2007

An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention and Learning

Niels Taatgen
Carnegie Mellon University

Hedderik van Rijn
University of Groningen

John Anderson
Carnegie Mellon University

Published In
Psychological Review, 114, 3, 577- 598.

An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention and Learning

Niels Taatgen, Hedderik van Rijn and John Anderson

Carnegie Mellon University and University of Groningen

Abstract

A theory of prospective time perception is presented that extends existing theories by incorporating it as a module in an integrated theory of cognition, allowing predictions about attention and learning. First, a time perception module is established by fitting existing datasets (interval estimation, bisection and impact of secondary tasks on attention). The module is subsequently used as a part of the ACT-R architecture to model a new experiment that combines attention, learning, dual tasking and time perception. Finally, the model predicts learning and attention in a new experiment. The model fits and predictions demonstrate that the proposed integrated theory of prospective time interval estimation explains detailed effects of attention and learning during time interval estimation.

Keywords:

Prospective time estimation, cognitive model, divided attention, instance learning, multi tasking.

An Integrated Theory of Prospective Time Interval Estimation

The ability to estimate short time intervals routinely plays an important role in everyday life. Time estimates are important in situations where we take an action and expect a response, e.g., when we click on a link in a web-browser or when we judge whether we should brake for a yellow traffic light. It also affects multi-tasking situations where we have to switch between tasks after specific intervals, for example when using a mobile phone in a car (Kushleyeva, Salvucci & Lee, 2005; Salvucci, Taatgen & Kushleyeva, 2006). This type of time-interval estimation in real life is often implicit, automated and tightly interwoven with perception, learning and decision-making. It is called *prospective time estimation*, because the start of the interval is marked by an event that starts the estimate. This can be contrasted with retrospective time estimation, in which one is asked to estimate its duration after the time interval has passed. Prospective time estimation often is implicit in nature, because for most tasks the timing aspect is secondary to the real task being performed. For example, Grosjean, Rosenbaum and Elsinger (2001) found that participants in a choice-reaction time experiment adapt to the interval between stimuli without being aware of it. The implicit aspect sets it apart from many other forms of reasoning about time that involve explicit reasoning and problem solving (see Michon & Jackson, 1985 for an overview). E.g., in retrospective time estimation, an explicit reasoning process might be used that involves recalling events that took place between the onset of the interval and its end (Zakay & Block, 2004). Because prospective time estimation is so implicit, it might be understood best as a component of the human cognitive architecture, while other forms of time estimation might be better understood in terms of knowledge, i.e., strategies that are stored in memory but that are not part of the architecture itself.

Despite the fact that time estimation is, in general, only a component of complex task performance, it is usually studied in isolation. Zakay (1990) identifies four paradigms to study interval estimation: (a) verbal estimation: after exposure to a time interval report how much time has elapsed; (b) interval production: produce an interval of a certain duration, for example a minute; (c) interval reproduction: perceive an interval of a certain duration, and then reproduce it; and (d) interval comparison: compare two intervals and report which is longer. In each of these paradigms time estimation is the explicit focus of the task. It is therefore quite possible that, analogous to the observation that memory studies using explicit recognition and recall do not necessary tell the whole story on everyday implicit memory usage, explicit time estimation paradigms do not provide a complete picture of the role of time estimation within the cognitive system. They will, for example, not cover implicit timing in the real-life task of sending SMS messages with a mobile phone. In order to send a text message with a numerical keyboard, multiple letters are mapped onto a single number key. For example, the letters “DEF” are all under the number 3. In order to type the letter “E”, the “3” has to be pressed twice. In order to enter two consecutive letters from the same key (for example to type “DE”) a pause must be inserted (“3”-pause-“3”-“3”) to disambiguate it from three key-presses that signify the third letter (“F”). An important aspect in this task is learning: even if the manual of the phone states that the interval is one second, learning the exact interval is partly characterized by trial-and-error. Only after sufficient practice does pressing the key at the right moment become fully automated. In this task, timing is essential for accurate performance, but it is only instrumental in reaching the goal. Although it is likely that people will initially reason explicitly about the behavior of the telephone, timing the key presses will eventually become automated and will require no or little attention. Understanding this process is outside the scope of explicit interval estimation.

In this paper, we present a model of prospective time estimation as a module in a larger theory of cognition. We describe how this module interacts with other aspects of cognition to explain a wide variety of phenomena associated with time estimation. This embedded approach is necessary to fully understand the role of timing in both laboratory settings, and tasks like sending SMS messages, driving a car, and other complex skills. We embed the timing model in ACT-R (Anderson et al., 2004), a cognitive architecture that supplies mechanisms for learning, attention, perception and motor behavior, and which has been applied to many different tasks with a wide range of complexity. Before we present our own model, we will review the existing models of time estimation.

Existing Theories of Prospective Time Estimation

Two theories address interval estimation: the *internal clock theory* and the *attentional counter theory*. Each of these theories has been formalized in one or more models, *internal clock models* and *attentional counter models*. Figure 1 illustrates and example model of each of the theories. The main difference between the two theories is the role of attention.

Figure 1a depicts the *pacemaker/accumulator internal clock model*, as described by Matell and Meck (2000). Following Gibbon (1977), they identified a series of models that share the property that the internal clock itself is unaffected by outside processes. In one of these models, an internal pacemaker produces a steady stream of pulses. An accumulator counts these pulses, but only after a gate is opened by a start signal. After the time interval is finished, the accumulated value of pulses is stored in memory. When an interval of equal length has to be reproduced, a start signal is sent to the gate, and pulses are counted until the same number of pulses is reached that was stored in memory. However, this model cannot account for differences in timing accuracy in tasks where attention is (partly) directed away from the timing process. The attentional counter theory (depicted in Figure 1b, Zakay & Block, 1997; Hicks et al, 1976; Thomas & Weaver, 1975) was developed to explain how attention influences time estimates. The model associated with this theory is an extension of the pacemaker/accumulator internal clock model. In addition to the components in that model, it assumes that the accumulator is only updated when attention is directed to the timing process, opening a second gate. As soon as attention is directed elsewhere, the accumulator is not augmented until attention is returned. This way, attention determines the frequency by which the accumulator is updated.

The internal clock theory

The internal clock theory is part of a long tradition of studying time perception in animals and in psychophysics. In some of Pavlov's experiments, the reinforcement was delayed by a particular time interval. When dogs were trained with the delay, they would anticipate this and only start salivating at the end of the interval (Pavlov, 1927). Many other animal studies have shown rats, dogs, pigeons and other animals to be capable of learning the temporal structure of tasks. Studies in psychophysics have shown that time perception shares characteristics with other forms of perception, most notably Weber's law. The consequence of this law is that uncertainty in a time estimate scales with the magnitude of the interval, which is also called the *scalar property* of time estimation (Gibbon, 1977).

Matell and Meck (2000) give an overview of three possible models of the internal-clock theory: a *pacemaker-accumulator*, a *process-decay,* and an *oscillator/coincidence detection* model. All three models are based on an internal clock that is not affected by attention. Figure 1a depicts an example of a pacemaker-accumulator model in which an accumulator counts the pulses generated by a pacemaker. In a process-decay account, decay of activation in memory is used to estimate elapsed time. In the oscillator/coincidence detection account, which is favored by Matell and Meck because of its neurobiological feasibility, stimuli can synchronize neurons in a certain area of the cortex, effectively acting as a starting sign. As each of the neurons produces its own particular pattern of activation over time, each interval is associated with a unique pattern of activation, which can serve as a basis for later comparison. Although Matell and Meck's three models differ in their neurobiological plausibility, they are equivalent with respect to time-estimation related predictions. Because we are, in the context of this article, primarily interested in the behavioral characteristics of time estimation, we consider these implementations as belonging to one family.

Contrary to the attentional counter theory, internal clock models do not require any attention, and errors in time estimation are due to noise in the system. In a typical interval timing experiment participants were trained on an interval of either 8, 12 and 21 seconds by being exposed to it multiple times, which they then had to reproduce (Rakitin et al, 1998). Each participant produced 80 estimations, and was given feedback about the true duration every few trials. Figure 2 shows the distributions of the responses. Although the variance increases for larger intervals, the peaks of each of the distributions align with the duration of the target interval.

Consistent with Weber's law these distributions exhibit the scalar property: the variance in the estimation increases approximately linearly with the length of the interval. If we were to divide the times on the x-axis in Figure 2 by 8, 12 and 21 respectively, and readjust the proportions on the y-axis, the three curves would fall on top of each other (see Rakitin et al, 1998, Figure 2). This scalar property can be reproduced by adding noise to each generated pulse.

The generator-accumulator model assumes that by means of showing the intervals at the beginning of the experiment, an accurate imprint of the time interval is stored for later comparison. Therefore, this model can be considered an account of expert behavior, and the residual variance as reflecting an inherent inaccuracy in the timing system itself that cannot be overcome by training.

The attentional counter theory

In experiments associated with time estimation, the estimation task itself is almost always accompanied by an unrelated secondary task. The purpose of the secondary task is to prevent explicit counting, because counting makes time estimation much more accurate (e.g., Rakitin et al., 1998, Experiment 2). However, the nature of this secondary task turns out to have an influence on the estimation of the interval. If the secondary task is very demanding, people's estimation of duration tends to be shorter than when the secondary task is less demanding (Zakay & Block, 1997). The explanation given by the attentional counter theory is that fewer pulses accumulate when another task demands attention, leading to a shorter estimate. Many experiments have confirmed this finding (Block & Zakay, 1997, offer a meta-analysis of 20 experiments).

