers to them as 'lapses'. We escape this ambiguity by using the term *procedural error*. Procedural error is defined as the violation of the (optimal) path to the current goal by a non-optimal action. This can either be the addition of an unnecessary or even hindering action, which is called an *intrusion*. Or a necessary step can be left out, constituting an *omission*.

Memory for Goals Theory

An explanation of procedural *error* must incorporate the generation of *correct* behavior as well. A very promising theoretical model of sequential action is the Memory for Goals theory (MFG; Altmann & Trafton, 2002). The MFG proposes that subgoals, i.e., atomic steps towards a goal, are represented in human memory, thereby underlying memory effects like time-dependent and noisy *activation*, *interference*, and associative *priming*.

Within the MFG theory, errors arise when the activation of a goal is not high enough to surpass interfering goals or even falls below a general retrieval threshold. Cognitive models based on the MFG assumptions have been shown to explain procedural errors in the HCI domain, namely omissions, very well (e.g., Trafton, Altmann, & Ratwani, 2011; Hiatt & Trafton, 2015; Halbrügge et al., 2015; Li, Blandford, Cairns, & Young, 2008).

The Role of the Environment

The MFG theory is clearly focused on managing task sequences in memory, i.e., within their head. This has been critized for neglecting the role of the environment (e.g., Salvucci, 2010). As embodied beings, humans strive to reduce cognitive complexity by exploiting the content and structure of the external world.

Recent research has shown that better predictions can be achieved and new error domains can be covered by extending the MFG with an activation process that relies on external cues (Halbrügge et al., 2015; Hiatt & Trafton, 2015). We propose that when a user cannot retrieve the next goal, they revert to an externalization strategy and randomly search the visual scene (i.e., the UI) for interactive elements. Whenever an element is found, the user tries to retrieve a goal that relates to this UI element. Because visually attending the element increases the activation of related goals through priming, a goal that had previously been forgotten may now surpass the retrieval threshold. As a consequence, the planned sequence of actions may be resumed. But because visual priming is independent of the currently planned sequence, it may also help retrieve an outdated goal from a previous trial, or the planned sequence may be resumed at an incorrect position.

Task- and Device-Orientation

The best known examples of procedural error are postcompletion errors (e.g., forgetting the originals in the copy machine; Byrne & Davis, 2006) and initialization errors (e.g., forgetting to reset Caps Lock before typing a password; Gray, 2000). Common to both of them is that these errors happen during procedural steps that do not directly contribute to the users' actual goals (i.e., making copies; logging into a system). This common property of goal-irrelevance of a sub-task has been coined *device-orientation* (Ament, Cox, Blandford, & Brumby, 2013; Gray, 2000), its opposite is analogously called *task-orientation*. The concept of device-orientation is based on the MFG by assuming that device-oriented tasks are "more weakly represented in memory". While Ament et al. (2013) discuss different encoding or lack of rehearsal as possible reasons for lower activation of device-oriented tasks, we are assuming lack of priming in this case (Halbrügge & Engelbrecht, 2014). We will call this the *task priming assumption* in the following.

The Role of the Application Logic

The elimination of device-oriented tasks is a reasonable design strategy for error reduction, but does not succeed in all cases. When a device-oriented step can not be avoided (e.g., the removal of the bank card from a teller machine), it is often made obligatory to make sure that users can not omit it. In the ATM example, this would mean not handing out the money before the user has taken back their card.

Previous research has shown that the impact of whether a subtask is obligatory or not is much higher than whether it is device-oriented or not, and the interaction of both needs to be taken into account when researching procedural error in real-world settings (Halbrügge et al., 2015). It is also worth noting that memory-based processing as put forth by the MFG can hardly explain why obligatory tasks are less prone to omissions. The *knowledge-in-the-world* assumption fills this gap.

