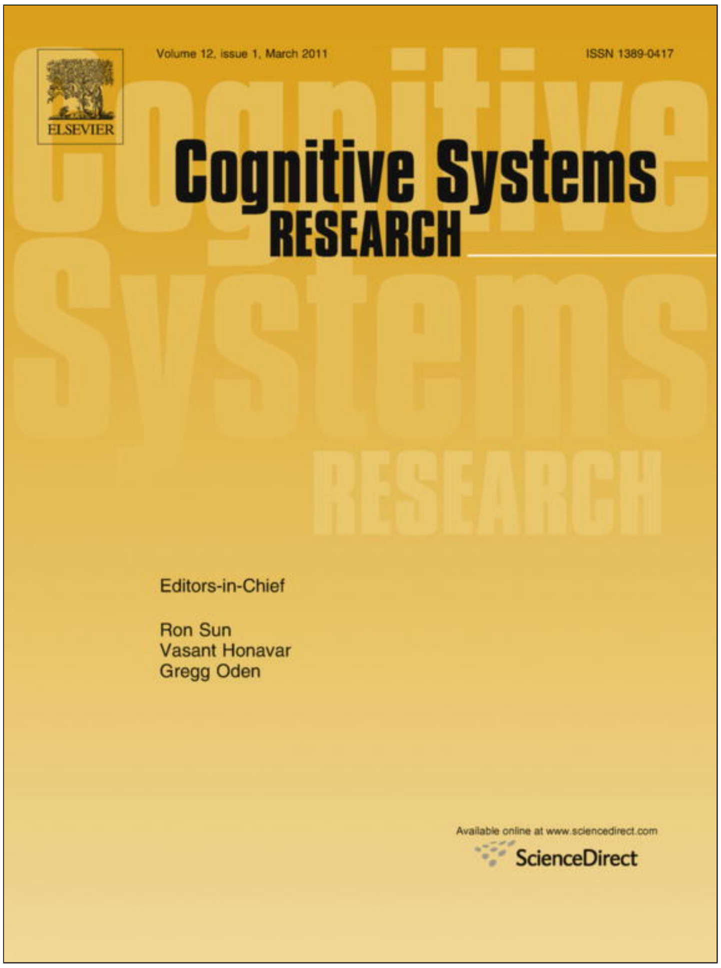
Volume 12, issue 1, March 2011
ISSN 1389-0417
ELSEVIER
Cognitive Systems
RESEARCH
Editors-in-Chief
Ron Sun
Vasant Honavar
Gregg Oden
Available online at www.sciencedirect.com
ScienceDirect

Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 12 (2011) 19–32

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A cognitive modeling account of simultaneous learning and fatigue effects

Action editor: Ron Sun

Cleotilde Gonzalez [a,*], Brad Best [a], Alice F. Healy [b], James A. Kole [b], Lyle E. Bourne Jr. [b]

[a] *Dynamic Decision Making Laboratory, Social and Decision Sciences Department, Carnegie Mellon University, United States*
[b] *Department of Psychology and Neuroscience, University of Colorado at Boulder, 345 UCB Boulder, CO 80309-0345, United States*

Received 20 March 2010; received in revised form 28 June 2010; accepted 30 June 2010
Available online 14 July 2010

Abstract

Current understanding of sources of fatigue and of how fatigue affects performance in prolonged cognitive tasks is limited. We have observed that participants improve in response time but decrease in accuracy after extended repetitive work in a data entry task. We attributed the increase in errors to accumulating fatigue and the reduction in response time to learning. The concurrent effects of fatigue and learning seem intuitively reasonable but have not been explained computationally. This paper presents a cognitive computational model of these effects. The model, developed using the ACT-R cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998), accounts for learning and fatigue effects through a time-dependent modification of architectural parameters. The model is tested against human data from two independent experiments. Best fit to human accuracy and total response time was found from a modulation of both cognitive and arousal processes. Implications for training and skill acquisition research are discussed.

Keywords: Training; Fatigue; ACT-R; Computational model; Prolonged work; Learning

1. A cognitive modeling account of simultaneous learning and fatigue effects

Fatigue is a complex concept that overlaps multiple areas of science, from physiology and sports medicine to psychology and therapy. Although fatigue is usually associated with decrements in performance, the construct of fatigue is imprecise, not well defined, and might have effects beyond simple performance decrements. There are at least three types of fatigue that are active areas of scientific research: sleep deprivation that produces a disruption in the normal circadian rhythms (Gawron, French, & Funke, 2001; Gunzelmann et al., 2007), physical fatigue that results from increased time spent performing physical work (Gawron et al., 2001), and mental fatigue that results in a reduction of the capacity to perform an activity as a result of extended time spent on mental work (Bartlett, 1953). The type of fatigue that we address in this paper is mental fatigue (hereafter, *fatigue*). Although a real-world situation could involve both physical and mental activities, Healy, Kole, Buck-Gengler, and Bourne (2004) have concluded that mental fatigue, not physical fatigue, is often responsible for the increase in errors in repetitive tasks.

On one hand, fatigue effects might be attributed to limitations on cognitive processes such as attention. For example, some models assert that cognitive resources are needed during task performance and that there is a limited amount of such resources to expend in the task (Wickens, 1984). Thus, monotonous and prolonged perceptual processing depletes this pool of resources, making it hard to maintain attention (Parasuraman, 1986), often resulting in habituation (Mackworth, 1969). On the other hand, fatigue effects might be explained with arousal theories that argue that

* Corresponding author. Address: Carnegie Mellon University, 5000 Forbes Ave. Porter-Hall 223-C, Pittsburgh, PA 15217, United States. Tel.: +1 412 268 6242; fax: +1 412 268 6938.
E-mail address: conzalez@andrew.cmu.edu (C. Gonzalez).

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2010.06.004

performance decrements are due to the lack of stimulation needed to maintain alertness (Ballard, 1996). Often, sustained repetitive tasks are boring (Hoffman, Sherrick, & Warm, 1998), producing decreases in arousal resulting in performance decrements (Mackworth, 1969). There is clearly a need to better understand these two possible explanations: fatigue as a cognitive process (e.g., attention) and fatigue as an arousal process.

Computational cognitive models can help formalize and integrate the factors that result in human fatigue, as well as serve as prediction performance tools. A key difference between the models developed by others and the one we present in this paper is that the present model derived directly from a behavioral pattern we have observed in a number of experiments incorporating extended task performance, which resulted in both beneficial and deleterious effects on performance (Healy et al., 2004; Kole, Healy, & Bourne, 2008). Beneficial effects, demonstrated as a decrease in response latency over time, result from general skill acquisition and from specific learning or repetition priming attributable to the repeated occurrence of stimuli and responses. Deleterious effects, demonstrated as an increase in errors over time, have causes that are less clear, but might be attributed to *fatigue* or fatigue-like processes such as boredom, task disengagement, or loss of attention that builds across trials. Although most of the cognitive models would be able to (in different ways) explain the increase in errors over time, none of the existing models of fatigue seems able to account for the concurrent and gradual effects that extended task practice has on learning and fatigue.

We propose that mental fatigue is generated by the integration and tradeoff of cognitive and arousal processes. In this research we aim to provide a concrete definition of these processes and their modulation of skill acquisition in repetitive tasks. Specifically, we propose a construct of fatigue developed and tested through an empirically-based cognitive modeling approach. We make use of mechanisms defined in the ACT-R cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998) to reproduce the simultaneous speedup and increase in errors observed in a data entry task (Healy et al., 2004; Kole et al., 2008). One could present an explanation of the concurrent learning and fatigue effects as the result of strategic control, where a human decides to adopt diverse strategies to place his or her own performance at different gains of speed at the cost of reducing accuracy (and vice-versus) (Pachella, 1974). However, rather than a strategic shift reflecting a speed-accuracy tradeoff adopted by the participants, we view the change in speed and the change in accuracy as reflecting a modulation of fatigue on the learning process. We will demonstrate that the concurrent learning and fatigue effects are the result of permanent learning critically present in extended repetitive tasks: the result of a more or less irreversible modification of the execution process, modulated by the level of arousal.