Here, we will focus on an experiment by Zakay (1993), where participants had to estimate and reproduce a single interval (12s) once. Each participant was required to perform a second task. This task had to be performed either during the presentation of the interval, when participants have to determine the duration, or during the reproduction of the interval, when participants had to use their perceived duration for reproduction. The attentional counter theory predicts that a demanding task during the reproduction leads to an estimate that is too long, because during the reproduction the timer is slowed down, and that a demanding task during the presentation leads to an estimate that is too short, because fewer pulses have been counted during presentation (Figure 3b). The secondary tasks were, in increasing complexity: (a) ET (Empty Time), no secondary task; (b) W (words), reading color words printed in black, (c) CW (color words), the Stroop task, naming the color of color words printed in incongruent ink; or (d) CWA (color word associations), like the Stroop task, but now participants had to name a word associated with the ink color. In the relatively easy ET and W conditions there is no effect of the secondary task on time estimates, but in the more demanding CW and CWA tasks, time estimates are affected in the way the attention counter theory predicts (Figure 3a).

A weakness of the attentional counter model is that the impact of attention cannot be quantified precisely. The model has no basis to assess the proportion of time spent with the attentional gate opened, versus closed. At the same time, one might question whether a timing model should account for the precise, quantifiable performance of non-timing aspects of a task. Another weakness of the model is that it has no account for how time intervals are learned, and how this learning might be influenced by differences in attention. In experiments supporting the attentional counter theory participants only make a single estimate of a time interval they have perceived once (single estimate is one of the criteria that Block & Zakay, 1997, use for inclusion in their meta analysis). The problems of the model can only be addressed by integrating the timing mechanisms with other aspects of cognition. With no connection to specified attentional processes, the attentional component is a free parameter in the model that can be set to any possible value to match the data.

Overview of the article

Internal clock models mainly focus on the estimation of the interval itself, and only have an approximate theory on how this connects to other aspects of cognition. The attentional counter model incorporates one of these aspects, attention, but the extent of the model is limited to explicit, one-shot time estimates. Even in those situations the model's predictions are mainly qualitative.

We take the approach to consider time estimation as a module in a more extensive cognitive system. In this approach, the timing module accounts for all primary aspects of time estimation. Secondary aspects of time estimation, such as attention and learning, are explained by how timing interacts with the rest of the cognition. The impact of attention is explained by general models of divided attention. Existing theories of learning and skill acquisition account for specific learning effects found in interval estimation.

In order to be able to understand interval estimation in a broader cognitive context, we have designed a time estimation module embedded within the ACT-R cognitive architecture (Anderson et al., 2004). The basis for the module is a pacemaker-based internal clock, with functional characteristics similar to the internal clock accounts proposed by Matell and Meck (2000). Interaction with the rest of the system allows explanations for the role of attention and learning without the need to incorporate the latter in the timing module itself.

We build our case as follows: On the basis of the distributions of time responses found by Rakitin et al. (1998) we have constructed an internal clock module that reproduces these distributions. We then validated this module by reproducing the results of the bisection experiments of Penney, Gibbon and Meck (2000). To test the module in a context where attention, learning, perception and motor actions interact, we have constructed a task in which keeping track of time intervals is only a single aspect of what participants have to do. The goal of this experiment was to test the quantitative accuracy of the model, but also to find support for the role of attention and learning as it is hypothesized by the incorporation of the temporal module in ACT-R. To this end, we have first conducted the experiment, and have then constructed the model to fit the data. To test whether our account also holds when the to-be-explained data are not known beforehand, we changed the setup of the experiment to make attending the time harder than in the first experiment. Before conducting the experiment itself we applied the model to this new task to predict the outcome. The experiment then confirmed the predictions made by the model.

Because it is impossible to discuss all the implementation-details of the models in article, we have made the code available on the Internet on http://act-r.psy.cmu.edu/models.

The Basic Internal Clock Module

In this section we will establish the internal clock module and will only make minor assumptions about processing in the rest of the architecture. The larger architecture and its impact on time perception will be discussed in the next section. As discussed earlier, the internal clock module is based on the pacemaker-accumulator model for pragmatic reasons. According to the pacemaker-accumulator model, a pacemaker generates pulses at certain intervals, which are counted by an accumulator. A reset event sets the accumulator to zero, after which it starts counting pulses anew. We found that to capture the scalar property, the interval between the pulses had to increase with the interval that has to be estimated, like a metronome that ticks slower and slower as time progresses. The interval estimate is based on the number of pulses the accumulator has counted. Because the gradual slowing of the pacemaker occurs in both perception and reproduction, it will lead to the same estimate, although less precise with longer time intervals as the time between pulses gets longer. In our approach, the temporal system is considered a module with an internal process that runs independently from other cognitive processes. The cognitive system as a whole (i.e., the ACT-R architecture) has only access to the result of this process, i.e., the current value of the accumulator.

More precisely, the duration of the first pulse is set to some start value:

$t_0 = startpulse$

Each pulse is separated from the previous pulse by an interval that is *a* times the interval between the previous two pulses. Noise from a logistic distribution is added to each pulse. The distribution of this noise is determined by the current pulse-length, modified by a parameter *b*.

$$t_{n+1} = at_n + \text{noise}(mean = 0, sd = b \cdot at_n)$$

These equations have three parameters, *startpulse*, *a*, and *b.* As the behavior of the timing module is assumed to be independent from the task the architecture currently executes, these parameters should be estimated to fit a single "benchmark-task" and then be left untouched. As a benchmark-task we used the Rakitin et al. (1998) experiment presented in Figure 3. We have left the parameters at the estimated values for all other models presented here.

A model of perception and reproduction of time intervals

The basic model for perceiving and reproducing intervals is simple. During the perception of an interval the accumulator is reset and the pacemaker is started at the beginning of the interval. The value of the accumulator is read at the end of the interval and stored in memory. If there are multiple presentations of the interval, the stored values are averaged to get a more accurate estimate[1]. Reproducing an interval means starting the timer, waiting until the accumulator has reached the stored value, and then making the response. Figure 4 illustrates this process. When the light turns on, the value of the counter, equaling five pulses, is read out and stored in memory. When the interval has to be reproduced, the timer is started again, and when counter reaches the stored value, the model assumes that the same amount of time has passed. In this case, the reproduced time is slightly more accurate than the stored time would have suggested, which is caused by the noise in the timing system.

In Experiment 3 of Rakitin et al. (1998) participants were first trained on a certain time interval (8, 12 or 21s). Training consisted of 10 trials in which a blue rectangle appeared on the screen, and changed to magenta when the time interval had elapsed. In the 80 test trials participants had to predict the interval by pressing a key when they expected the rectangle to change color. To make sure that the participant's representation of the interval would not drift, the rectangle changed color when the interval had elapsed in 25% of the test trials. The results are based on the remaining 75%, in which the rectangle stayed blue. Participants were instructed not to count during the experiment.

The model of this experiment closely resembles the example in Figure 4. In the learning phase the number of pulses in the interval is estimated (the model takes the average of the ten presentations), and during the testing phase the model waits until the appropriate number of pulses has passed and then makes a response. Based on least-square estimations of fit between model and data, we estimated the following values for three model parameters: 11 ms for *startpulse*, 1.1 for *a*, and 0.015 for *b*. Figure 2 shows the fit between this model and the three conditions of the Rakitin et al. (1998) experiment.

The fit between model and data is overall very good. This is no surprise considering there are three parameters to fit the data. The only aspect of the data the model does not predict perfectly is the shape of the tail of the distribution for the 8 and 12 second conditions. However, in similar experiments (e.g., Experiment 1 in Rakitin et al., 1998) the tails of the distributions are much shorter. We therefore decided against complicating the temporal module with a mechanism to simulate this aspect of the data, and made the assumption that these tails were produced by factors outside of the temporal module itself.

Application of the module to bisection experiments

Another paradigm in time perception concerns so-called bisection experiments. In these experiments, participants are trained on two time intervals, one short interval and one long interval. After this learning phase, they are exposed to new time intervals that are either equal to the short or the long interval or somewhere in between. Participants are then asked to judge whether the presented interval is closer to the short or to the long interval. To test whether the estimated parameters fit this timing paradigm equally well, we will model Experiment 2 from Penney, Gibbon and Meck (2000). In this experiment, three short-long pairs of intervals were used: 3 and 6 seconds, 2 and 8 seconds, and 4 and 12 seconds. In the training phase 10 tones of either the short or the long duration were presented to the participant. After that, participants were tested for 100 trials, 30% of which were anchor point intervals (short or long), and 70% were tones of different durations in between the anchor points.

The model uses the training phase to determine the timing for both the short and the long interval. During testing, it counts the number of pulses during the presented interval, and then compares the value to the number of pulses associated with both anchor intervals. If the value is closer to that of the short interval, it answers “short”, if it is closer to the long interval, it answers “long”. The parameters for the model are identical to those used to fit the interval estimation experiment earlier.

Figure 5 shows the results of the experiment and the model. A typical result in bisection experiments is that an interval exactly in between short and long is judged to be long considerably more often than chance. For example, in the 2-8 second version of the task (Figure 5b), 5 seconds is judged to be long 80% of the time. This happens because 2 seconds corresponds to an average of 33 pulses, 5 seconds to an average of 42 pulses, and 8 seconds to an average of 47 pulses. In terms of pulses, 5 seconds (42 pulses) is much closer to 8 seconds (47 pulses) than to 2 seconds (33 pulses). As can be seen in Figure 5, this mechanism yields very similar results as observed by Penney, Gibbon and Meck (2000).