Experiment

As errors are relatively infrequent, but proper statistical analysis needs sufficiently big case numbers, researchers have developed several strategies to artificially increase error rates in the laboratory. Among these are interruptions by the experimenter (Li et al., 2008; Trafton et al., 2011), secondary tasks (Byrne & Davis, 2006; Ruh, Cooper, & Mareschal, 2010), or special UIs that make users feel lost easily (Hiatt & Trafton, 2015). Because we want to maximize external validity, we rejected all of these options and chose to observe human error during repeated interaction with a real world application instead. We selected a kitchen assistance system from a "smart home" environment for the experiment. The assistant aims at helping with the preparation of meals by suggesting recipes, calculating ingredient amounts and maintaining shopping lists. A screenshot of the recipe search screen of the kitchen assistant is displayed in Figure 1.

Methods

Participants 24 members of the Technische Universität Berlin paid participant pool, 15 women and 9 men, aged between 18 and 54 (M=29.4, SD=9.4), took part in the experiment conducted in January 2015. As the instructions were given in German, only fluent German speakers were allowed.

Materials A personal computer with 23" (58.4 cm) touch screen and a 10" (25.7 cm) tablet were used to display the interface of the kitchen assistant. While the large screen operated in landscape mode, portrait mode was used for the tablet. We created two variations of the UI of the kitchen assistant that were targeted at the large screen and the tablet, respectively. All user actions were recorded by the computer system.

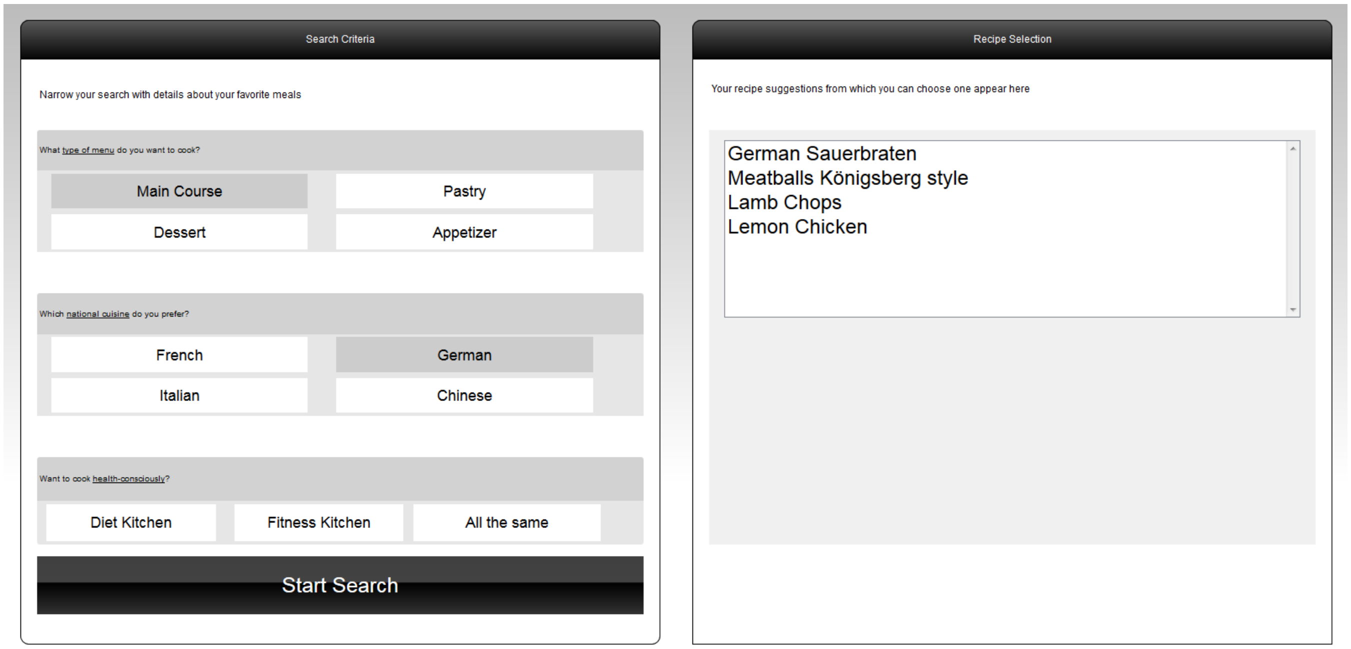

Figure 1: Screenshot of the English version kitchen assistant used for the experiment.