Task speedup with extended practice has been studied extensively in ACT-R models of learning and skill acquisition (e.g., Anderson, Fincham, & Douglass, 1999; Ritter & Schooler, 2001; Taatgen & Lee, 2003). The increase in errors that results from prolonged work, however, is rarely explained using cognitive modeling (Gunzelmann et al., 2007), and the simultaneous speedup and increase of errors with extended practice has not been explained with cognitive models at all. Thus, in this paper our theoretical account is that the concurrent speedup and increase in errors together arise from a combination of fatigue-induced changes in speed of learning and in processing capacity. These two mechanisms trade off and ultimately make people both faster and less accurate. A model of fatigue developed in ACT-R will be used to predict human data across two behavioral data sets. We will demonstrate how ACT-R can account for the effects of fatigue through time-varying changes to a subset of some central parameter values. Using the ACT-R mechanisms, we will demonstrate that fatigue results from the manipulation of motivation or level of arousal, which modulates cognitive functions, such as the maintenance and prioritization of goals and attention to relevant versus irrelevant information. Finally, we will discuss implications of the model for research on skill acquisition and training in general.

2. Cognitive models of fatigue

Computational cognitive models of fatigue are not common. Initial work in this area was done by Jongman (1998). She proposed an ACT-R model of mental fatigue in the Sternberg (1966) memory search task that consisted of performance-related decrements to cognitive control (in this particular case defined as the inhibition of interfering processes and stimuli, W parameter) and to different levels of affective control (defined as motivation, G parameter). The model worked on two different strategies that were advantageous for both accuracy and speed. Interestingly from our perspective is that the model predicted an early switch from faster and more accurate responses to slower and less accurate responses as the motivational parameter (G) decreased. The model also predicted slower and less accurate responses as the cognitive control parameter (W) decreased. Thus, Jongman's ACT-R model does not predict the simultaneous effects of extended task practice we found in behavioral studies of data entry (Healy et al., 2004). Rather, the model predicts that a decrease in either W or G would result in deterioration in both response time and accuracy. The simultaneous effects of W and G relate of course to the nature of the task being modeled. In the Sternberg task used by Jongman (1998) there are limited opportunities for learning, and the interaction of learning and fatigue could not be examined as well as in the extended repetitive task that we use in the present research (from Healy et al. (2004)).

More recently, considerable work on computational models of fatigue has been reported by Gunzelmann and colleagues (Gunzelmann, Byrne, Gluck, & Moore, 2009; Gunzelmann & Gluck, 2009; Gunzelmann, Gluck, Van

Dongen, O'Connor, & Dinges, 2005; Gunzelmann et al., 2007; Gunzelmann, Gross, Gluck, & Dinges, 2009). Although their emphasis is on modeling fatigue effects as a result of sleep deprivation rather than fatigue from time on task, their use of ACT-R's architectural parameters to modulate performance highlights some important relationships with the model we propose in this paper. A common way to capture fatigue effects in these efforts has involved a decrement in a parameter associated with arousal in ACT-R (the *G* parameter). This parameter, which is explained in detail later, has been associated with arousal and motivation in ACT-R (e.g., Belavkin, 2001). Interestingly, in addition to decreased arousal, the work of Gunzelmann and colleagues also explored mechanisms that account for the individuals' compensation for that decreased arousal in a psychomotor vigilance task (Gunzelmann et al., 2009).

An example of an alternative mechanism for modeling fatigue is that proposed by Gunzelmann et al. (2007). They modeled sleep deprivation effects by directly decrementing activation of declarative knowledge as time on task increased. This mechanism accounted for an increase in errors and an increase in response time, which is the behavioral pattern they observed during sleep deprivation. As per ACT-R mechanisms, decreased activation produces more omission errors and directly relates to a slowing down in fact retrieval (Anderson et al., 2004). Thus, the use of this mechanism of directly decrementing activation of declarative knowledge alone would not account for both the speedup in response time and the increase in errors that we want to capture in our model.

Other alternative mechanisms in ACT-R that may potentially capture the effects of prolonged work presented here come from our own initial work on fatigue modeling (Fu, Gonzalez, Healy, Kole, & Bourne, 2006; Gonzalez, Fu, Healy, Kole, & Bourne, 2006). In this work we increased activation noise within ACT-R combined with the motivational parameter, *G*. Although we still believe this is a reasonable way to reproduce the beneficial and deleterious effects of fatigue, we also think that the *W* parameter has stronger support from the literature on cognitive control and resource models to explain the speedup process. Also, *W* has a clearer correspondence to cognitive variables involved in prolonged work, compared to activation noise. For example, performance decrements resulting from prolonged, repetitive work can be explained in terms of reduced attention and other cognitive resources (Smit, Eling, & Coenen, 2004), but they are rarely explained as an increase in noise. Also, the *W* parameter has been used to model key aspects of language processing (Lewis, 2006), the effects of individual differences in working memory (Lovett, Reder, & Lebiere, 1999), and individual differences in a memory search task (Chuderski, Stettner, & Orzechowski, 2006). The direct correspondence of noise to cognitive constructs is less clear.

Following previous work, we propose an ACT-R model of mental fatigue in which both, arousal (*G*) and cognitive factors such as attention control (*W*), influence performance, as in Jongman (1998). However, in our model, the decreased attention control and arousal (*G*) combined with a *production compilation* mechanism (explained later) that learns new production rules on the basis of existing ones, produces the simultaneous speedup and increase in errors. We demonstrate the effects of both *G* and *W* in our model and present predictions of both the isolated and combined effects of these parameters in combination with the production compilation mechanism.

In what follows, we will detail our account of the impact of prolonged work on attention and arousal processes within ACT-R. We will demonstrate that the best model-to-human data fits (in two different laboratory experiments) are found when both attention and arousal processes decay with extended time on task. Inferior model-to-human data fits are found when only arousal or only attention processes decay, or when neither attention nor arousal processes decay as work progresses. However we also find that the decrease in arousal (*G*) coupled with production compilation leads to speed up. As *G* decreases, this leads to a faster shift to selection of newly compiled productions.

3. The ACT-R model of fatigue

ACT-R (Anderson & Lebiere, 1998) is a unified computational theory of cognition that has accounted for hundreds of phenomena from diverse areas in cognitive psychology including perception, attention, learning, memory, problem solving, and decision making. ACT-R is an architecture composed of interacting modules for declarative memory, perceptual systems such as vision and audition, and motor systems such as manual movement, all synchronized through a central production system. Although we used ACT-R 6.0, we relied on the utility mechanism from ACT-R 5.0 (available in ACT-R 6.0), because the *G* parameter is an essential component from models of fatigue existent in the literature (Gunzelmann et al., 2009; Jongman, 1998; Fu et al., 2006; Gonzalez et al., 2006). For a discussion of the differences between ACT-R 5.0 and 6.0 and how they would impact our model, please see Appendix A.

ACT-R combines a symbolic level with a sub-symbolic system. The symbolic level is implemented as a production system that enables the specification of complex cognitive functions through productions (if–then rules, or procedural knowledge) and chunks (declarative knowledge). The sub-symbolic level involves a set of mathematical procedures that help tune the system to the statistical structure of the environment. The combination of these aspects provides both the broad structure of cognitive processes and the graded characteristics of cognition such as adaptivity, robustness, and stochasticity.

Speedup can be captured in ACT-R in many different ways (Anderson et al., 1999; Ritter & Schooler, 2001; Taatgen & Lee, 2003). The mechanism we used is *production compilation* (Taatgen & Lee, 2003), which is a form of procedural learning that can be described as combining multiple productions (or steps) into a single production. This single production has the same effect as its components but is more efficient. This approach has been used successfully to model skill acquisition in diverse areas including use of the English past tense (Taatgen & Anderson, 2002), problem solving in the balance-beam task (van Rijn, van Sumeren, & van der Mass, 2003), and air traffic control (Taatgen & Lee, 2003).

We show that prolonged work effects are captured by the combination of two ACT-R sub-symbolic parameters, G and W, that map onto motivational fatigue in the form of lack of arousal (the G parameter) and mental fatigue in the form of lack of attention control (the W parameter) in combination with the production compilation mechanism of ACT-R.

The W parameter comes from the following Activation Equation defined in the ACT-R architecture (Anderson & Lebiere, 1998):*Activation Equation*:

$$A_i = B_i + \sum_j W_j S_{ji} - D_i + \varepsilon \tag{1}$$

According to the ACT-R theory, the activation A_i of a chunk i reflects the estimate of how likely the chunk would match to a production at the current point of time as well as the speed of retrieval of that chunk (Anderson & Lebiere, 1998; Anderson, Reder, & Lebiere, 1996). The activation of a chunk is determined by the base-level activation B_i, the associative activation S_{ji}, the mismatch penalty value D_i and noise ε. The base-level activation B_i of the chunk i reflects the recency and frequency of use of the chunk. S_{ji} reflects the impact of contextual values on the chunk activation, and D_i reflects the degree to which the chunk matches a context (i.e., the extent to which a given chunk is dissimilar to previously presented chunks in each chunk position). The parameter ε is a variable noise value associated with a chunk.