An alternative model of Zakay's (1993) results

In the introduction we discussed the experiment by Zakay (1993) in which time estimates were influenced by the difficulty of a secondary task. This effect was reason for Zakay and others to propose the attentional counter theory, introducing a modulating effect of attention on the timing mechanism. However, an important aspect in the experiment is that all participants made only a single estimation of an interval. This might have made them prone to all sorts of "startup" mistakes. One possible mistake is that the temporal module is accidentally used for one of the secondary tasks. The temporal module could be used implicitly to estimate the inter-stimulus intervals in the secondary task (as in the Grosjean et al., 2001, experiment). This would effectively reset the accumulator to zero. The probability for this becomes larger as the secondary task becomes more demanding. A reset during the presentation means fewer pulses are counted, leading to shorter estimates, while a reset during reproduction is restarted somewhere halfway. Simple put, people might just "forget" estimating the time, and cope with this by starting anew. In order to reproduce the Zakay data, we set the interruption probability for each pulse to 0% for empty time (ET), to 1% for word reading (W), to 3.5% for the Stroop task (CW), and to 5.5% for Stroop with association (CWA). Based on these estimates, the model produces the results in Figure 3, essentially producing a fit that is very similar to what the attentional counter theory would predict.

Although the model based on the temporal module presented here requires the estimation of disruption parameters for each of the four tasks, it is similar to an attentional counter model in explanatory power, as that account would also require an estimate of the proportion of attention that can be directed towards interval estimation for each task. In fact, the two explanations are remarkably similar on the surface, despite the fact that the internal structures that produce them are quite different. The fact that two conflicting models can both fit the data shows that the validity of both theories (attentional counter and the temporal module) cannot be decided on the basis of this case alone, and that additional tests are needed.

Summary

We have described a temporal module that was designed to reproduce the distribution of estimates in a simple timing experiment. A model based on this module proved to be capable of explaining a second class of timing experiments, the bisection experiments, without adjusting any parameters of the underlying module. To fit the Zakay (1993) data we made some assumptions about the possibility that timing is disrupted. Based on these cases we can conclude that the temporal module is quite successful in explaining data from existing time-estimation tasks.

The simplicity of the tasks in this section allowed explanations in terms of the temporal module and a few additional assumptions. Although it is successful in the sense that it can offer explanations for experiments that have been studied in the context of both the internal clock and the attentional counter theory, it has not yet offered any details of larger integration with other aspects of cognition. Both the attentional counter theory and the temporal module can explain Zakay's data, but neither provides an account in terms of attention as a general cognitive process. Furthermore, most studies on timing have neglected the effects of learning. Do time estimations get better over time? And if so, how does attention modulate this effect?

To answer these questions we have designed an experiment that incorporates both learning and attention. To account for the data from this experiment, we have incorporated the temporal module into the ACT-R architecture. ACT-R already provides mechanisms for learning and attention. This provides us with an appropriate test-bed for assessing our claim that timing is in integral aspect of cognition, and that the interplay of different cognitive mechanisms result in the observed timing effects.

Experiment 1

The goal of the first experiment is to study the effects of attention and learning on interval estimation in a fairly complex task, in which time estimation, at least from the participant's perspective, is just one of the many prerequisites to achieve accurate performance. The experiment is supposed to mirror real-life situations in which people have to discover the temporal structure of a situation or a device, such as the SMS example in the introduction. In the case of sending SMS messages responding too early leads to an error, while responding late only costs more time. The present experiment is analogous to that situation.

The general idea of Experiment 1 is that participants work on two simultaneous tasks which are either both hard (verify additions), or both easy (recognize letters). Points are awarded for each correct response. A time interval has to be estimated as part of one of the tasks. The duration of this interval has to be determined by trial-and-error while doing the other task. One aspect of this experiment is that the primary goal from the perspective of the participant is to respond to the stimuli (because that scores points directly), and estimating the interval is only secondary (because it helps in scoring more points). To be successful at the task, it is necessary to spread attention over all the subtasks. To study the specific effects of attention, the task will switch at some point from easy to hard in one of the conditions and vice versa in one of the other conditions. This change in task difficulty modifies the amount of attention that can be spent on estimating the interval. Because the task involves many repetitions it is also allows the study of the effects of learning.

Although the model was constructed and fitted to the data after the experiment, we did have some expectations of the results on the basis of the timing module and general ACT-R characteristics.

A first expectation is a learning effect in the sense that participants will become increasingly better at estimating the interval and improve their score. The basis for this is that ACT-R generally learns from experience by storing and retrieving examples of past behavior. As the model gains experience it will be able to approximate the time interval with increasing accuracy.

A second expectation, and contrary to the attentional counter theory's prediction, is that the interval transfers perfectly from one task to the other, i.e., when the duration of the interval is learned during one task, this knowledge can be used to estimate the interval for a different version of the task that places higher or lower demands on attention. As a consequence we expect that the effects of changing task difficulty on time estimates will be small. The basis for this is the temporal module itself: the estimate it makes is not influenced by attention, and given the many repetitions the start-up problems that we hypothesized as affecting performance in the Zakay (1993) task should not occur.

Method

Participants.

Thirty-two students from the Carnegie Mellon University (17 males and 15 females) volunteered to participate in the experiment. Volunteers were paid for their participation.

Experimental task.

The display is divided into two halves (Figure 6a). The left half contains a high profit box, the right half a low profit box. Stimuli, to which participants may respond, appear in each of these boxes. Stimuli are buttons with, depending on the condition, either an addition with one-digit numbers or a letter. Additions are either correct or wrong by one, and letters are either “A” or “B”. Participants have to respond to correct additions and “A”s by clicking them with the mouse when they are in the right box, or by pressing space on the keyboard when they are in the left box. Incorrect additions and “B”s are to be ignored.

Figure 6 shows a screenshot of the task in the Addition condition, and a description of the 13-second trial in Letter condition. Stimuli in the right box appear for 1200 ms and are separated by a random interval between 0 and 2000 ms. Stimuli in the left box do not appear automatically: they are only available during certain time periods, and after the participant has pushed the “Test High” button. The basic trial is 13 seconds long and built up as depicted in Figure 6b. To indicate the previous trial has ended, the text “End of high profit” appears in the left box. During the next 7 seconds the interface will only present stimuli in the right box. Correct responses (clicking on “A”s or correct additions) in the right box score 30 points. Clicking on the “Test High” button (we will call this the test button from here on) during this interval has no effect, except that 10 points are deducted from the score. When the test button is clicked after 7 seconds have passed (but before the end of the 13 sec trial), stimuli start appearing on the left side (and 10 points are deducted). Stimuli in the left box appear for 1200 ms each, and are separated by 300 ms intervals. Participants have to respond to targets in the left box by pressing the space key on the keyboard. Correct responses in the left box score 100 points. Because stimuli also keep appearing in the right box, it is possible to work on both boxes at the same time. After a total of 13 seconds has passed, the text “End of high profit” appears in the left box, and no more stimuli appear in the left box, starting a new 13-second trial. “End of high profit” also appears when the participant has not clicked the “Test High” button at all.

Optimal behavior is to click the test button exactly seven seconds after the word “End” appeared in the left box. Participants are not informed about the length of the interval, they have to discover it themselves.

Design and procedure.

The experiment has four between subject conditions. Each condition consists of three phases. In each phase the task is either Letter (easy) or Addition (hard). The four conditions are: three phases with the Letter task (LLL), three phases with the Addition task (AAA), two phases with the Letter task followed by one phase with the Addition task (LLA), and two phases with the Addition task followed by one phase with the Letter task (AAL). Each phase consists of five blocks of nine 13-second trials. Each block lasts 120 seconds. Participants score 30 points for each correct response in the right box, 100 points for each correct response in the left box, they pay 10 points for each click on the test button, receive 30 penalty points for an incorrect response in the right box, and 100 penalty points for an incorrect response in the left box. Participants were instructed on all the aspects of the task, except the duration of the interval.

Results

Although we will postpone discussing details of the model until the next section, we have already included the model fits in this section's graphs to make comparisons easier.

Performance Scores.

The performance scores, i.e., the number of points participants gather, can be used to assess whether there is a global learning effect, and whether the difficulty of the two tasks (addition and letter) differs sufficiently to have an impact. Figure 7 illustrates the effect of phase (i.e., learning) and condition on the performance scores[2]. The two dashed lines represent the conditions that start out with the more difficult addition condition (A), whereas the two solid lines represent the conditions that start out with the simpler letter condition (L). After two phases, the condition represented by the dashed line with triangle markers switches to the simpler letter condition (AAL), the condition represented by the solid line with plusses switches to the more difficult addition condition (LLA).

To test the effect of phase (learning), task difficulty and switches in task difficulty, we analyzed the averaged performance scores for the three phases. These averages were entered into analyses of variance with phase as within-subject factor, condition as between-subject factor and subjects as random factors. The results show a significant interaction between condition and phase ($F(3,28)=22.04$, MSE=1050888, $p<0.001$). In addition to that, there are also main effects of phase ($F(3,28)=79.9$, MSE=3807377, $p<0.001$), and of condition ($F(3,28)=7.0$, MSE=2963876, $p=.001$). A Welch two-sample t-test between the average scores on phase 1 and 2 contrasting the Addition and the Letter task shows a significant effect of task difficulty ($t=5.35$, df=29.9, $p<0.001$). Paired t-tests between phase 1 and 2 for each of the two tasks reveal a learning effect for both the letter and the addition task (for the AAA and AAL conditions an improvement of 345 points: $t(15)=4.06$, $p=0.001$, and for the LLL and LLA conditions and improvement of 456 points: $t(15)=4.56$, $p<0.001$).

The interaction is driven by the transition from phase 2 to 3. In the conditions where the task remained the same, LLL and AAA, there is no or hardly any improvement (no improvement, $t<1$ for the paired t-test, in the LLL condition, and a non-significant improvement of 199 points in the AAA condition, $t(7)=2.2$, $p=0.06$). The score differences in the AAL and LLA conditions, however, are pronounced: in the case of a switch from L to A, there is a decrement in score of 596 points ($t(7)=-6.4$, $p<0.001$), and in the case of a switch from A to L, there is an increment in score of 672 points ($t(7)=7.88$, $p<0.001$).

To summarize: learning mainly occurs between phase 1 and 2, and the difficulty of the task has a strong effect on the score in all phases. Because the task difficulty changes in two of the conditions, there is an interaction between phase and condition.