The participants' gaze was recorded using a SR Research Ltd EyeLink II head-mounted eye-tracker. Only the dominant eye was tracked; the sampling frequency was 250 Hz. Gaze recording was only applied during task blocks that used the large monitor because the visual angles between UI elements on the tablet were too small to obtain an unambigous mapping from gaze to element. Nevertheless, the eye-tracker was worn by the participants during the complete procedure as its scene camera view was recorded for subsequent error identification. The experiment was conducted in a laboratory with fixed lighting conditions.

Design We used a four-factor within-subject design, the factors being the physical device used, the UI variant, whether a sub-task was obligatory, and whether it was device-oriented as opposed to task-oriented. We collected user errors, task completion times, and gaze position. The participants completed a total of 46 tasks grouped into 4 blocks. Physical device, UI variant, and block sequence were varied randomly, but counterbalanced across the experiment.

Procedure After having played a simple game on each of the two devices to get accustomed with the respective touch technology, the participants received training on the kitchen assistant. The training covered all parts of the application that were used during the actual experiment. Each block of tasks began with relativly simple tasks like "Search for German main dishes and select lamb chops". Afterwards, the ingredients to a recipe were collected and some of them were added to a shopping list that is part of the kitchen assistant (e.g., "Create a shopping list for six servings and check-off garlic"). The instructions followed the experiment described in Halbrügge et al. (2015) closely, with only one new type of device-oriented non-obligatory tasks added in order to gain more insights about this special combination. When a participant made an error during a trial, they were not interrupted but informed after the trial and were given the chance to repeat this trial a single time. The complete procedure lasted approximately 45 minutes.

Results

Errors A total of 6921 clicks were recorded. We observed 85 (1.2%) omissions and 53 (.8%) intrusions. Because of the unequal number of observations per trial, mixed models were used to analyze the data (Bates, Maechler, Bolker, & Walker, 2013). Neither physical device nor UI variant showed relations to the error rate (mixed logit model with subject and task block as random factors, all $p > .3$). Obligatory task steps were less prone to errors ($z = -2.1, p = .034$) and device-oriented steps showed higher error rates than their task-oriented counterparts ($z = 7.8, p < .001$). Together with the highly significant interaction ($z = 4.4, p < .001$), this is mainly due to the relatively high omission rate for device-oriented non-obligatory task steps (e.g., checking off previously selected ingredients; see Figure 2).

Eye-Tracking The gaze of three participants could not be recorded because of calibration failure. For the remaining 21 participants, the raw screen coordinates were mapped to dynamic areas of interest (AOI) around the UI elements using a computational bridge between the eye-tracker and the web browser that rendered the UI (Halbrügge, 2015). Individual gaze positions were collapsed into fixations using a hidden Markov model (HMM) approach. Based on common rule-of-

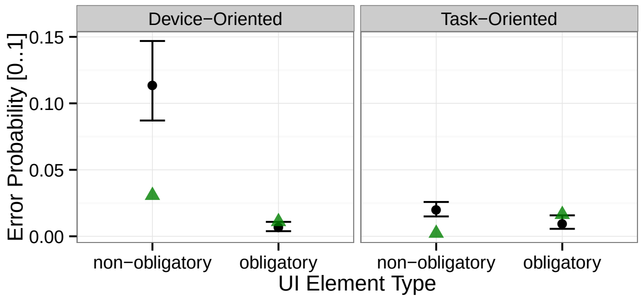

Figure 2: Error probabilities for the experiment (logarithmic scale). Error bars are 95% confidence intervals using the Agresti-Coull method. ▲ denote predictions of the revised model (400 runs).

thumb values of maximum 100 deg/s for fixations and minimum 300 deg/s for saccades (Salvucci & Goldberg, 2000), the parameters for the states of the HMM and the respective transition probabilities were estimated from the recorded data.