S_{ji} contributes to the sum of the source activation that a chunk receives from the elements currently in the focus of attention. W_j represents the attention weightings of each of the element's j cues that are part of the current goal chunk, and the S_{ji} component represents the strengths of association that measures how often the chunk i is needed when cue j is an element of the goal. ACT-R assumes that there is a limited total amount of attention (W, the sum of all W_j) that one can distribute over source objects (Anderson et al., 1996, 2004). W is an ACT-R parameter that reflects the salience or attention given to the cues of a chunk. This salience helps create contrast between relevant and irrelevant information for the current goal, aiding in the maintenance of information necessary for task performance. Thus, W influences the attention to relevant and irrelevant information (Anderson et al., 1996; Lovett et al., 1999). Higher values of W facilitate the retrieval process by increasing spreading activation, whereas lower values of W reduce activation, increasing the likelihood that incorrect items will be retrieved from memory. We suggest that variations to ACT-R's W parameter represent the cognitive component of fatigue. With prolonged work, a decrease in W would produce attention reduction to relevant stimulus information, impacting negatively the recall of information and thus, producing more errors.

In addition to W, we will show that fatigue effects are influenced by motivational factors, represented in ACT-R by the G parameter. The G parameter comes from the Utility Equation (Anderson et al., 2004; Anderson & Lebiere, 1998):*Utility of a production*:

$$U_i = P_i G - C_i + noise \tag{2}$$

In contrast to the W parameter that impacts memory retrieval, the G parameter has its main impact on the production cycle. The utility (U_i) of a production i is the product of the probability of a production's success (P_i) and the value of the goal (G), minus the expected cost of executing the production (C_i), plus noise. P_i is an estimate of the probability that if a production i is chosen the current goal will be achieved, and C_i is an estimate of the cost (measured in seconds) to achieve that goal. Both, P_i and C_i are learned statistically with the execution of production i using a Bayesian mechanism of the ratio of successes to the sum of successes and failures started with some prior value (Anderson et al., 2004). The estimated value of P is simply the ratio of successes to the sum of successes and failures, where successes and failures occur when the production has succeeded or failed to accomplish the current goal, respectively. G is a global parameter that represents the value of the current goal and produces dynamic effects in the functioning of ACT-R. The properties of the G parameter have been studied by Belavkin (1999) and interpreted as a "motivation" parameter. Higher values of G result in higher dependence on the expected probability of success (P) and less on the costs (C). That is, the system pays less attention to the cost and relies more on the expected probability of success. In contrast, lower values of G result in lower utility of a production because the system puts little effort into the task regardless of the probability of a successful outcome.

Because ACT-R specifies that only one production can fire at a time, one production rule must be selected from all the ones that match the conditions for execution (a process called conflict resolution). The probability of selecting one production rule out of j possible ones is defined by the following equation (Anderson & Lebiere, 1998):*Probability of production selection*:

$$\mathrm{Prob}(production_i) = \frac{e^{\frac{U_i}{t}}}{\sum_j e^{\frac{U_j}{t}}} \tag{3}$$

where t is a noise parameter and the summation is over the j alternative productions. As expected, the utility of a production influences the probability of its being selected during the conflict resolution process.

We suggest that ACT-R's G parameter represents the arousal component of fatigue. A decrease in G would reflect decreased motivation, lower effort, and a decline in the maintenance of goals (Belavkin, 1999). The G parameter has also captured decreases in arousal that lead to decreased cognitive performance (see Ritter, Reifers, Klein, Quigley, & Schoelles, 2004, for a discussion of the impact of moderating arousal on cognitive performance), and it has been used to account for effects of emotion in several tasks (Belavkin, 2001; Jongman, 1998). In the work of Gunzelmann et al. (2009), decreases in G produced situations where production utilities interacted with the utility threshold, the minimum utility a production must have to fire, and thereby produced lapses in behavior. This interaction was necessary to capture the effects of extreme sleep deprivation, where participants sitting at a computer screen after several days without sleep simply failed to respond to a presented stimulus (as if it never appeared). In our model, reductions in G in combination with the production compilation mechanism produce the speedup with increased time on task, since it accelerates the transition to newly compiled productions.

In summary, in its current state, ACT-R is equipped with mechanisms that can be used to explore the integration of arousal and attention states of fatigue in prolonged task performance.

4. ACT-R model of fatigue in the data entry task

In this section, we explain how the arousal and attention ACT-R architectural parameters (G and W, respectively) were manipulated to obtain the fatigue effects found in two published experiments. We first introduce the task and then describe the ACT-R model implementation for the task.

4.1. The data entry task

Data entry, a discrete, repetitive, and long-duration task, has been used for various purposes in empirical and modeling research (Gonzalez et al., 2006; Healy et al., 2004; Kole et al., 2008). In this task, participants type digits shown on a display screen using a computer keyboard (key row or keypad). In the experiments used as the basis for our models, four-digit numbers were shown in the center of a computer display and remained there until the participant typed the *Enter* key. Participants used the keypad to the right of the letter keys to enter the digits. Participants were instructed to type the number as quickly and accurately as possible. Participants' responses were not displayed and no supplemental feedback was provided. We selected this data entry task for four key reasons: First, the task has partially separable cognitive and motoric components, allowing us to investigate the cognitive processes involved somewhat independently from the motoric processes. Second, the task is such that even the most motivated participant might be strained by repetition and prolonged work, allowing us to investigate the arousal processes involved. Data entry could be described as being non-exciting without generating much debate, a feature that is extremely valuable to the study of the motivational aspects of fatigue. Third, the task shows interesting, opposing effects of prolonged work on response time and error rate (Healy et al., 2004), a challenge that current cognitive models of fatigue cannot explain (e.g., Jongman, 1998). Finally, data entry is a very common task in human life. Data entry is performed at every cash register, every telephone keypad, and every computer, by a variety of people with a wide range of abilities. Thus, the results from this research are potentially widely applicable to these situations.

4.2. ACT-R fatigue model in data entry

The declarative representations for the data entry task in ACT-R involved two main chunk structures (representations of declarative knowledge). The first structure represents four-digit numbers, where each of the digits in a chunk serves as a cue. Typing the four-digit number is the main goal in the model, involving either all the four digits or two digits at a time depending on the strategy used, as will be explained later. The second chunk structure in the model represents the keypad, with the number keys 1–9 and the *Enter* key being the cues of the chunk.

The general ACT-R productions used in the model of the data entry task are shown in Fig. 1. This model includes three main steps: (a) encoding of the stimulus on the computer screen as well as preparing the required response, (b) entry of the encoded stimulus using the keypad, and (c) entry of the "Enter" key. The first step further unpacks to include reading of individual digits as well as preparing the proper motor program to press the desired keys (e.g., retrieving key locations), whereas the second and third steps include only motor execution.

Although the task is quite simple, it still requires maintenance of encoded stimuli in working memory between encoding and typing, potentially decomposing a task into subgoals (e.g., entering a digit at a time), and the interaction with skilled actions (keyboard entry), which is simu-

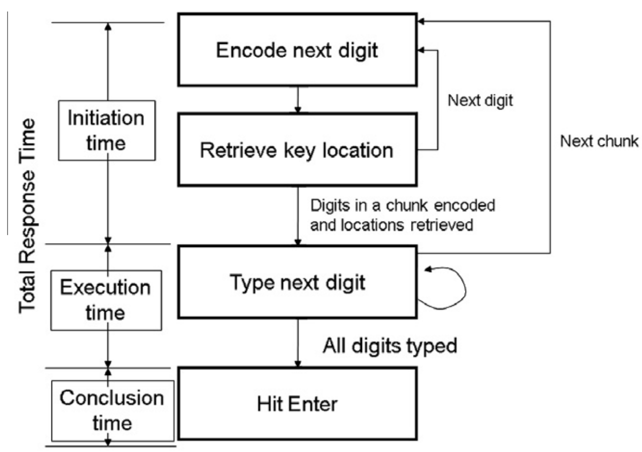

Fig. 1. Steps in the ACT-R cognitive model of the data entry task.

lated through the application of individual ACT-R productions (e.g., typing the nine key on the keypad).