Time Estimation.

The differences we found in the scores can be due to many factors, only one of which is the accuracy of the time estimate. We therefore look more closely at how the quality of the time estimate is influenced by learning and condition.

The two solid lines in Figure 8a plot the distributions of the moments at which participants first click the test button within a trial. These moments are defined as the deviation from the optimal time, that is, the time at which new high profit stimuli become available (which is seven seconds into the trial, so -7 is the beginning of the trial). A negative value represents a click that is too early, and a positive value a click that is too late. The data are averaged over the two conditions that start with the letter task (LLL and LLA) and the two conditions that start with the addition task (AAA and AAL), and are plotted separately for phase 1 and phase 2. The higher peak for phase 2 suggests that participants are more accurate in phase 2, indicating that a more accurate estimate is learned. The proportions plotted in these, and all subsequent, histograms are taken over the total number of trials in the phase (instead of the total number of first clicks). This means that trials in which no attempt at all is made to make a time estimate also weigh into the proportions (we will discuss these non-response trials later). The dotted line plots the distribution that would be expected if this were a pure interval estimation experiment like the Rakitin et al. (1998) experiment (see Figure 2). We derived this expectation by scaling, according to the scalar property, the distribution for the 8-second interval from the Rakitin experiment to the 7-second interval of this experiment. The wider empirical distributions indicate that participants do worse than ideal, that is, they deviate more from the optimal time, which could be expected because participants first have to discover the duration of the interval.

To determine the effects of learning and task difficulty on the quality of the time estimates, we analyzed the absolute values of the deviations from optimal time. As taking the absolute value introduced skewedness in the distribution of the data, we log-transformed the absolute deviations. Analyses of variance with phase as within-subject factor, condition as between-subject factor and subjects as random factors revealed only a main effect of phase ($F(28,3)=7.88$, MSE=0.905, $p=0.009$). Paired t-tests showed that this effect is due to learning between phase 1 and 2 (an improvement of 356ms, $t(31)=3.16$, $p=0.004$), but not between phase 2 and 3 ($t<1$).

In the discussion of the Zakay model we assumed that inaccuracies in time estimation were partly caused by participants forgetting about time estimation, and restarting it at some point if they do. In the present task forgetting to estimate results in making no estimate at all. This can be assessed by analyzing how often participants fail to make any estimate in a given trial. Figure 9 shows the proportions of non-responses by phase and condition.

These data were subjected to a logistic regression (Harrell, 2001) with proportion non-responses as response variable and phase and condition as predictors. In addition to these main predictors, an interaction between phase and condition was included to test for differential effects of task in different phases of the experiment. A significant main effect was found for phase ($Chi^2= 22.96$, df=4, $p<0.001$), indicating a decrease in non-responses due to learning. A main effect was also found for condition ($Chi^2= 21.33$, df=6, $p=0.002$), but no effect was found for the interaction. Combined with the analyses of the estimation accuracy this suggests that the task difficulty does not influence the accuracy of the time estimate, but it does cause people to sometimes omit making an estimate.

Changes in the accuracy of time estimation due to changes in task

An interesting question is what happens to the distributions of time responses when the task difficulty changes. According to the attentional counter theory, a major change in estimate should occur after the shift. This follows from the idea that fewer ticks reach the accumulator when the task is more difficult. One could compare this with a slower or faster ticking clock (Figure 3b). If the estimate of the interval is based on the clock that ticks fast, and the clock is slowed due to a more difficult task, the estimate should be too long. In the other direction, if the estimate is based on a slow clock that speeds up due to an easier task, the estimate should be short. Figure 10a shows what this theory would predict if the clock ticks 25% slower on the Addition task than on the Letter task.

Figure 10b compares the empirical distributions of the first clicks before and after a switch in difficulty. Although the changes in distribution are slight, there is a shift to the left in the AAL condition, indicating that participants click earlier after the change to an easier task, while there is a shift to the right in the LLA condition, indicating that participants click later after a change to a harder task. A paired t-test of the mean click-time the phase 2 and 3 in the AAL condition reveals that this shift is significant, from 701ms to 371ms ($t(7)=-2.61$, $p=0.035$), as is the shift between phase 2 and 3 in the LLA condition, from -624ms to 102ms ($t(7)=3.63$, $p=0.008$). Although the shift in the AAL condition can still be attributed to a learning effect, the shift towards later responses in the LLA condition seems to support the attentional counter theory. In order to obtain a better idea of the nature of the shift, we compare the average response times before and after the shift in task difficulty (Figure 11). The attentional counter theory would predict that the effect of the change would be most pronounced on trial 91, right after the shift, because that is where update rate of the accumulator has suddenly changed, and the participant has had no opportunity to adjust. Instead of a peak in click time that levels off afterwards, the click time increases gradually after the shift. Moreover, the average click time converges to that in the AAA condition but does not exceed it. This indicates that responses in the Addition task are all slightly later than those in the Letter condition, and independent of switches in the task. A comparison of the average response moment in phases 1 and 2 for the Letter and the Additions tasks confirms this (Mean response time for the L task is -472 ms, mean response time for the A task is 594 ms).

To summarize, task difficulty has hardly any impact on the accuracy of time estimation as evidenced in the timing of the first click. Instead it has an effect on how often a time estimate is missed, which is similar to our account of Zakay’s (1993) results. Time estimates for the Addition task are all slightly later than those for the Letter task, without affecting the absolute accuracy of the estimate.

The results up to here imply that task difficulty and shifts in task difficulty have an impact on the task performance and the role of interval estimation, but not in the way the attentional counter theory would predict them. The attentional counter theory would predict significant and immediate shifts in timing after a task change (the AAL and LLA conditions). The shift in the LLA condition was, however, small, and seems to be more related to global properties of the tasks than to changes between tasks.

Dual Tasking.

A possible explanation for the small impacts of task on time estimation is that the Addition task is too easy: Zakay only found attention effects in the more difficult secondary tasks (Zakay, 1993). If both the Addition and the Letter task are easy, enough time is left to keep track of time. But if that were the case, participants would also have enough time left to do a secondary task when they do not have to attend the time. A measure of dual tasking can be obtained by looking at these high-profit periods. As stimuli in the left box produce higher scores, we assume that people will only react to stimuli in the right box if they have spare capacity to do so. We therefore took as a measure of dual tasking the proportion of target stimuli in the right box to which the participant responded while there were also stimuli in the left box. Participants turn out to be able to achieve a level of 86% dual tasking in the Letter task, but only 47% in the Addition task (Figure 12). This shows that the Addition task does indeed require much more attention than the Letter task (Welch two-sample t-test between the average dual tasking scores in phase 1 and 2 for the two tasks: $t = 4.252$, df = 29.954, $p < 0.001$). According to the attentional counter theory this difference should have an impact on time estimation.

Taken together, our results are not consistent with the attentional counter theory, as this theory would predict a larger and different impact of the task-difficulty and switch manipulations. On the other hand, the internal clock theory does not cover this experiment because it has no explicit theory about how the clock interacts with aspects of cognition outside time management. The results also show pronounced effects of learning, which are covered by neither theory. The results are, however, consistent with the expectations that we formulated at the beginning of this section: a clear learning effect, and an unbiased transfer of the time estimate between tasks of different difficulty, as can be seen in Figure 10b.

The Integrated Model of Prospective Time Interval Estimation

In this section we will discuss how the temporal module fits into a cognitive architecture, and allows fitting the data from experiment 1. In addition to the temporal module, the cognitive part of the model builds on the ACT-R cognitive architecture, more specifically on earlier models of instance learning (Logan, 1988; Lebiere, Wallach & Taatgen, 1997) and of central bottleneck theories of divided attention (Pashler, 2004; Anderson, Taatgen & Byrne, 2005).

The ACT-R architecture and the role of attention

Figure 13 shows a general overview of the ACT-R architecture including the temporal module (Anderson et al., 2004). The center of the architecture is *procedural memory* (the production system) in the middle of the diagram. The production system has access to all the other modules in the system through *buffers*, each of which can only hold a single item of information. For example, the *temporal buffer* holds the current value of the accumulator, the *visual buffer* hold the currently attended visual stimulus, and the *retrieval buffer* holds the last element retrieved from *declarative memory*. The basic cycle of the central production system is that the contents of all the buffers are matched against the rules stored in procedural memory. A single rule is then chosen on the basis of its *utility*, and this rule carries out its set of actions that it communicates to the other modules through their respective buffers.

The temporal module's output is only one of many buffers that the production system can match, and if it is busy with another subtask in a multi-tasking situation, it may fail to integrate the information from the temporal module with other processing. More specifically, when dual tasking, the model might be busy attending visual stimuli and responding to them with motor responses. Part of this process involves declarative memory to determine whether the stimulus is a target or a foil. As a consequence, attempts at reasoning about time (which also involves declarative memory) can be postponed or disrupted, acting like a system with a central bottleneck (Pashler, 1994; Anderson, Taatgen & Byrne, 2005).

In this experiment the model has to divide its attention between three tasks: attending and responding to the left box, attending and responding to the right box, and attending the time. Only two of these tasks are relevant at the same time: either both the left and the right box have to be attended, or both the right box and the time. The model is mainly event-driven, and will respond to changes on the screen. When a new stimulus appears on the screen, the model will attend to it and initiate a response. The only exception is when the model is busy with a stimulus in the left box. In that case it will ignore stimuli in the right box until it is done with the stimulus in the left box. A stimulus in the left box can on the other hand interrupt processing in the right box. This reflects the fact that the score for the left box is 100 points, and for the right box only 30 points. Attending to the time interval is initiated whenever the model has no stimulus to process. Because retrieving a past experience takes time, especially when these experiences are relatively new and still have a low activation, attending to the time will be interrupted when a new stimulus appears on the screen, making it necessary to restart the time estimation process once the stimulus has been processed.