In order to examine the knowledge-in-the-world assumption, the gaze recordings were split into segments based on the clicks within a trial (e.g., a trial with five clicks yielded four segments). Because of the varying length of these segments, we use the fixation rate (number of fixations divided by segment length) as dependent variable. A mixed model with subject and sub-task as random factors yielded significantly higher rates for segments before erroneous clicks compared to correct clicks ($F_{2,1952.5} = 5.3$, $p = .005$, see Figure 3). This could still mean that the users were recurrently fixating the ultimately clicked UI element instead of searching the screen. We addressed this by counting the fixations on the AOI corresponding to the element that concluded its segment. Because the resulting counts are extremely right-skewed (median = 1), we divided them into two groups for statistical analysis. A logit mixed model with subject and sub-task as random factors revealed that the probability to fixate the ultimately clicked element at least once during a segment was actually lower before errors ($z = -2.5, p = .012$).

Discussion

The results partially reproduce the findings of our previous studies. Compared to Halbrügge et al. (2015), the combination of non-obligatory device-oriented tasks yielded a much higher error rate. This is probably due to the changed set of user tasks that seems to have included device-oriented task steps that were much harder to remember than the ones used before.

The eye-tracking data confirms the knowledge-in-the-world assumption nicely. The higher fixation rate before erroneous clicks matches the proposed process of (random) search for 'inviting' elements on the surface of the UI. In principle, this could also be caused by memory-based processing, e.g., a re-fixation strategy to strengthen the activation of the current subgoal through visual priming. But the AOI-based

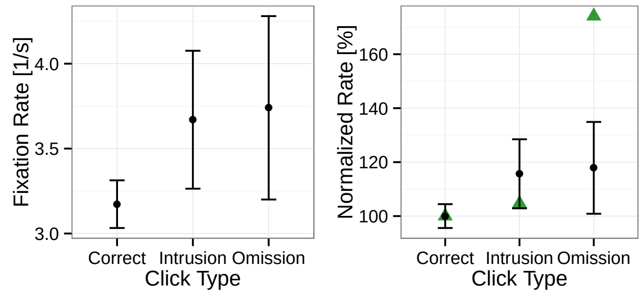

Figure 3: Fixation rate before performing a correct vs. an erroneous click. Error bars are 95% confidence intervals. Right: Rates have been normalized to the baseline of correct clicks to allow a comparison between data and cognitive model. ▲ denote predictions of the revised model (400 runs).

analysis shows that the frequent fixations before errors are on other elements than the one that is eventually clicked.

Cognitive Model

To gain more insight in the implications of our theoretical assumptions, the *task priming* and the *knowledge-in-the-world* assumptions have been implemented in a cognitive model based on the ACT-R framework (Anderson et al., 2004). A simplified flow-chart of the model is displayed in Figure 4. The model had been fit to the error rates of the previous experiment and reached good fits in this domain (Halbrügge et al., 2015).[1] The new data presented here allows constraining the existing model. The model fit to the error rates of the current data is still good with R^2=.76, but somewhat increased RMSE=.044. The eye-tracking results confirm the theoretical assumptions at least qualitatively. But what about the actual quantitative predictions of the model in the visual domain?

As ACT-R's visual module operates on an abstract attention layer (as opposed to raw eye movements), the model's visual behavior cannot be compared directly to the human sample. Nevertheless, effects that were significant in the human data should be present in the model results as well. The initial model produces fixation rates that are actually lower for erroneous trials compared to correct trials, in contrary to both the theory and the human data. This happens because the memory test that is part of the knowledge-in-the-world strategy (try-retrieve-goal-for-element in Figure 4) is not restricted in time. If it fails to retrieve a matching goal chunk, this is only signaled after approximately one second. As a result, the visual search process is slowed down considerably.