There are two major component processing stages in the data entry task: (a) encoding plus response preparation and (b) response execution. Encoding involves perceptual processes, response preparation involves the mental construction of a program or plan for entering the sequence, and response execution involves the actual motoric button presses. It is assumed that both encoding and response preparation for digit sequences or chunks are completed prior to the execution of the first keystroke of the sequence or chunk (i.e., encoding of chunks and response preparation are not interleaved with execution). Thus, the time to complete the first keystroke in a sequence or chunk reflects both stages, whereas the time to complete the remaining keystrokes reflects primarily the response-execution stage (e.g., Proctor & Dutta, 1995). Research involving motor and cognitive operations at the millisecond level, suggests that motor and cognitive operations overlap. That is, to account for human performance at the millisecond level one needs to assume that perceptual and motoric processes are interleaved rather than executed in sequence (Veksler, Gray, & Schoelles, 2007). We will show that our simplifying assumption of response preparation before execution of a chunk is confirmed by the model and human data (from Healy et al. (2004)), and thus encoding and execution of keypresses are performed separately in the data entry task.

The *production compilation* mechanism predicts that individual steps of cognition will be combined into more efficient steps when possible. For example, given the abstract productions $A \rightarrow B$ and $B \rightarrow C$, the combined production $A \rightarrow C$ achieves the same outcome with fewer steps and in less time. In the case of the data entry task, the production representing the search for the key on the keypad is combined together with the production representing the step of actually typing a key, reducing response time. The production compilation mechanism initially creates this potentially more efficient production with some uncertainty of its actual utility (see Eq. (2)); the actual utility of the new production is learned over time in a competitive learning environment including the original (less efficient) productions, producing a gradual speedup as the transition to the new production occurs. This compilation uses the utility of the production as described earlier (see Eq. (2)), where the value of G will influence how soon the transition to the new, more efficient production will occur. As G decreases, the utility of productions, given by the utility equation (Eq. (2)), becomes more heavily dependent on the cost parameter C than on the production success probability parameter P. This change leads the system to more quickly adopt new productions that have a higher risk (low P) and lower time cost (low C), resulting in a faster transition to newly compiled productions than would otherwise be achieved. Thus, decreasing G produces speedup relative to a model with a constant G parameter.

According to the Activation Equation (Eq. (1)), as W decreases, the target chunk (i.e., the actual digits that were encoded) receives less and less of an activation boost when a retrieval attempt is made. This decrease makes it more likely that an incorrect item will be retrieved from memory, and provides for errors of commission in entering the key sequence (with the most likely mistake being a repetition of a previously keyed and highly active chunk, or yet-to-be keyed and highly active chunk, but with other intrusions also being possible).

To obtain fatigue effects observed in the two experiments to be considered here, G and W were manipulated to decrease gradually and in real time over the time in the task. This degradation of G and W from their initial values takes place in very small amounts, on every cycle of the ACT-R production system. The decrement was implemented in a function that is called automatically when productions fire. The G degradation occurs at a rate of 0.009 per s, updated in every cycle of task performance. The W degradation occurs at a rate of 0.000175 per s. These specific values of G and W degradation were obtained through a calibration process done by comparing the data produced by the model to the human data and using goodness-of-fit measures (r and RMSE) to determine the accuracy of the model. The calibration process was done with data from the first experiment reported in Section 5. Then, the same model was used to predict data from the second experiment. The comparison process to human experiments is explained in Section 5.

The initial value of G is the default value for the architecture (20.0), but the initial value of W depends on the strategies implemented in the model. Based on observations of human participants (Buck-Gengler, Raymond, Healy, & Bourne, 2007; Raymond, Buck-Gengler, Healy, & Bourne, 2007; Raymond, Fornberg, Buck-Gengler, Healy, & Bourne, 2008), we implemented two strategies in this model (which were used equally often among the participants): one strategy involved encoding and entering pairs of digits one pair at a time (two-at-a-time strategy), and the other strategy involved encoding all four digits before entering them all at a time (four-at-a-time strategy). Given these two strategies (entering digits as either one string of four numbers or two pairs), accuracy within the model is strongly impacted by the W parameter, which weights the attention given to each element or chunk, in this case corresponding to a pair or sequence of four digits: The same value of W produces very different performance when coupled with the two different strategies. This outcome occurs because the elements in the focus of attention (i.e., the target chunk) are different (two or four digits at a time) depending on the strategy. Thus, because W is distributed among two source objects in the case of the two-at-a-time strategy (W is distributed into two W_j) or among four source objects in the case of the four-at-a-time strategy, the two-at-a-time strategy would result in higher attention weightings (fewer errors) compared to the four-at-a-time strategy.

The particular effect of W in our model depended on both the initial value of W and the two different strategies used. As shown in ACT-R modeling work of individual differences, higher values of W are associated with better working memory (e.g., Lovett et al., 1999). Thus, the use of the two- or four-at-a-time strategies would depend on the level of W. The default ACT-R value of W is 1.0, and to reproduce the patterns of accuracy demonstrated by human data we used two initial values of W: 1.2 for the four-at-a-time strategy and 0.7 for the two-at-a-time strategy for an average W of 0.95. In both experiments reported next, half of the simulated participants were run with each strategy because about half of the experimental participants demonstrated each strategy, although none of the groups seem to have chosen one strategy exclusively (Raymond et al., 2007, 2008). The choice of the strategy occurred at the start of the learning process. Simulated participants were equally split between the two strategies and, hence, the two W values.

In the next sections, we will show how this model was calibrated to the human data from the first experiment. Then the same model was used to make predictions of human data from another experiment. We will also demonstrate the results from an independent manipulation of G and W in a model experiment.

5. Accounting for human experimental data

We examined two laboratory experiments using the data entry task reported in Healy et al. (2004). In these experiments, participants were presented with a list of four-digit numbers and were asked to type in the digits using the keypad on the keyboard. The experiments varied specific aspects of the task. The common structure involved two session halves, each half using a series of five blocks of 64 numbers with a break between session halves. The experiments presented participants with a practice block of four trials before proceeding with the task itself. Experimental conditions used numerals as stimuli (e.g., "4 6 1 9"). The factors varied across the experiments included the following:

Repetitions: In Experiment 1, items were presented multiple times during the experiment to investigate practice and repetition priming effects. Specifically, a list of 64 numbers was used five times (permuted) in the first half of the experiment (each of Blocks 1–5), and then a new list of 64 numbers was used five times in the second half. In Experiment 2, there was no repetition of stimulus items and all the numbers used in the second half were different from those used in the first half of the experiment.

Dominant hand use: Participants in Experiment 1 used their non-dominant hand (i.e., left-hand for all participants, none of whom reported to be left-handed) in both session halves. In Experiment 2, half of the participants (again, all of whom were right-handed) used their dominant hand for the first session half, half started with their non-dominant hand, and then half of each of these groups switched hands for the second session half.

The ACT-R cognitive model was run under the same conditions as the laboratory experiments. We collected the number of simulated participants from the model that corresponded to the number of human participants, so that we could compare the two sets of data on common ground. Then, we compared the model predictions to human behavior for RT and proportion correct. To measure the model's ability to reproduce human results, we calculated correlations and the RMSEs between the model output and the experimental data. The number of data points used for these statistics corresponded to the averages per block for each of the conditions (i.e., 10 data points for each measure).

5.1. Healy et al. (2004, Experiment 1)

This study explored the impact on data entry speed and accuracy of the repetition of study items. Following the practice block, the same list of 64 four-digit numbers was used in five different permutations in the first session half, once per block, and then a second list of 64 four-digit numbers was used in five different permutations in the second session half, again with 64 trials per block. A break was provided between the fifth and sixth block, which was varied in length, but because no important effects of the variable length factor were detected, the results reported here for Healy et al. (2004) were collapsed across these levels.

The dependent variables included both total response time (TRT) and accuracy measures. Only correct trials were included in the TRT analyses. All four digits of a number had to be entered in the correct order, with no extra digits typed, for the trial to be scored as correct. TRT was the total time to enter all four digits plus the concluding *Enter* keystroke.

The results comparing model predictions to human behavior for average TRT and proportion of correct responses are summarized in Fig. 2. The primary observation by Healy et al. (2004) is that prolonged work resulted in both learning and fatigue effects, with learning effects dominating in the speed measure and fatigue effects dominating in the accuracy measure. With continued repetition across blocks, the proportion of correct responses decreased over time. On one hand, this result suggested deterioration in performance as a result of fatigue. On the other hand, with continued repetition, participants completed correct trials more and more quickly, suggesting improvement in performance as the result of learning.

