Int. J. Human-Computer Studies 79 (2015) 51–65

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Brief Lags in Interrupted Sequential Performance: Evaluating a Model and Model Evaluation Method ☆

Erik M. Altmann [a], J. Gregory Trafton [b]

[a] Department of Psychology, Michigan State University, East Lansing, MI
[b] Naval Research Laboratory, Washington, DC

A R T I C L E I N F O

Article history:
Received 6 March 2014
Received in revised form
10 December 2014
Accepted 27 December 2014
Communicated by E. Motta
Available online 5 January 2015

Keywords:
Task interruption
Sequence errors
Cognitive modeling
Goodness-of-fit testing

A B S T R A C T

We examined effects of adding brief (1 second) lags between trials in a task designed to study errors in interrupted sequential performance. These randomly occurring lags could act as short breaks and improve performance or as short interruptions and impair performance. The lags improved placekeeping accuracy, and to interpret this effect we developed a cognitive model of placekeeping operations, which accounts for the effect in terms of the lag making memory for recent performance more distinct. Self-report data suggest that rehearsal was the dominant strategy for maintaining placekeeping information during interruptions, and we incorporate a rehearsal mechanism in the model. To evaluate the model we developed a simple new goodness-of-fit test based on analysis of variance that offers an inferential basis for rejecting models that do not accommodate effects of experimental manipulations.

1. Introduction

1.1. Background

Many everyday tasks have two important characteristics that interact to elevate the chances of a performance error. One is sequential constraints: A set of steps has to be performed in some prescribed order and an error occurs when a step is skipped or repeated. For example, in the medical domain, one might forget to record a dose of medication in a log (a skipped step), which could then lead to administering a second dose (a repeated step). Sequential constraints are common in medicine, equipment maintenance, computer programming and technical support, data analysis, legal analysis, accounting, and many other home and workplace environments. Sequential constraints also play a role in such basic cognitive processes as language production, event counting, serial recall, and problem solving. To perform correctly under sequential constraints, the cognitive system has to keep track of where it is in the sequence and select the correct next step when one step is complete, a process we refer to as *placekeeping*.

The second characteristic is the possibility of interruption: In the middle of a task the phone might ring, an email might arrive, or a glitch or subgoal of some kind might arise in the primary task. Experience suggests that interruptions like this often lead to "where was I?" moments afterwards, and in fact interruptions generate substantial performance costs at the point where the interrupted task is resumed (e.g., Altmann and Trafton, 2007; Hodgetts and Jones, 2006; Monk et al., 2008).

That said, errors in sequential performance can be a challenge to study, both in general and after interruptions, because they are relatively infrequent in most tasks that it makes sense to have people perform. In routine procedures like making coffee, for example, error rates in one study reached only 4% even in the condition where interruptions were timed to be most disruptive (Botvinick and Bylsma, 2005). To obtain enough errors to analyze, researchers have variously studied neurological patients (Cooper et al., 2005) and used diary methods to expand the temporal window during which errors can occur (Reason, 1990). In laboratory tasks, a common approach is to structure the task environment to increase memory load. This can be done by including "post-completion" steps (Li et al., 2008), which are difficult to remember to begin with (Byrne and Bovair, 1997), or by including an ongoing task that makes it easy to forget to return to the interrupted task (Dodhia and Dismukes, 2009). Perhaps the most common device is to eliminate any cues in the task display that could tell participants where they were in the task sequence (e.g., Brumby et al., 2013; Gray, 2000; Trafton et al., 2011).

In recent work we developed a new task to study errors in interrupted sequential performance (Altmann et al., 2014). As in other interruption tasks there are no external placekeeping cues, but we also designed the stimulus materials and decision rules to generate enough perceptual and cognitive load that placekeeping

☆This research was supported by grants from the Office of Naval Research, N000140910093 and N000141310247 to the first author and N0001412RX20082 and N0001411WX30014 to the second author.

E-mail addresses: ema@msu.edu (E.M. Altmann), greg.trafton@nrl.navy.mil (J.G. Trafton).

http://dx.doi.org/10.1016/j.ijhcs.2014.12.007
1071-5819/

operations have to compete with task steps for system cycles, and to generate enough variability from trial to trial that processing does not become routine. The task is also continuous, producing many opportunities for error and many opportunities to interrupt participants between steps of the primary task.

The error data generated by this task are rich enough to be analyzed as a function of multiple experimental factors and interactions (Altmann et al., 2014). For example, interruption effects are substantial, but there are also enough errors on trials not preceded by interruptions to shed some light on placekeeping under baseline conditions. Errors also form gradients as a function of the "offset" of the incorrect step from the correct step within the sequence, and the shapes of these gradients interact with interruption effects. All told, the empirical patterns are complex enough to provide strong constraints on a theory of the underlying mechanisms.