Revised Model

In order to speed up the knowledge-in-the-world strategy, we make use of ACT-R's temporal buffer (Taatgen, Van Rijn, & Anderson, 2007). A timer is started with the initiation of the memory test and a single production was added that aban-

[1]The source code of the model is available for download at http://www.tu-berlin.de/?id=135088

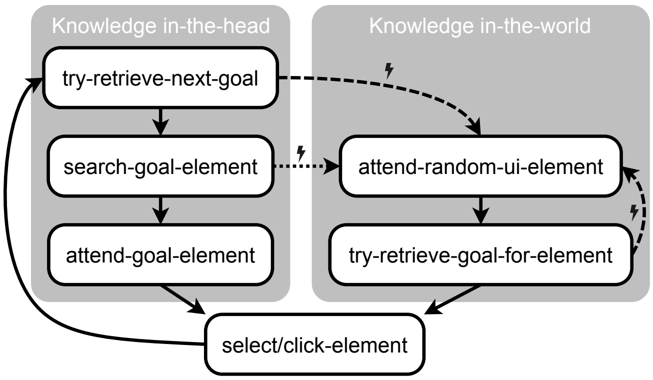

Figure 4: Simplified flow chart of the cognitive model. Dashed arrows denote retrieval errors, the dotted arrow denotes visual search failure.

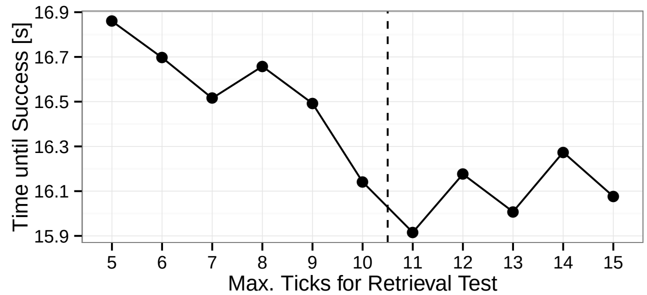

Figure 5: Average time until successful task completion when using the knowledge-in-the-world strategy as function of the maximum time ticks spent on the retrieval test. The vertical dashed line demarks the average wait tick value chosen for the model. (3000 model runs)

dons the retrieval after a fixed amount of time ticks. Based on the observed fixation rate between 3 and 4 per second before erroneous clicks, we set the tick threshold to approximately 250 ms (9 to 12 ticks using ACT-R standard parameters).

Comparing the model predictions based on the fixation rates is tricky because ACT-R only models shifts of attention and assumes completely stable gaze otherwise, which results in unnaturally low fixation rates. An exploratory approach to this based on the relative change compared to the 'correct click' baseline is shown in Figure 3. The corresponding fit is unimpressive with R^2=.41, but should be interpreted with care. Of higher importance is the qualitative result that error trials show increased fixation rates.

Errors After having optimized the model for visual behavior, the fit in the error domain somewhat degrades with R^2=.70 and RMSE=.044 (see Figure 2).

General Discussion

The current paper presents an empirical study and a cognitive user model of procedural error. The study has been designed to foster external validity, featuring a real-world application in a household scenario and drawing participants mainly from non-student populations. The model extends the activation-based Memory for Goals theory (Altmann & Trafton, 2002) by highlighting the importance of external cues during sequential action. Internal cues are divided into task-oriented, i.e., steps that directly contribute to the users' goals, and device-oriented ones (Ament et al., 2013). Together with the assumption that only task-oriented steps receive additional priming from the user's overall goal, this allows not only to explain our data, but also provides an acceptable explanation of how device-orientation actually affects sequential behavior. The role of external cues as formulated in the knowledge-in-the-world assumption is fostered by the eye-tracking results.

Eye-tracking has been used before to predict special subtypes of procedural error (Ratwani & Trafton, 2011) with good success. Ratwani and Trafton's system uses a cognitive model based on the MFG theory in combination with a statistical classifier that uses eye-tracking data as input to predict and prevent postcompletion error. The approach presented in this paper differs therefrom in two important ways. Firstly, we are targeting not only omissions of the final step within a sequence, but other types of omissions and intrusions as well. And secondly, we are trying not only to predict *when* a user makes an error, but also *why* this happens and have therefore incorporated the necessary visual processes into our cognitive model.