Learning the time interval through instance learning

Because the duration of the interval is initially unknown, the model has to determine it by trying out various intervals. ACT-R models generally deal with such a situation through *instance learning* (Logan, 1988; Lebiere, Wallach & Taatgen, 1997). Instance learning assumes that previous experiences, in this case of a specific time interval, are stored in (declarative) memory. For example, memory could contain an experience that 30 pulses is too short, and another that 50 pulses is correct.

When the model sees "End of high period" in the left box, signaling the start of the interval, it starts generating time pulses as illustrated in Figure 4. Whenever the model has time in between processing stimuli it will attempt to retrieve a previous experience of pressing the test button at approximately the present time. If a successful experience is retrieved, the model will initiate a click on the test button. If a "too short" experience is retrieved, the model will do nothing. Finally, if no experience at all is retrieved for the present time, the model randomly decides to click the button or not. Experiences are then stored in ACT-R's declarative memory, which has an activation-based mechanism to model forgetting. However, if two experiences are identical (i.e., concerning the same judgment for the same number of pulses), their activations are combined. Because of the decay in memory, a particular experience sometimes has to be repeated a number of times before it can be retrieved at all, and retrieval will become faster with frequent use (see Anderson et al., 2004, for the details of declarative retrieval).

After the button has been clicked, the model judges whether the button click was successful. If stimuli appear in the left box, the present time is stored as successful, but if nothing happens the present time is stored as “too short”. Note that because of the nature of the task, late test clicks are judged as successful, even if they are 4 seconds late, but early presses are judged as failures (“too short”), even if they are early by only 100 ms. As the model accumulates more experiences it will become more accurate at estimating the right interval, but only within the boundaries of the accuracy of the temporal module itself (i.e., what is depicted by the dotted pure interval estimation distribution in Figure 8). In addition to that, experiences of button clicks around the seven-second mark are a mixed set of successes and failures, adding noise to the timing process.

Model results

We used the same parameters for the temporal module as in all earlier models. The parameters that control the timing of the interaction with the interface (time to attend stimuli on the screen, timing of mouse actions) were left at their ACT-R default values. We estimated the parameter that controls the threshold at which elements in declarative memory are forgotten and the probability that the test button is clicked when no previous experience could be found to fit the score and time estimation data.

The model was run 100 times for each of the four conditions. The results of the model were already shown in Figures 7-10 and 12 to make comparisons with the data easier. The fitted scores match the participants’ scores (Figure 7) with respect to the main aspects in the data: a learning effect between phase 1 and 2, and differential effects of task difficulty on the change from phase 2 to 3. The improvement in the model is due to an improved accuracy in estimating the time interval allowing for more high-profit responses. This improved accuracy is evident in the distribution of the first clicks in Figure 8, where the peak of the distribution is higher in phase two, indicating more clicks are closer to the target click-time at zero. The model explains the effects of task difficulty by the fact that retrieving an addition fact from declarative memory takes longer than just retrieving that the letter “A” is a target. The increased retrieval time for the Addition task makes the model miss more targets (as is evident in the non-response data in Figure 9 and in the dual-tasking proportions in Figure 12).

Figure 8 and 10 show that the distribution of the time estimates of the model is very similar to that found in the data. The qualitative effects found in the analysis are also present in the model fits: there is an effect of learning on timing accuracy in terms of a 370 ms improvement in the absolute values of the deviations from optimal time between phase 1 and 2 (compared to 356 ms in the data), but only a 120 ms improvement between phase 2 and 3 (there was no improvement at all in the data). The differences in time estimation accuracy between the two tasks were very small in the model: on average the deviation on the Letter task was 149 ms shorter than on the Addition task: this corresponds well to the non-significant 94 ms difference in the data.

The shifts in average click time due to changes in the task are also produced by the model: after the task has changed from Letter to Addition, the average first-click time is 402 ms later (726 ms in the data), and after the task has changed from addition to letter, the average first-click time is 158 ms earlier (330 ms in the data). The model’s explanation for these shifts is that processing an Addition stimulus takes more time than a Letter stimulus. When the moment arrives to make a click, the model still has to complete its response to the current stimulus, resulting in slightly earlier clicks for the Letter task, and slightly later clicks for the Addition task.

The main impact of the difficulty manipulation is on the proportion of trials in which no response is made at all. This is also captured by the model. Although the graphs in Figure 9 are hard to compare due to the noisiness of the data, the model does exhibit the two main effects of condition and learning without an interaction that are present in the data. Finally, Figure 7 shows that the model also correctly captures the dual-tasking results, confirming that accuracy of time estimation and the amount of attention that can be devoted to it are relatively independent.

The two expectations we formulated before the experiment have been confirmed: there is a clear learning effect on the time estimates that participants make, and the accuracy of the time estimate transfers very well from one task to the other. The task difficulty manipulation had very little impact on the accuracy of time estimation: the distributions of the first-click times are very similar for all conditions. Instead, the impact of task difficulty is an increase in the proportion of non-response trials.

The key difference between the attentional counter theory and the ACT-R model is that the former predicts that a harder secondary task creates a shift in the time estimate, while the latter predicts that a harder secondary task makes it less likely that there will be a response at all. In the view of the attentional counter theory a shift to an easier secondary task would produce responses that are too early and vice versa, while the ACT-R model predicts that there is a change in non-response trials instead, which is confirmed by the data.

Although the empirical results are not consistent with the attentional counter theory, it can be argued that it not a strong test of the ACT-R model yet, because the task difficulty did not have a large impact on time perception, and the model was fitted to the data. To build a stronger case we have conducted a second experiment such that a strong effect of task difficulty could be expected on the basis of the model, but one that is different from what the attentional counter theory would predict. Moreover, instead of fitting the model to the data, we made a model prediction, thereby avoiding the criticism that insufficiently constrained cognitive models can be made to fit any dataset (Roberts & Pashler, 2000).

Experiment 2

The second experiment is identical to Experiment 1 with one major change: there are now always stimuli in the left box, instead of only during a high-profit period. Correctly identifying the high-profit period now only increases the score for each hit in the left box. Because participants can work on the two tasks all the time, there are fewer opportunities to estimate the time interval. In Experiment 1 participants managed to achieve a dual tasking score of 47% when the task was Addition, so they were not able to fully attend to two tasks. We therefore expect that they have little attention left to do even a third task, the estimation of the interval, leading to poor performance on time estimation. As we will see in the model predictions, we expect that in the Addition task so little attention can be devoted to time estimation that it is impossible for the learning mechanisms to pick up the duration of the interval.

The Model

As indicated in the introduction we have used the model for Experiment 1 to make a prediction for Experiment 2 before doing the actual experiment[3], with one slight modification: In Experiment 1 the feedback for a successful press on the test button is that stimuli start to appear in the left box, while in Experiment 2 it is the appearance of “HIGH” above the left box. We adjusted the model to be able to interpret the changed feedback correctly. Otherwise all parameters and production rules in the model have been kept the same. The main qualitative predictions are:

- The Addition task is so hard that it is almost impossible to learn the correct interval. We expect the accuracy of the presses on the test button to be at chance level in the conditions that start with the Addition task.
- The Letter task leaves some time to attend to learn the interval, and therefore the model predicts that participants are able to learn the interval and make reasonable estimates, although at a lower level of accuracy than in Experiment 1.
- Time estimation has to compete with two other tasks (the left and the right box). The assumption in the model is that a new visual stimulus will interrupt any ongoing reasoning about time. Although reasoning about the time interval can be restarted as soon as the stimulus has been processed, time has passed in the meantime making it necessary to start over again. In practice this means that in many trials there will be no attempt to click the test button at all. This is especially true with the Addition task, where the model predicts that in the majority of the cycles there will be no attempt at estimation, but also with the

Letter task, where the model predicts that no attempt will be made in approximately a quarter of the opportunities.

- When the task shifts from Letter to Addition in the LLAA condition, the model predicts no strong shift in the interval estimate (as opposed to the attentional counter theory). The model predicts that participants will even do better on the Addition task after two blocks of the Letter task than after two blocks of the Addition task because the Letter task offers better learning opportunities. The attentional counter theory would predict the opposite, because it predicts that the estimate of the interval during the letter task does not transfer to the addition task.
- When the task shifts from Addition to Letter in the AALL condition, the model predicts that participants will start picking up the length of the interval in the same way as when they had they just started out with the Letter task.

Method

Participants.

Forty students from the Carnegie Mellon University (21 males and 19 females) volunteered to participate in the experiment. Volunteers were paid for their participation.

Experimental task.

The task was identical to the task in Experiment 1 with the following modifications. In Experiment 1 stimuli appear in the left box only after the participant has clicked the test button in the high-profit period. In Experiment 2 stimuli appear in the left box all the time, but will only yield 100 points after the participant has clicked the test button in the high profit period, otherwise they yield 30 points, the same score as for the right box. Stimuli in the both boxes appear for 1200 ms and are separated by a random interval between 0 and 2000 ms. Feedback with respect to the high-profit period is given by displaying "HIGH" above the left box after the participant has clicked the test button in the high profit period. The "HIGH" text is removed at the end of a high-profit period.

Design and procedure.

The experiment had four conditions: four phases of five blocks with the letter task (LLLL), four phases of five blocks with addition task (AAAA), two phases of five blocks with the letter task followed by two phases of five blocks with the addition task (LLAA), and two phases of five blocks with the addition task followed by two phases of five blocks with the letter task (AALL). To prevent participants from counting the stimuli in the left box to help their time estimate, the inter-stimulus intervals of the left box are randomized using the same randomization process as is used for the right box. Otherwise the procedure was identical to the procedure of Experiment 1.

Results

In this section we will discuss the results of Experiment 2 alongside the predictions of the model. As the model predictions are overall similar to the experimental data., we will only highlight the aspects of the model that are of particular interest.

Performance Scores.