The ACT-R model as described in the previous sections produces these same main effects: learning reflected in decreasing response time, but fatigue reflected in decreasing accuracy with prolonged work. The speedup resulted from the *production compilation* mechanism that combined the search for the key together with the step of actually typing a key, reducing the response time. The predicted pattern for TRT produced an r of 0.96 and an RMSE of 0.03 when compared to human results. Similarly, the predicted pattern for accuracy (proportion of correct responses)

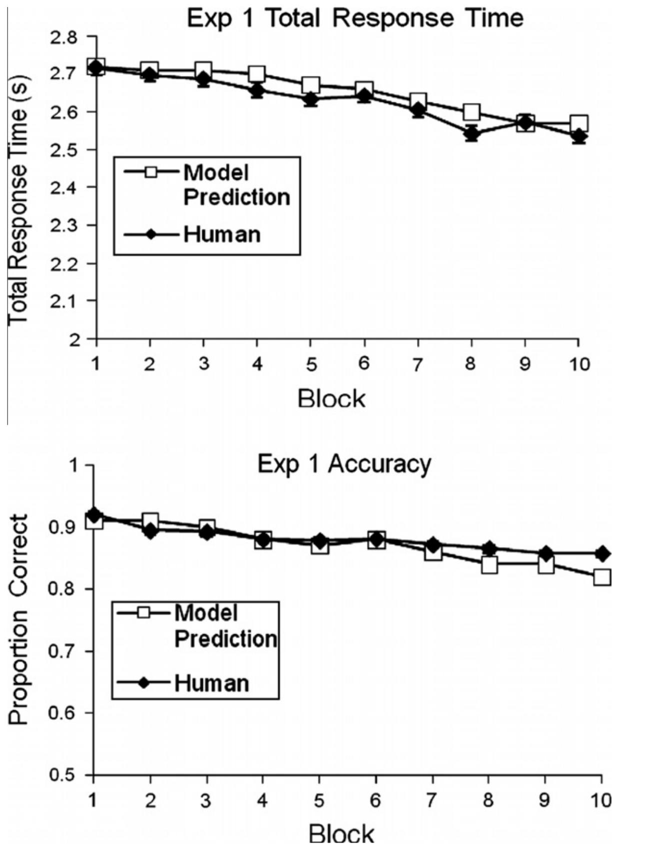

Fig. 2. Model predictions and human behavior comparisons for total response time (top panel) and proportion of correct responses (bottom panel). The number of simulated participants is the same as the number of human participants ($n = 32$). Human data come from Experiment 1 in Healy et al. (2004). Error bars represent standard errors of the mean.

produced an r of 0.92 and an RMSE of 0.02 when compared to human results. The human TRT data, when split by even and odd participants and the two groups are compared to each other, produce an r of 0.99 and an RMSE of 0.04 for TRT and an r of 0.79 and an RMSE of 0.03 for accuracy. Thus, the model fits the data close to the way the data fit themselves in terms of TRT, and perhaps over-fits in terms of accuracy (i.e., some of the variance accounted for may be noise in the data), although only half of the participants ($n/2$) were used in the fits of the data to themselves.

The deleterious effect of prolonged work in the model was produced through changes in G and W during the course of performance. As explained earlier, we implemented a simultaneous decrease of G and W. Decreasing the G parameter, which can be interpreted as decreasing the amount of time worthwhile to pursue a particular goal, makes the relative cost of component actions (i.e., the time to complete steps) more important in the conflict resolution process (see Section 1 for more detail). Thus, as G is decreased, the model biases its efforts towards shortcuts produced through ACT-R's compilation mechanism that speedup task performance, thereby reducing the total RT as shown in Fig. 2.

In the current study the G parameter appeared to be insufficient on its own to produce the exhibited pattern of human behavior (this issue will be discussed in detail later). In particular, the decrease in accuracy exhibited by human participants is atypical of the empirical body of skill-learning studies, which ACT-R attempts to capture. Rather than an overall increase in accuracy over time, the current participants demonstrated maximal accuracy in the initial block, followed by a steady decrease throughout task performance. A corresponding decrease in the W parameter achieves exactly this effect through increasing the confusions in memory for relevant (target digits) and irrelevant (other digits not in the four-digit number) elements.

The assumption that encoding occurs separately from the motoric processes led to the prediction that the first keystroke would consume most of the total time, whereas the rest of the keystroke times would be relatively short and roughly, but not exactly, equal to each other. Note, however, that the third keystroke would take longer than the second and fourth when the two-at-a-time strategy is used. Fig. 3 shows the response time per digit. As predicted, Digit 1 consumes most of the TRT, followed by Digit 3, and then by the Enter key, Digit 2, and Digit 4. The model predicts these differences in response time per key and compares well to human results (Digit 1 $r = 0.90$, RMSE = 0.11; Digit 2 $r = 0.80$, RMSE = 0.06; Digit 3 $r = 0.50$, RMSE = 0.03; Digit 4 $r = 0.95$, RMSE = 0.08; Enter key $r = 0.92$, RMSE = 0.09). Furthermore, as the model (and humans) learn, the response time of Digit 1 remains slower than the response time of the other digits, again supporting the assumptions of the model.

5.2. Healy et al. (2004, Experiment 2)

A possibility resulting from Experiment 1 discussed in Healy et al. (2004) is that the decrease in accuracy is produced by the motor activity of typing the digits and the

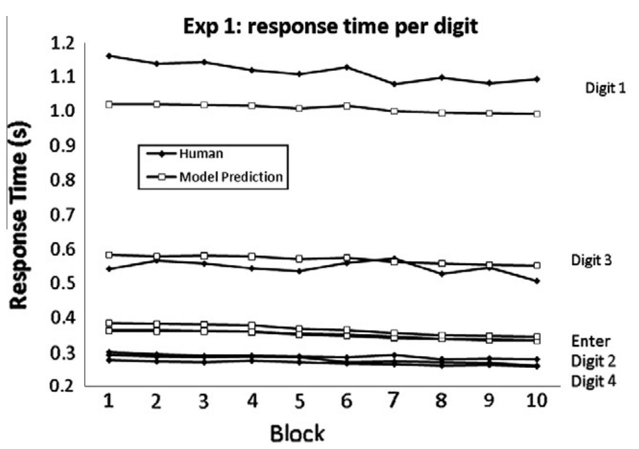

Fig. 3. Model predictions and human behavior comparisons for response time per digit. The number of simulated participants is the same as the number of human participants ($n = 32$). Human data come from Experiment 1 in Healy et al. (2004).

Enter key rather than by a cognitive activity. Healy et al. (2004) followed Experiment 1 with a manipulation of the dominant and non-dominant hands, which for half of the participants switched across session halves separated by a 5-min delay. Half of the participants used their dominant hand for the first session half, half started with their non-dominant hand, and then half of each of these groups switched hands for the second session half. In addition, this study eliminated repetition of numbers across blocks and used an entirely new set of numbers on each block. This manipulation was intended to rule out the possibility that decreases in TRT were due to prior practice with the stimulus items. Using stimuli that did not repeat allowed for the isolation of fatigue and general learning effects from individual stimulus learning effects.

Healy et al. (2004, Experiment 2) replicated the primary observations from the prior Section 5.1, that prolonged work produced both speedup and increase in errors. As with the prior experiment, with continued data entry across blocks, the error rate increased while TRT decreased. However, the error proportion in this study was somewhat different, ranging from approximately 11% in initial blocks to approximately 13% in later blocks, whereas the initial study was characterized by a broader range (approximately 8–15% across the experiment). The fact that error proportion increased across blocks in Experiment 2 even when there was a change in the hand used for typing convinced Healy et al. (2004) that the deleterious effects of prolonged work were mostly due to cognitive rather than motoric factors, and thus, the ACT-R modeling focused on the cognitive factors only.

The same ACT-R model of fatigue that was developed and calibrated for Experiment 1 was used for this task, but with varied stimuli corresponding to this experiment, and with modified actions corresponding to dominant and non-dominant hand use of the keypad. In particular, hand dominance was modeled by manipulating costs (i.e., action times) specific to key presses to account for slower motor actions with the non-dominant hand. Because the model for Experiment 1 estimated the use of the non-dominant hand, this model required a modification of the default time it takes to fire a production in ACT-R. In the first model this action time was set at 155 ms for the non-dominant hand productions of the key presses (the default action time of a production in ACT-R is 50 ms). In this model this action time was set instead at 100 ms for the dominant hand productions of the key presses. The corresponding decrease in response time of 55 ms per keystroke was estimated by dividing the difference in TRTs between dominant and non-dominant hand condition by the number of keystrokes. All other parameters retained their estimates from the previous study.