How do the strategies that we propose relate to the soft constraints hypothesis (SCH; Gray & Boehm-Davis, 2000; Gray, Sims, Fu, & Schoelles, 2006)? According to this hypothesis, users select (micro-) strategies based on a temporal cost-benefit tradeoff. The reduction of the waiting time for the retrieval test that is part of the knowledge-in-the-world strategy can already be viewed as the application of such a tradeoff. After a few hundred milliseconds, it may be more beneficial to try a new visual target than to wait for the successful retrieval of a matching goal chunk. We explored this hypothesis by varying the maximum time ticks the model waits until abandoning the retrieval test and comparing the success rates of the knowledge-in-the-world strategy depending on the ticks. Keeping all other model parameters fixed, the results show that the wait time chosen based on the fixation rates is also close to a local minimum of the time spent per successful click (see Figure 5). This criterion represents a time/benefit tradeoff as proposed by the SCH.

While providing good fits to the data, the model also has several limitations. Firstly, the model only covers expert behavior. The initial formation of the task sequence by novice human users is beyond its capabilities. The model also does not account for errors caused by the UI design violating general expectations of its users towards computer systems.

Secondly, the empirical basis of the model is limited. So far, we have only collected data using a single paradigm. Changing the experimental tasks resulted in changed error rates, but the overall pattern remained stable and the model fit is still satisfactory. While the current addition of eye-tracking data represents a significant extension of its empirical foundation, the generalizability of the model still needs further investigation.

Conclusions

We have presented a model of procedural error grounded in the Memory for Goals theory. The model consists of two interdependent cognitive strategies, one relying only on memory and another that is mainly using cues from the environment. The underlying assumptions, namely the *task priming assumption* and the *knowledge-in-the-world assumption*, are confirmed by an error analysis of newly gathered data without further parameter fitting. Eye-tracking data recorded at the same time qualitatively backs our theoretical assumptions. A revised cognitive model provides reasonable fit to both the error and the gaze data.

Acknowledgments

This work was supported by DFG grant MO 1038/18-1 ("Automatische Usability-Evaluierung modellbasierter Interaktionssysteme für Ambient Assisted Living").

References

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

Ament, M. G., Cox, A. L., Blandford, A., & Brumby, D. P. (2013). Making a task difficult: Evidence that device-oriented steps are effortful and error-prone. *Journal of experimental psychology: applied*, *19*(3), 195.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036–1060.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. (R package version 1.0-5)

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

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

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*(4), 322.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological review*, *113*(3), 461.

Halbrügge, M. (2015). Automatic online analysis of eye-tracking data for dynamic HTML-based user interfaces. In K. Gramann & T. O. Zander (Eds.), *Berliner werkstatt mensch-maschine-systeme.* Berlin: Technische Universität Berlin.

Halbrügge, M., & Engelbrecht, K.-P. (2014). An activation-based model of execution delays of specific task steps. *Cognitive Processing*, *15*, S107-S110.

Halbrügge, M., Quade, M., & Engelbrecht, K.-P. (2015). A predictive model of human error based on user interface development models and a cognitive architecture. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of the 13th international conference on cognitive modeling* (p. 238-243). Groningen, the Netherlands: University of Groningen.

Hiatt, L. M., & Trafton, J. G. (2015). An activation-based model of routine sequence errors. In N. A. Taatgen, M. K. van Vugt, J. P. Borst, & K. Mehlhorn (Eds.), *Proceedings of the 13th international conference on cognitive modeling* (p. 244-249). Groningen, the Netherlands: University of Groningen.

Li, S. Y., 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.

Norman, D. A. (2002). *The design of everyday things*. Basic books.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics, IEEE Transactions on*, *13*, 257–266.

Ratwani, R. M., & Trafton, J. G. (2011). A real-time eye tracking system for predicting and preventing postcompletion errors. *Human–Computer Interaction*, *26*(3), 205–245.

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

Ruh, N., Cooper, R. P., & Mareschal, D. (2010). Action selection in complex routinized sequential behaviors. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 955.

Salvucci, D. D. (2010). On reconstruction of task context after interruption. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 89–92).

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78).

Taatgen, N. A., Van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: the role of cognition, attention, and learning. *Psychological Review*, *114*(3), 577.

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