Figure 14 shows the scores achieved in the four conditions and four phases of the experiment. We again analyzed the averaged performance scores by phase to determine the effects of the change in task difficulty. An analysis of variance shows a significant interaction between phase and condition ($F(36,3)=40.5$, MSE=5647061, $p<0.001$), as well as main effects of phase ($F(36,3)=13.5$, MSE=1875016, $p<0.001$) and condition

($F(36,3)=8.93$, MSE=8018033, $p<0.001$), indicating strong effects of learning and task difficulty. Paired t-tests within each condition on the performance on phase 2 versus phase 3 indicate no significant performance difference for the AAAA condition, and no significant difference for the LLLL condition (AAAA: mean difference 86.4, $t=1.845$, $df=9$, $p=0.100$, LLLL: mean difference 6.8, $t<1$, df=9). In contrast, the effect of task change is pronounced in both AALL and LLAA conditions (AALL: mean difference: 824, $t=-6.262$, $df=9$, $p<0.001$, LLAA: mean difference -770, $t=7.709$, $df=9$, $p<0.001$), showing a performance improvement for the AALL condition, and decreased performance for the LLAA condition. All of these results are consistent with Experiment 1.

Welch two sample t-test comparisons between phase 1 of the LLLL condition and phase 3 of the AALL condition, as well as phase 2 of the LLLL condition and phase 4 of the AALL condition show no differences (both $t<1$), indicating that participants in the AALL condition have no benefit from doing the Addition task before they switch to the Letter task.

The model predictions are quite accurate, with the exception of the prediction that performance on the LLAA condition is better than the AAAA condition in phase 3 and 4. The empirical data show no significant difference between conditions.

Time estimates.

Figure 15 shows the distributions first-click times for the first phase of the experiment. Compared to Experiment 1 the distributions are much flatter, indicating that on many trials no response at all is made. This lack of response is made explicit in Figure 16, which shows the non-responses over the whole experiment. Consistent with the prediction of the model, participants have great trouble making accurate time estimates at all in the Addition condition. Of the 294 estimates that were made with the Addition task in the first phase, 148 responses were within 3 seconds of the optimal time, and 146 responses were more than 3 seconds early or late, suggesting that accuracy was indeed at chance level because according to the internal-clock distribution it is possible to make almost all estimates within these 3 seconds. In the Letter condition participants are able to make slightly better estimates, but are performing at a lower level than in Experiment 1. Nevertheless 408 of the 590 responses made are within 3 seconds of the optimum, and only 182 are outside of it. This clearly shows that the task manipulation in this modified experiment has a major impact on time estimation, which is correctly predicted by the model. The explanation the model offers is that in the Addition task the time estimation process is interrupted so often (much more often than in Experiment 1) that the few experiences it gets are spaced apart too far to produce a stable representation of the duration of the interval (in ACT-R terms: the activation of the experiences has dropped below the retrieval threshold at the time they are needed). When it does push the test button it is therefore a blind guess.

A logistic regression of the non-response data with proportion non-responses as response variable and condition, phase and the interaction between condition and phase as predictors now reveals a significant interaction between condition and phase ($\text{Chi}^2 = 8.89$, df = 3, $p=0.03$), as well as main effects of phase ($\text{Chi}^2 = 14.48$, df= 4, $p=0.006$) and of condition ($\text{Chi}^2 = 33.81$, df=6, $p<0.001$).

These results confirm what was found in Experiment 1: the difficulty of the task influences how often participants make a response at all. However, the magnitude of the effect is much larger in this experiment. In addition, there now also is an interaction between phase and condition (it was absent in Experiment 1 because by phase 3 the non-response rates were all low). The interaction in this experiment is due to the fact that in two of the conditions the task changes half way, and this affects the non-response rate (fewer responses when the task changes to Addition and more responses when the task switches to Letter). The model predicts both of these effects correctly, even though the exact fit between model and data is hard to assess because the data are again quite noisy.

Instead of a shift in non-response trials, the attentional counter theory would predict a shift in the estimate with a change in task difficulty. In Experiment 1 we saw a small shift in the estimate, so according to the attentional counter theory this shift should be much larger in Experiment 2. However, as can be seen in the results (Figure 17), there is no shift at all.

A paired t-test on the mean first-click time in phase 2 and 3 in the AALL condition reveals that there is no significant shift in first-click time, from -503 ms to -965 ms ($t<1$), nor between phase 2 and 3 in the LLAA condition, from 245 ms to - 345 ms ($t(7)=1.66$, n.s.).

Discussion

In the introduction to Experiment 2 we stated that the ACT-R model makes five general predictions. The first prediction was that participants would not be able to make accurate time predictions in the Addition task, and that the estimates they make are at chance level. This prediction was confirmed by the data: in the first phase with the addition task participants often failed to make a prediction at all, and if they made one it was at chance level.

The second prediction was that participants would be able to make proper time estimations in the Letter task, although at a lower level of accuracy than in Experiment 1. This prediction was also confirmed by the data.

The third prediction of the model was that the main impact of the increased difficulty is an increase of the number of non-responses, trials in which participants make no attempt at all to give an estimate. This prediction was also confirmed by the data.

The fourth prediction concerns the shift from Letter to Addition in the LLAA, where the model predicts no change in the time estimation after the shift. Moreover, it stated that participants would perform better in phase 3 and 4 than participants in the AAAA condition. This prediction was only partially confirmed: there was indeed no change in the estimated interval after the shift, but participants' performance turned out to be the same in the AAAA and LLAA conditions with respect to phase 3 and 4.

The fifth prediction concerned the shift in the AALL condition, in which the model predicts that participants' performance in phase 3 and 4 is just like performance in phase 1 and 2 of the conditions that start with the letter task (LLAA and LLLL). This last prediction was also confirmed by the data. For all five predictions (apart from the mispredicted part of prediction 4) the model was not only able to make a correct qualitative prediction, but also an accurate quantitative prediction.

Although the five predictions may not sound particularly counter-intuitive, many of them are clearly inconsistent with the attentional counter theory. The dominant factor in Experiment 2 is the proportion of non-response trials, which are presumably due to not paying attention to time estimation at the right moment. Because the attentional counter theory has a specific theory about attention in terms of limiting the number of ticks reaching the accumulator, it cannot deal with situations in which attention manifests itself differently, in this case by not responding at all. Instead it would predict shifts in the time estimates: when the task switches from Addition to Letter it predicts that estimates would be early, and in the switch from Letter to Addition it predicts that estimates are late. Whereas we found small effects in those directions in Experiment 1, they were smaller and no longer significant in Experiment 2, while according to the attentional counter theory they should have been larger. Instead the shifts have large impacts on the number of non-response trials, something that is not predicted by the attentional counter theory.

General Discussion

In this paper we have presented a theory of time estimation that is integrated in a larger theory of cognition with a focus on attention and learning. The core of the theory is a simple pacemaker-accumulator module that counts pulses as time passes, similar to the theory described by Matell and Meck (2000). The main twist in the mechanism is that the duration of the pulses increases with the interval, allowing the module to produce the scalar property of the variance in the estimate. The behavior of the module is controlled by three parameters that we estimated on the basis of the Rakitin et al. (1998) data and which were confirmed in accurately predicting bisection data (Penney et al., 2000).

In most time estimation studies, interval estimation is the main task. However, time estimation should also be studied in contexts where time estimation itself is secondary to a main task, because this corresponds to the natural role time estimation plays in everyday life. The success of the model does not depend on the actual mechanism of time estimation itself, but on the way it interacts with other aspects of cognition. The variant of the pacemaker-accumulator mechanism we have chosen accurately models the scalar property in the variance in time estimation. Any mechanism with the same properties can in principle replace it, for example an oscillator/coincidence mechanism (Matell & Meck, 2000), or a process-decay mechanism (Staddon, 2005). Once integrated in a large framework it can be used to model complex tasks in which time estimation is only a component, and make accurate predictions on the outcome.

The ACT-R architecture models attention, or more specifically divided attention, by having a control structure of the task that determines which cognitive modules participate in determining the next action. This control structure is necessary to prioritize the subtasks and to prevent interference, and produces behavior similar to central bottleneck theories of attention (Pashler, 1994). As a consequence of this, time estimation is often disrupted by new visual stimuli, producing the effects of attention we found in the experiments. As we have demonstrated, this more general theory of attention offers a better explanation of the data than the more specialized attentional counter theory.

Learning of times estimates was modeled with an instance retrieval strategy: experiences with a certain time interval are stored in declarative memory, and can be retrieved for future decisions. Accumulating experiences improve the estimate, and increase its activation in memory, which speeds up its retrieval. Although we have simplified the process in the first two models, this learning process plays a role in all models discussed in this article.

Although the model for the dual-task timing task was specifically designed for one task, the principles of learning and attention are general enough to be extended to other tasks. We have already successfully modeled operating a typing device while driving a car based on the temporal module introduced here and the same principles of instance learning and dual tasking (Salvucci, Taatgen & Kushleyeva, 2006). Other potential situations in which timing is relevant are discovering how to interact with new devices, for example determining how long to turn the key before the car engine starts, how long to wait after pushing the power button on a camera before it is ready to make a photo, or how long to wait with putting the meat in the pan while the oil heats.

An open question that remains is whether there is only a single timer, or whether our cognitive system can time multiple things at the same time. For example, in some of the conditions in the bisection experiments (Penney et al., 2000), participants had to estimate two slightly staggered intervals at the same time, and performed almost identical to when they have to estimate only a single interval. Although it is possible that separate timers track both intervals, it is also possible a single timer is used to track all the intervals in between events, and that explicit reasoning is used to find the answer. As long as there is no clear evidence for multiple timers the more parsimonious assumption is that there is only a single one.