Results, summarized in Fig. 4, demonstrate that the ACT-R model predicts the main effects found in both Experiments 1 and 2: learning reflected in decreasing response time, but fatigue reflected in decreasing accuracy with prolonged work. As discussed earlier, this fatigue effect (reduction of accuracy with prolonged work), was largely produced through reductions in motivation (modeled by decreasing the *G* parameter) and reductions in cognitive control (modeled by decreasing the *W* parameter). Reductions of *G* led to a more rapid shift to faster alternative actions that had been produced by the compilation mechanism. Newly compiled actions are, by their nature, untried, and therefore have a lower initial utility. However, these newer actions are also faster than the initial actions. Thus, decreasing *G* produces a more rapid switch to the faster, but riskier, new productions. The reduction in *W* makes memory retrieval and target key selection more error-prone with repeated task performance. As expected from the *G* and *W* manipulations, these parameters need to combine with each other (see Section 6 for an explanation of why the combination of parameters is needed). Thus, a decrease in *G* coupled with a decrease in *W* produces faster response time and lower accuracy with prolonged work.

The predicted TRT produces a gap of about 0.2 s between model and human data, resulting in an *r* of 0.82 and an RMSE of 0.27 when compared to human results.

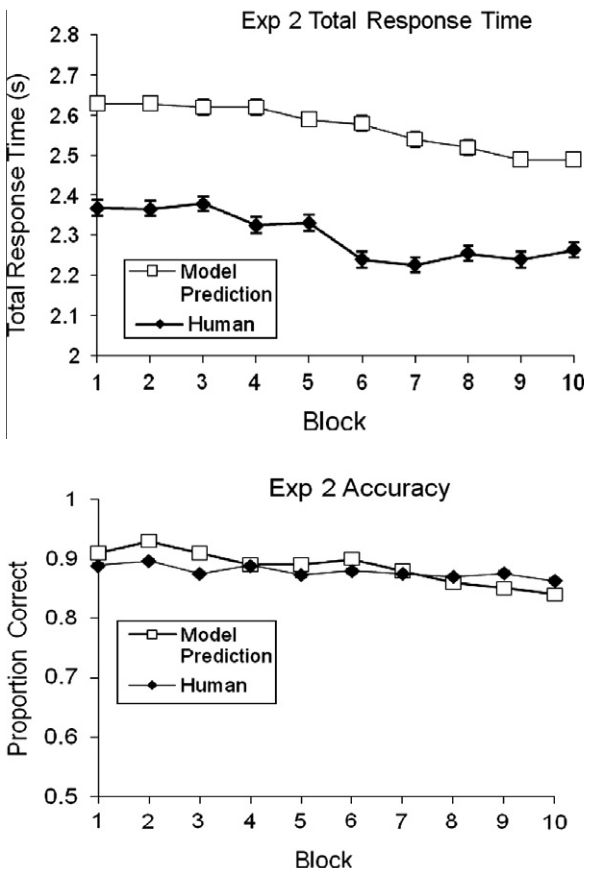

Fig. 4. Model predictions and human behavior comparisons for total response time (top panel) and proportion of correct responses (bottom panel). The number of simulated participants is the same as the number of human participants ($n = 32$). Human data comes from Experiment 2 in Healy et al. (2004). Error bars represent standard errors of the mean.

The predicted pattern for errors produced an r of 0.78 and an RMSE of 0.02 when compared to human. Recall that the same model developed and calibrated with data in Experiment 1, without modification, was used to predict the human data in the new conditions defined in Experiment 2. Thus, the model does a relatively good job when fitting new human data, under different experimental conditions. The human performance data, when split by even and odd participants, produces an r of 0.91 and an RMSE of 0.14 for response time and an r of 0.61 and an RMSE of 0.02 for accuracy. This finding indicates that the model's fit to the human TRT data is worse than that of the fit of one half of the human TRT data to the other half. However, the model fits human accuracy data better than one half of the human accuracy data fit the other half. This finding suggests that large individual differences may be responsible for the worse fits of the model in Experiment 2. However, our goal here is to demonstrate that the original model of Experiment 1 is able to predict the effects of novel human experimental conditions without having to re-fit the conditions of the new experiment.

We expected that different motor action costs would interact with the manipulation to the G and W parameters, and thus the new motor actions added to this experiment (i.e., for switching hands) may be a reason for the worse fit of the model to Experiment 2 compared to Experiment 1. For example, the assumption that encoding occurs separately from the motoric processes might more relevant in this experiment under the assumption that a switch in hands would slow down motoric but not encoding processes. Fig. 5 shows the response time per digit. Again our prediction holds that Digit 1 consumes most of the TRT, with Digit 3 consuming more time than the remaining digits, as shown in the human data (Digit 1 $r = -0.22$, RMSE $= 0.12$; Digit 2 $r = 0.86$, RMSE $= 0.05$; Digit 3 $r = 0.83$, RMSE $= 0.11$; Digit 4 $r = 0.94$, RMSE $= 0.06$; Enter key $r = 0.98$, RMSE $= 0.07$).

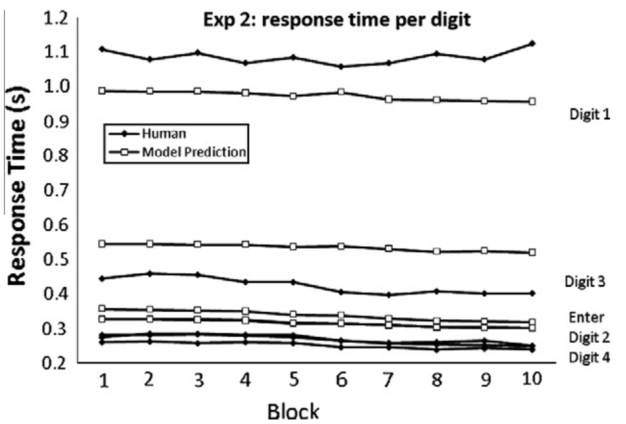

Fig. 5. Model predictions and human behavior comparisons for response time per digit. The number of simulated participants is the same as the number of human participants ($n = 32$). Human data come from Experiment 2 in Healy et al. (2004).

6. A model experiment: predictions of the G and W parameters

The effects presented in the previous section were obtained from a single model and one specific manipulation of parameters: the gradual degradation of both the G and W parameters given repeated trials on the data entry task. The specific values and decline of the G and W parameters were obtained after tuning the model to the human data in Experiment 1 and measuring the closeness of the model-to-human data as explained earlier. The same model was then used to make predictions in Experiment 2, changing only those parts of the model directly related to the new experimental conditions (i.e., hand used and repetition of the numbers). The fact that decreasing both G and W over time yielded the desired learning and fatigue effect suggests that, as we predicted, the integration of both attention and arousal processes together produced the learning and fatigue effects. In this section, we test this proposition, using the computational model to make predictions while manipulating the decreases of G and W orthogonally.

In the simulation we varied the G and W parameters in four conditions: No G decay and no W decay ($\sim G \sim W$); No G decay but W decay ($\sim GW$); G decay but no W decay ($G \sim W$); both G decay and W decay (GW). As described above, G decayed from a value of 20.0, gradually by 0.009 per s (updated on every ACT-R cycle), whereas W decayed from a default value of either 1.2 or 0.7 (depending on the strategy) by 0.000175 per s. When the parameters did not decay they were kept at the same initial value (20.0 for G and 1.2 or 0.7 for W) during all the blocks of each of the experiments.

The model was run for each of the two experiments described in Sections 5.1 and 5.2: Healy et al. (2004) Experiments 1 and 2. That is, the model in each of these two experiments was run four times, one for each of the four conditions ($\sim G \sim W$, $\sim GW$, $G \sim W$, and GW), using the same number of participants as in the original experiments (32 for each of Experiments 1 and 2 of Healy et al. (2004)). The results of these runs were averaged per block, with the two experiments contributing equally, for each of the four conditions. These averages of TRT and accuracy per block for each of the four conditions are shown in Fig. 6. Fig. 6 also shows the overall TRT and accuracy from human data averaged over the two experiments. As can be seen from this figure, the G parameter clearly drives the TRT improvement. The two conditions in which G does not decay (i.e., $\sim G$), regardless of W, present higher TRT over the last six blocks, compared to the conditions with G decrement (i.e., G). Also, it seems clear from the figure that the W parameter drives the decline in accuracy. The two conditions in which W does not decay (i.e., $\sim W$), regardless of G, present a higher proportion of correct responses over the last eight blocks, compared to the conditions with W decrement (i.e., W).