1.2. Present study

In the present study we address an interrelated set of applied, theoretical, and methodological goals concerning interrupted sequential performance. The applied question is whether slowing people down a little can improve placekeeping accuracy. There is considerable evidence that people can trade speed for accuracy strategically (e.g., Wickelgren, 1977), and there is evidence from interruptions research in particular that linking errors to a high time cost improves accuracy (Brumby et al., 2013). Of interest here is whether a lower bound on the time between events—not an upper bound on time to respond, as in deadline procedures, but a brief lockout period in which there is no processing to be done—has the side effect of improving accuracy. To address the question we added brief (1 second) lags randomly between trials of our task, and compared performance on trials preceded by a lag with trials preceded immediately by another trial.

We also wanted to investigate rehearsal as a placekeeping strategy during interruptions. Rehearsal is a core strategy in memory procedures (e.g., Baddeley et al., 1975; Reitman, 1974), but beyond an earlier study of ours (Trafton et al., 2003) there seems to be little research evaluating the empirical prevalence of rehearsal in context of task interruption. Here we include a self-report measure asking participants to indicate, after the experimental session, if they used any strategies to keep their place in the interrupted task.

Our theoretical goal is to develop a cognitive model of placekeeping mechanisms that explains the effect of brief lags and the role for rehearsal if we find evidence for it, and that accounts for the complex empirical patterns in data from our task more generally. As we suggested above, placekeeping seems to be a general capability expressed in many different tasks, so such a model could inform our understanding of errors in many different contexts.

The basic theoretical premise in our model is that placekeeping involves two interacting memory systems, one that stores episodic information about what steps were recently performed, and another that stores a long-term associative representation of the task sequence. When the cognitive system has finished performing one step in a sequence, it selects the next step by first remembering what step it just performed, then using that memory to index into the associative representation of the task sequence to find that step's successor. Skipped or repeated steps arise from errors in these two retrieval operations.

In context of this basic theoretical framework, several cognitive mechanisms could lead to improved accuracy after brief lags, each by sharpening memory for the most recently performed step and thereby improving accuracy in looking up the next step. Cowan (1999) proposed that an item does not decay as long as it remains in the focus of attention. One possible illustration of this mechanism is that participants in discrimination learning tasks hold tightly to their most recent hypothesis over a long series of trials if given no feedback to update it (Frankel et al., 1970). In context of our lag condition, if information about the most recently performed step remains in the focus of attention during the lag, then it will not decay—even as information about earlier steps that is not in the focus of attention does decay. Thus, after a lag, the most recently performed step will be more active in memory in relation to earlier steps, leading to more accurate selection of the next step.

Another mechanism is a "strengthening" process that has played a role in previous models of goal-directed performance (Altmann and Gray, 2008; Altmann and Trafton, 2002; Trafton et al., 2011). Strengthening hypothetically takes some time but could be deployed during a brief temporal lag to maintain the activation of relevant control information. A related construct is the attentional refresh process found in some models of working memory (Barrouillet et al., 2004; Oberauer and Lewandowsky, 2011). Strengthening and attentional refreshing are more active and strategic whereas Cowan's (1999) mechanism is more passive and structural, but each mechanism points to the same outcome, which is improved accuracy after a brief lag.

There is also some reason to expect the opposite outcome. In previous work with our task, interruptions as brief as 2.7 seconds reduced accuracy (Altmann et al., 2014), and 1 second is not that much shorter than 2.7 seconds. Moreover, there is evidence that unpredictable onset of events impairs placekeeping. Using an event-counting task, Carlson and Cassenti (2004) found higher error rates when the timing between event onsets was random than when it was rhythmic, for a given average time between events. In our task, if placekeeping operations are triggered by completion of a step, then an unpredictable lag between that step and the next could increase the chance of an anticipatory error. In this case a successful model would have to spell out the timing and coordination of the underlying control operations in detail.

Finally, our methodological goal is to develop and evaluate a simple method for testing whether a model adequately accounts for effects of experimental manipulations. The method involves fitting the model to the data from each individual participant, to generate a distribution of model-data residuals across participants for each cell of the experimental design. If these distributions cluster around zero in all cells of the design, this would indicate that the model is able to track all the experimental effects. If the distributions differ significantly from zero in at least some cells, this would indicate that the model was unable to track a specific experimental main effect or interaction—the conditions of which should help us identify the underlying theoretical problem. The decision rule for testing model-data residuals comprises a set of F ratios derived in part from the analysis of variance (ANOVA) applied to the empirical data.

In sum, our goals in this study are as follows. Empirically, we would like to investigate the effect of brief lags between trials that could function either as short breaks that help performance or as short interruptions that hinder it. We would also like to examine the role of rehearsal as a strategy for maintaining placekeeping information during interruptions. At a theoretical level, we would like to develop a cognitive model of placekeeping, focusing in this study on explaining effects of brief lags and rehearsal. At a methodological level, we would like to demonstrate a simple procedure for testing model fit and inferentially rejecting models that include incorrect assumptions.

In the remaining sections we present the experiment (Section 2), then describe the model (Section 3), and then describe our goodness-of-fit test and apply it to different model versions (Section 4). In the General Discussion (Section 5) we discuss the external validity of our task, relate our goodness-of-fit test to Bayesian methods, and discuss limitations of our modeling approach. In the Appendix we describe the model mathematics and assumptions in detail.