A related question is whether the timer can be stopped and started again later. Fortin, Bédard and Champagne (2005) found that estimations of time intervals that were interrupted depended on where in the interval the interruption was placed. This makes it unlikely that the timer can just be stopped and restarted, and that instead an explanation in which explicit reasoning about time determines the estimate is more appropriate to explain these results. Although we do not offer a theory of explicit reasoning about time in this article, the ACT-R architecture in general has many mechanisms that can help in building such theories. However, we consider it unlikely that a single theory of explicit reasoning about time can cover all phenomena. Instead, each particular phenomenon will have to be explained by assuming task-specific strategies, for example retrograde time estimation for which explicit reasoning instead of an internal mechanism is responsible (Block & Zakay, 1997).

As a final note, a criticism of fitting cognitive models is that given enough free parameters any dataset can be modeled (Roberts & Pashler, 2000). We have tried to counter this criticism by estimating the three parameters of the temporal module for the very first task only, and using those parameters for all other models. With more complex tasks the models also become more complex, and made it necessary to estimate additional, non time-estimation related, parameters. For the dual-task timing model, we used the data from Experiment 1 to estimate the non-temporal parameters. This model was then used to make a true prediction for Experiment 2. Although Experiment 1 guided this prediction, the predictions for Experiment 2 were novel, e.g. that time estimation in the addition task completely breaks down. Given the success of the model we are confident that the temporal module can be used in all situations in which intervals in the order of 1-30 seconds have to be estimated.

References

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

Anderson, J.R., Taatgen, N.A. & Byrne, M. (2005). Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, & Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31*(4), 749-761.

Block, R. A. & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*(2), 184-197.

Fortin, C., Bédard, M.C. & Champagne, J. (2005). Timing during interruptions in timing. *Journal of Experimental psychology: Human Perception and Performance, 31*(2), 276-288.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279-335.

Grosjean, M., Rosenbaum, D.A. & Elsinger, C. (2001). Timing and reaction time. *Journal of Experimental Psychology: General, 2*, 256-272.

Harrell, F. E., Jr. (2001). *Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Hicks, R.E., Miller, G.W., Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology, 89*, 719-730.

Kushleyeva, Y., Salvucci, D. D., & Lee, F. J. (2005). Deciding when to switch tasks in time-critical multitasking. *Cognitive Systems Research*, *6*, 41-49.

Lebiere, C., Dieter, W. & Taatgen, N.A. (1998). Implicit and explicit learning in ACT-R. In Frank Ritter & Richard Young (eds.), *Proceedings of the Second European Conference on Cognitive Modelling* (pp. 183-189). Nottingham, UK: Nottingham University Press.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Matell, M.S. & Meck, W.H. (2000). Neuropsychological mechanisms of interval timing behavior. *BioEssays, 22*, 94-103.

Michon, J.A. & Jackson, J.L. (1985). Introduction: the psychology of time. In J.A. Michon & J.L. Jackson (Eds.), *Time, mind, and behavior* (pp. 2-17). Berlin: Springer.

Pashler, H. (1994). Dual-Task interference in simple tasks: Data and Theory. *Psychological Bulletin, 16,* 220-244.

Pavlov, I.P. (1927). *Conditioned Reflexes*. Oxford, UK: Oxford University Press.

Penney, T.B., Gibbon, J., & Meck, W.H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1770-1787.

Rakitin, B.C., Gibbon, J., Penney, T.B., Malapani, C., Hinton, S.C. and Meck, W.H. (1998). Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 15-33.

Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358-367.

Salvucci, D.D., Taatgen, N.A. & Kushleyeva, Y. (2006). Learning when to switch tasks in a dynamic multitasking environment. *Proceedings of the seventh international conference on cognitive modeling* (pp. 268-273). Trieste, Italy: Edizioni Goliardiche.

Staddon, J.E.R. (2005). Interval timing: memory, not a clock. *Trends in Cognitive Sciences, 9*(7), 312-313.

Thomas, E.A. & Weaver, W.B. (1975). Cognitive processing and time perception. *Perception & Psychophysics, 17*, 363-367.

Zakay, D. (1990). The evasive art of subjective time measurement: some methodological dilemmas. In R.A. Block (Ed.), *Cognitive models of psychological time* (pp. 59-84). Hillsdale, NJ: Erlbaum.

Zakay, D. (1993). Time estimation methods–do they influence prospective duration estimates? *Perception, 22,* 91-101.

Zakay, D. & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science, 6*(1), 12-16.

Zakay, D. & Block, R. A. (2004). Prospective and retrospective duration judgments: an executive-control perspective. *Acta Neurobiologiae Experimentalis, 64*, 319-328.

Author Note

Niels A. Taatgen, Psychology, Carnegie Mellon University and Artificial Intelligence, University of Groningen; Hedderik van Rijn, Artificial Intelligence, University of Groningen; John R. Anderson, Psychology, Carnegie Mellon University.

This research was supported by Office of Naval Research grant N00014-06-1-0055. We thank Stefani Nellen for her helpful comments on the manuscript, and Daniel Dickison and Lisette Mol for collecting the data.

Correspondence concerning this article should be addressed to Niels Taatgen, Department of Psychology, Carnegie Mellon University, Pittsburgh PA 15213. E-mail: taatgen@cmu.edu.

Footnotes

[1] For the models that will be discussed later a more sophisticated learning technique, instance learning, will be used. However, for this experiment and the bisection experiment we will discuss next, instance learning and averaging produce similar results.

[2] All plots of empirical data are based on participant-averaged values as entered in the ANOVA analyses.

[3] On March 7, 2005, an email was sent to all the members of the ACT-R community with a web-link to the prediction. We started the experiment the week after that.

Figure Captions

Figure 1. The upper panel (a) depicts the pacemaker/accumulator version of the internal clock model; the lower panel (b) depicts the attentional counter model.

Figure 2. Distribution of estimates of intervals of 8, 12 and 21 seconds (Rakitin et al., 1998). Vertical lines indicate 8, 12 and 21 seconds. The solid lines are ACT-R model fits that will be discussed later.

Figure 3. (a) Results of Zakay's (1993) experiment (bars are standard deviations) and ACT-R model fits that will be discussed later, and (b) a depiction of the attention counter theory explanation.

Figure 4. Illustration of perceiving and reproducing a time interval

Figure 5. Comparison between model and data in three bisection experiments.

Figure 6. The dual timing task (a) a screenshot with the Addition task (b) example of a single trial with the Letter task.

Figure 7. Average performance scores and standard errors for the four conditions over three phases in Experiment 1. See text for more information.

Figure 8. Distribution of first-click times in two subsequent phases for both tasks. The dashed line is the expected distribution for a pure interval estimation experiment.

Figure 9. Proportions of non-responses. Bars indicate standard errors.

Figure 10. Distribution of first-click times in phase two and three for the AAL and LLA conditions. The dashed line represents the expected distribution if this was a pure interval estimation experiment (c.f., Figure 2).

Figure 11. Average click time in the LLA condition (solid line) ten trials before and after the change in task difficulty (indicated by the vertical dashed line). For comparison the AAA condition is plotted with a dashed line.

Figure 12. Proportion dual tasking in the three phases for four conditions. Bars are standard errors.

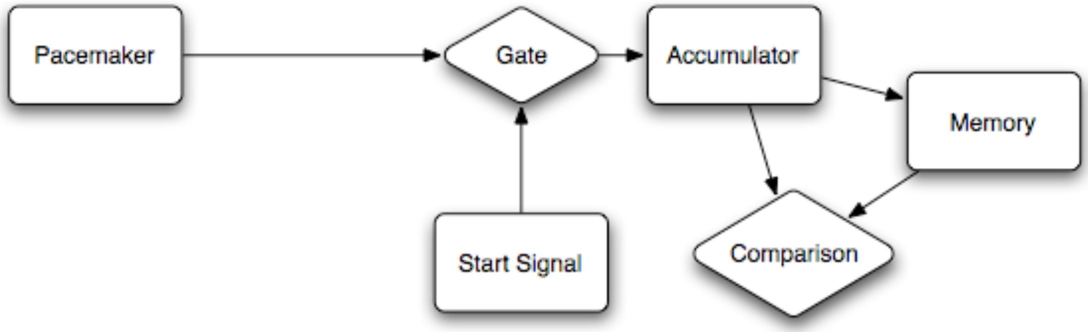
Figure 13. Interval timing as part of the ACT-R architecture

Figure 14. Average performance scores and standard errors for the four conditions over three phases in Experiment 1. See text for more information.

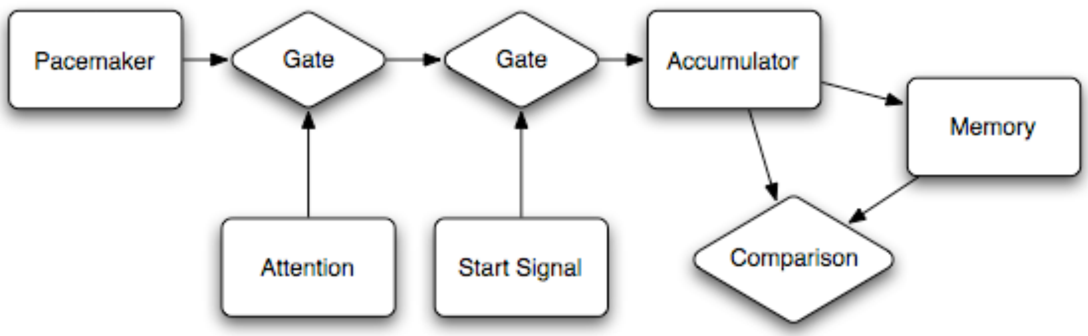
Figure 15. Distributions of the moment of the first click on the test button in the first phase of the experiment

Figure 16. Proportion of non-response by condition and block.

Figure 17. Changes in click time distributions for the conditions in which there is a change in task.



(a) Pacemaker/Accumulator Internal Clock Model



(b) Attentional Internal Clock Model

Figure 1.