Table 1 summarizes the fit to both human total response time and accuracy for each of the two experiments and for

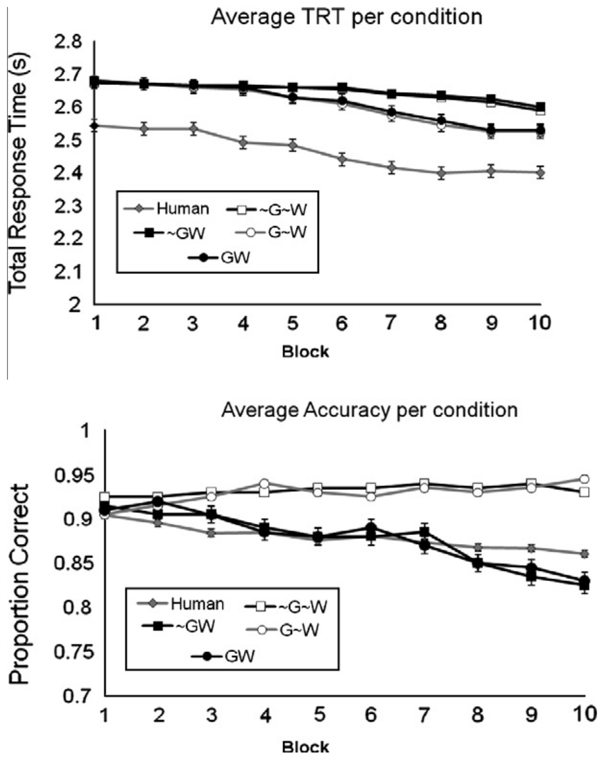

Fig. 6. Model's average total response time (TRT, top panel) and proportion of correct responses (bottom panel) per block, across the two experiments, presented for each of the four conditions of the model experiment. Each of the blocks represents an average of 64 simulated or human participants, obtained by adding the total participants in each of the two experiments from Healy et al. (2004). Error bars represent standard errors of the mean.

each of the four simulated conditions. As illustrated, neither the ACT-R baseline, *W* decay by itself, nor *G* decay by itself captured the TRT and accuracy effects of fatigue as well as did *G* and *W* joint decay.

7. General discussion and conclusions

This research highlights the increasing convergent evidence of the *G* parameter of the ACT-R architecture in accounting for performance degradations associated with fatigue. The initial work by Jongman (1998) as well as the most recent computational cognitive models of fatigue (Gunzelmann, Gluck, Van Dongen, O'Connor, & Dinges, 2005, 2009; Gunzelmann et al., 2007; Gunzelmann & Gluck, 2009) all have suggested the relevant role of arousal, and they have represented the fatigue effects through a decrement in the *G* parameter associated with arousal in ACT-R.

Furthermore, our research uniquely highlights that prolonged work on repetitive tasks results in both skill acquisition and performance decrements (Healy et al., 2004). We explain these concurrent effects by mechanisms that modulate the learning process. Across two studies, two unique forces, arousal and attention processes, integrate to produce the results found in human behavioral data. We used one ACT-R model that varied in two parameters, one mostly associated with arousal processing, the *G* parameter, and the other associated with attention control, the *W* parameter. By decrementing both *G* and *W* combined with the production compilation mechanism in the prolonged task, we explained simultaneous speedup and increase in errors from two previously published experiments. We extend the suggestion from earlier literature that emotional states can influence attention control and vice versus (Easterbrook, 1959; MacLeod & Mathews, 1988) by demonstrating how arousal and attention can have a joint but also an independent impact on RT and accuracy. When *G* and *W* are manipulated orthogonally in a model experiment we found that the condition that meets highest *r* and lowest RMSE values for both response time and accuracy, as compared to human data, is that in which both *G* and *W* parameters decayed jointly over time. However, *G* and *W* clearly have different impact in RT and accuracy: Arousal impacts RT whereas attention impacts accuracy in prolonged work on repetitive tasks.

Importantly, these interesting effects of the *W* and the *G* parameters need to be considered when designing training programs and when designing vigilance and training systems. First, the results from our experiment with the model, manipulating *W* and *G*, suggest that arousal (*G*) is the main driver of response time improvement. This is a counterintuitive prediction, because a reduction in arousal has been shown usually to result in fatigue effects typically having a negative impact on human performance (Gunzelmann et al., 2009). We showed that, when arousal is constant throughout the task, TRT was higher and did not decrease compared to when the arousal parameter decreased. It might seem paradoxical that when people maintain their arousal (e.g., through sustained motivation) they will produce higher and constant rather than lower and decreasing

Table 1
Summary of fit to human data in Healy et al. (2004) for the G and W model experiment.

Condition	Experiment 1				Experiment 2			
	TRT		Errors		TRT		Errors	
	r	RMSE	*r*	RMSE	*r*	RMSE	*r*	RMSE
∼G ∼ W	0.93	0.07	−0.75	0.06	0.70	0.31	−0.40	0.06
∼GW	0.91	0.07	0.91	0.02	0.69	0.31	0.71	0.02
G ∼ W	0.95	0.03	−0.84	0.05	0.83	0.27	−0.59	0.06
GW	0.96	0.03	0.92	0.02	0.82	0.27	0.78	0.02

response times with prolonged work. But in fact, the prediction suggests that when people are motivated and aroused they spend more time in the task. For example, our model shows that a given arousal state may influence willingness to engage in an effortful task. In fact, there is a well-established relationship between arousal level and span of attention (e.g., Easterbrook, 1959; Eysenck, 1982) that supports our proposed interpretation of the simultaneous learning and fatigue effects. Motivation and emotion influence attention selectivity by reducing attention to task-relevant items (Easterbrook, 1959), and there are many other ways in which motivation and emotion influence attention (Eysenck, 1982). The interaction of the production compilation mechanism with the utility learning mechanism allows the cognitive system to adapt to decreasing arousal by speeding up the transition to more efficient productions.

Second, our model experiment suggests that the cognitive part in the theoretical mapping is the main driver of the decline in accuracy. When attentional control decreased over time, there was a decrease in the proportion of correct responses. In contrast, when attentional control stayed constant over time, there was a slight increase in the proportion of correct responses. The maintenance of attentional control produces higher attention weightings in the digit chunks; however these weightings do not get distributed unless the production that uses a target chunk is fired. The prediction is that when people are able to maintain their attention selectivity, their goals, and their attention to relevant versus irrelevant information (by themselves or through the use of technology and design), they will maintain or improve accuracy in repetitive and prolonged work, rather than exhibit a decline in accuracy as found in these experiments.

Thus, training programs should determine what aspects of performance are more important for the task and context at hand, and be aware of the differences between response times and accuracy through the interactions produced by arousal and attention processes. If the goal is to reduce the number of errors in the task, designing technology and interventions to improve the attention control throughout the task is recommended. If the goal is to reduce response time, paying attention to how arousal will influence learning during task performance is recommended. Although many of the model's predictions remain to be tested empirically, this model suggests ways in which to improve or avoid decrements of performance in a prolonged, repetitive task.

Future research will focus on the empirical demonstration of the model's predictions, the better understanding of the generality of the fatigue construct proposed, and on ways to overcome the negative effects of fatigue through fatigue countermeasures. Also, for future modeling work, it is important to note that fatigue as a construct is not yet part of ACT-R. However, the existing ACT-R mechanisms G and W and their functions were sufficient to model these the simultaneous effects of fatigue and learning, and thus, the proposed theoretical mappings could be used in other models of fatigue. Including these ideas into the ACT-R architecture could offer several advantages: It would help make fatigue theories more precise and help extend the scope of the ACT-R architecture (Ritter, Reifers, Klein, & Schoelles, 2007).

In conclusion, this paper proposes a construct of fatigue that builds on the existing ACT-R architectural mechanisms. The proposition is that two aspects of behavior, arousal and attention processes, combine to produce concurrent effects of fatigue and learning, consisting of a decrease in both response time and accuracy with prolonged work.

Acknowledgements

This research was supported by a Multidisciplinary University Research Initiative grant from the Army Research Office (W911NF-05-1-0153). We are grateful to Frank Ritter, Kevin Gluck, and Glenn Gunzelmann for their helpful comments on previous versions of this paper.

Appendix A

In ACT-R 6.0 a new utility mechanism has been included, which is fundamentally different form the utility mechanism in ACT-R 5.0. We have implemented our model in ACT-R 6.0 but used the utility mechanism of ACT-R 5.0; the reasons should be obvious from reading the main text of this manuscript. This Appendix discusses the implications of this decision, particularly what would it take to modify the model presented in this paper to use the new utility mechanism in ACT-R 6.0, called the Temporal Difference, TD, algorithm.

In our model, we modified two parameters: W and G, where G was used as an arousal parameter, and it is not part of the TD mechanism of V 6.0 of the architecture. In contrast, W still exists in V 6.0 as a parameter in the Activation Equation (see Eq. (1) in main text). Thus, we focus this discussion only on the G parameter and suggest how we might be able to capture the same effect found in this manuscript with the new TD mechanism of V 6.0.

Decreasing G impacts production selection in the Utility equation of V 5.0 (see Eq. (2) in main text). A decrease in G makes the PG term smaller, so the cost of the production, C, becomes more important. This is part of what drives the switch to compiled productions (they are much lower in cost, even if they are not yet known to be accurate) in our model.

TD learning is designed to learn to estimate future rewards based on experience, and has a built-in credit assignment mechanism that reinforces the predicting stimuli. Because TD learning discounts future rewards in terms of their temporal proximity, it might already capture some of this pressure. Future rewards are discounted using a hyperbolic function. But also, a scale factor can be applied to objective rewards to influence their subjective value. This scale factor could make more distant rewards worth even less (see Fu and Anderson (2006) for a discussion of this, they used a scale factor of 1.0).

A second effect of decreasing *G* is that noise becomes more important (it becomes a greater proportion of the total utility as the PG term get smaller). Thus, some of the same effects that we capture in our model could be achieved by increasing the noise rather than by decreasing *G*. Taken together, increasing noise and (possibly) modifying the reward subjective scale factor might provide a similar behavior as in our present model. As far as we know, this possibility has not been tested, and we would need to see if the discounting of future rewards in the V 6.0 TD mechanism would account for the current interactions of *G* and the production compilation mechanism. This would be an interesting test for future research.

Regardless, we believe that by introducing the new TD mechanism, the ACT-R architecture is missing now a motivational, emotional or arousal parameter, *G*, which is an important element of the architecture that has been used to account for a very wide variety of motivational, emotional and fatigue effects in the last decade, as discussed in Sections 2 and 3 of the main text of this manuscript.

References

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

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1120–1136.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221–256.

Ballard, J. C. (1996). Computerized assessment of sustained attention: A review of factors affecting vigilance performance. *Journal of Clinical and Experimental Psychology, 18*, 843–863.

Bartlett, M. S. (1953). Approximate confidence intervals. *Biometrika, 40*, 12–19.

Belavkin, R. V. (1999). *Annual report: Studying the role of emotions in problem solving with cognitive architectures (Technical Report)*. Nottingham, UK: School of Computer Science and Information Technology, University of Nottingham.

Belavkin, R. V. (2001). The role of emotion in problem solving. In C. Johnson (Ed.), *Proceedings of the AISB '01 symposium on emotion, cognition, and affective computing* (pp. 49–57). York, England: Heslington.

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E. (2007). Modeling data entry in IMPRINT. In *Proceedings of the sixteenth conference on behavior representation in modeling and simulation* (pp. 205–206). Orlando, FL: Simulation Interoperability Standards Organization.

Chuderski, A., Stettner, Z., & Orzechowski, J. (2006). Modeling individual differences in working memory search task. In D. Fum, F. D. Missier, & A. Stocco (Eds.), *Proceedings of the seventh international conference on cognitive modeling* (pp. 74–79). Italy: Trieste.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.

Eysenck, M. V. (1982). *Attention and arousal, cognition and performance*. New York: Springer.

Fu, W. T., & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General, 135*(2), 184–206.

Fu, W., Gonzalez, C., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2006). Building predictive models of skill acquisition in a data entry task. In *Proceedings of the human factors and ergonomics society annual meeting (HFES 50th annual meeting)* (pp. 1122–1126). Santa Monica, CA.

Gawron, V. J., French, J., & Funke, D. (2001). An overview of fatigue. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 581–595). Mahwah, NJ: L. Erlbaum.

Gonzalez, C., Fu, W., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2006). ACT-R models of training data entry skills. In *Proceedings of the conference on behavior representation in modeling and simulation (BRIMS 2006)* (pp. 101–109). Baltimore, MD.

Gunzelmann, G., Byrne, M., Gluck, K. A., & Moore, R. L. (2009). Using computational cognitive modeling to predict dual-task performance with sleep deprivation. *Human Factors, 51*(2), 251–260.

Gunzelmann, G., & Gluck, K. A. (2009). An integrative approach to understanding and predicting the consequences of fatigue on cognitive performance. *Cognitive Technology, 14*, 14–25.

Gunzelmann, G., Gluck, K. A., Price, S., Dinges, D. F., O'Connor, R. M., & van Dongen, H. P. A. (2007). Decreased arousal as a result of sleep deprivation. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 243–253). New York, NY: Oxford University Press.

Gunzelmann, G., Gluck, K. A., Van Dongen, H. P. A., O'Connor, R. M., & Dinges, D. F. (2005). A neurobehaviorally inspired ACT-R model of sleep deprivation: Decreased performance in psychomotor vigilance. In B. G. Bara, L. Barsalou, & M. Bucciarellli (Eds.), *Proceedings of the twenty-seventh annual meeting of the cognitive science society* (pp. 857–862). Mahwah, NJ: Lawrence Erlbaum Associates.

Gunzelmann, G., Gross, J. B., Gluck, K. A., & Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science, 33*, 880–910.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E. Jr., (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10*, 188–199.

Hoffman, R. R., Sherrick, M. F., & Warm, J. S. (1998). *Viewing psychology as a whole: The integrative science of William N. Dember*. Washington, DC: American Psychological Association.

Jongman, G. M. G. (1998). How to fatigue ACT-R? In *Proceedings of the 2nd European conference on cognitive modeling* (pp. 52–57). Nottingham: Nottingham University Press.

Kole, J. A., Healy, A. F., & Bourne, L. E. Jr., (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology, 22*, 917–937.

Lewis, R. L. (2006). The surprising nature of working memory in language comprehension: Insights from cognitive architectures and psycholinguistics. In D. Fum, F. D. Missier, & A. Stocco (Eds.), *Proceedings of the seventh international conference on cognitive modeling* (pp. 16–17). Italy: Trieste.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). New York: Cambridge University Press.

MacLeod, C., & Mathews, A. (1988). Anxiety and the allocation of attention to threat. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 40*, 653–670.

Mackworth, N. H. (1969). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 17*, 203–325.

Pachella, R. G. (1974). The interpretation of reaction time in human information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halstead.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance*. *Cognitive processes and performance* (Vol. 2, pp. 1–43). New York: Wiley.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage Publications.

Raymond, W. D., Buck-Gengler, C. J., Healy, A. F., & Bourne, L. E., Jr. (2007). *Predicting differences in individual learning and performance in a motor skill task*. Paper presented at the 77th Annual Convention of the Rocky Mountain Psychological Association Meeting, Denver, CO.

Raymond, W. D., Fornberg, B., Buck-Gengler, C. J., Healy, A. F., & Bourne, L. E., Jr. (2008). Matlab optimization of an IMPRINT model of human behavior. In *Proceedings of the seventeenth conference on behavior representation in modeling and simulation* (pp. 26–34). Orlando, FL: Simulation Interoperability Standards Organization.

Ritter, F. E., Reifers, A., Klein, L., & Schoelles, M. J. (2007). Lessons from defining theories of stress. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 254–262). New York: Oxford University Press.

Ritter, F. E., Reifers, A., Klein, L. C., Quigley, K., & Schoelles, M. J. (2004). Using cognitive modeling to study behavior moderators: Pre-task appraisal and anxiety. In *Proceedings of the 48th meeting of The Human Factors And Ergonomics Society* (pp. 2121–2125). New Orleans: Human Factors and Ergonomics Society.

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 8602–8605). Amsterdam: Pergamon.

Smit, A. S., Eling, P. A. T. M., & Coenen, A. M. L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica, 115*, 35–42.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke?" A model of learning the past tense without feedback. *Cognition, 86*, 123–155.

Taatgen, N. A., & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors, 45*, 61–76.

van Rijn, H., van Sumeren, M., & van der Mass, H. L. J. (2003). Modeling development transitions on the balance scale task. *Cognitive Science, 27*, 227–257.

Veksler, B. Z., Gray, W. D., & Schoelles, M. J. (2007). From 1000 ms to 650 ms: Why interleaving, soft constraints, and milliseconds matter. In *Proceedings of ICCM- 2007 eighth international conference on cognitive modeling* (pp. 127–132). Oxford, UK: Taylor & Francis/Psychology Press.

Wickens, C. D. (1984). The multiple resources model of human performance: Implications for display design. In *Proceedings of the AGARD/NATO Report*. Williamsburg, VA: AGARD/NATO.