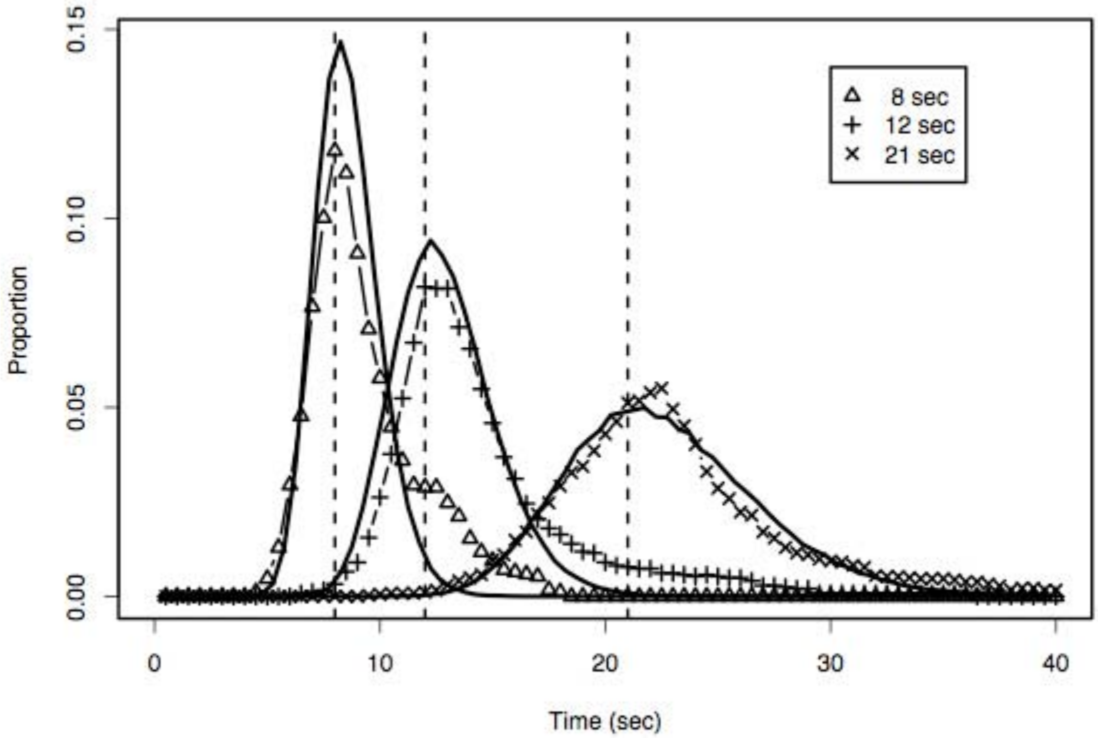
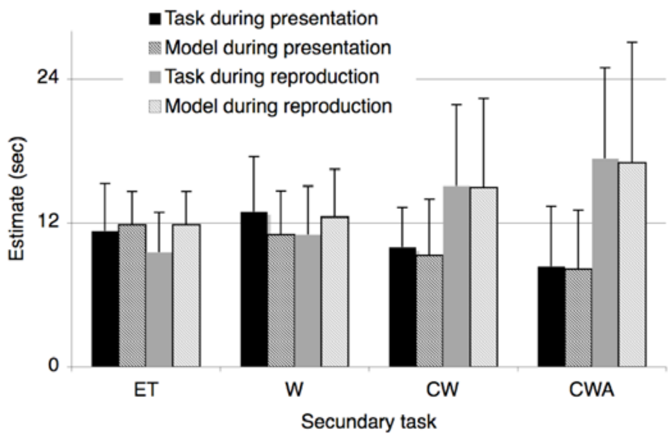
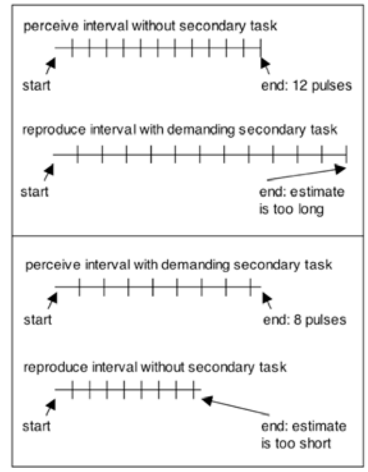

0.15
0.10
0.05
0.00
Proportion
0
10
20
30
40
Time (sec)
8 sec
12 sec
21 sec

Figure 2.


Task during presentation
Model during presentation
Task during reproduction
Model during reproduction
24
12
0
Estimate (sec)
ET
W
CW
CWA
Secundary task

(a)


perceive interval without secondary task
start
end: 12 pulses
reproduce interval with demanding secondary task
start
end: estimate is too long
perceive interval with demanding secondary task
start
end: 8 pulses
reproduce interval without secondary task
start
end: estimate is too short

(b)

Figure 3.

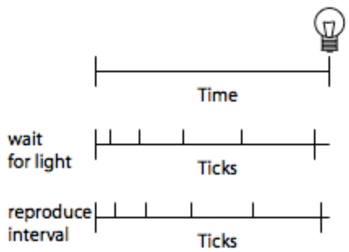


Figure 4.

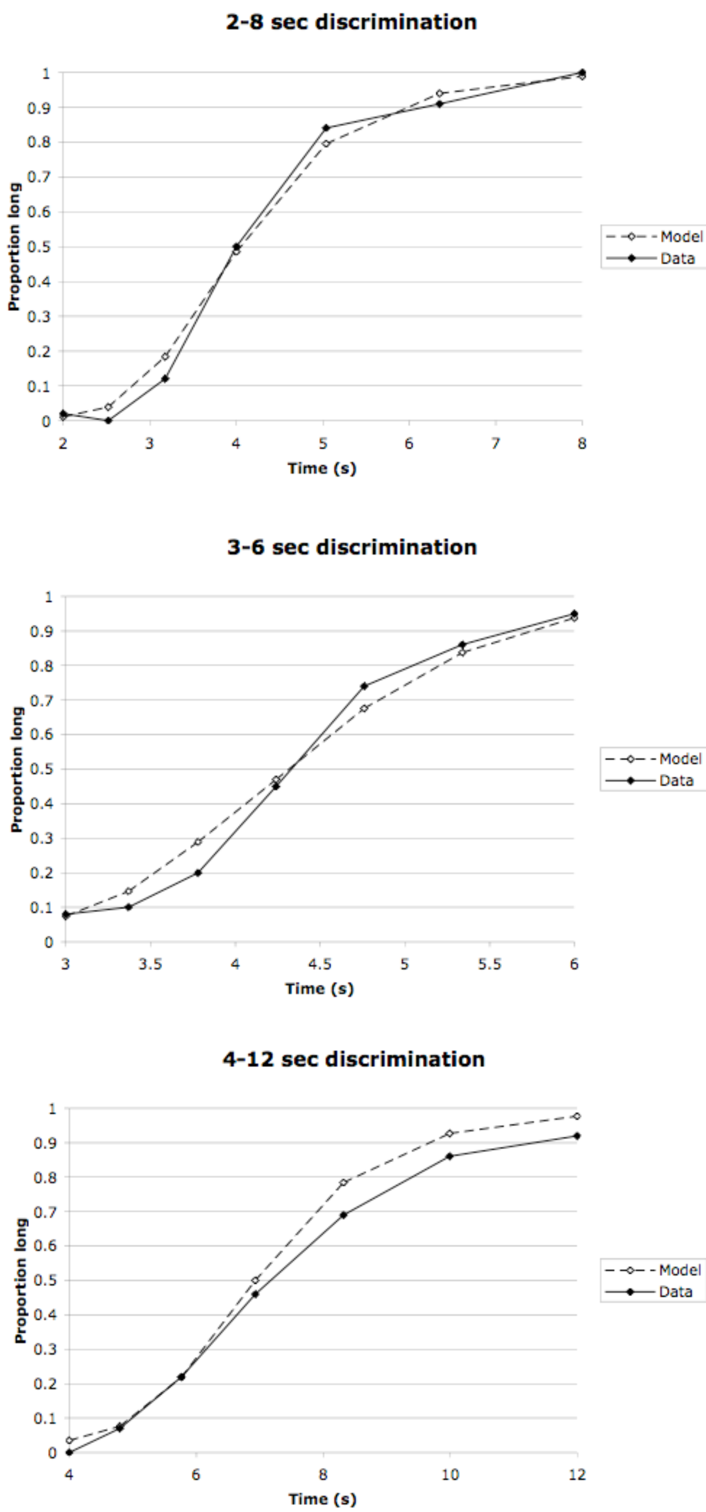
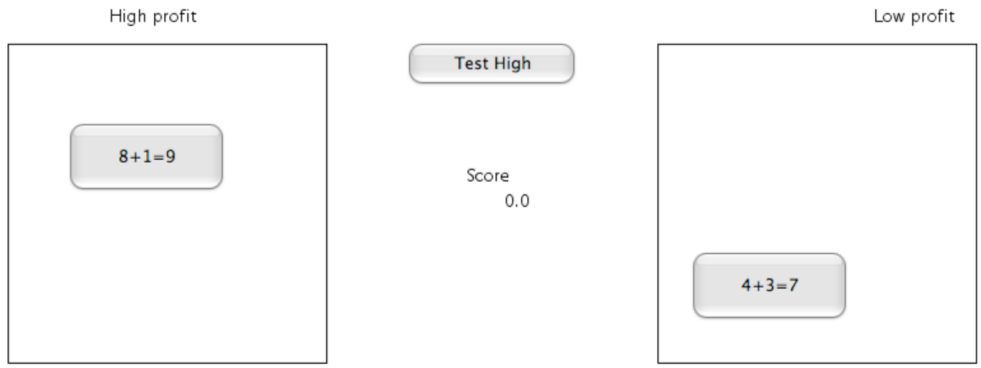
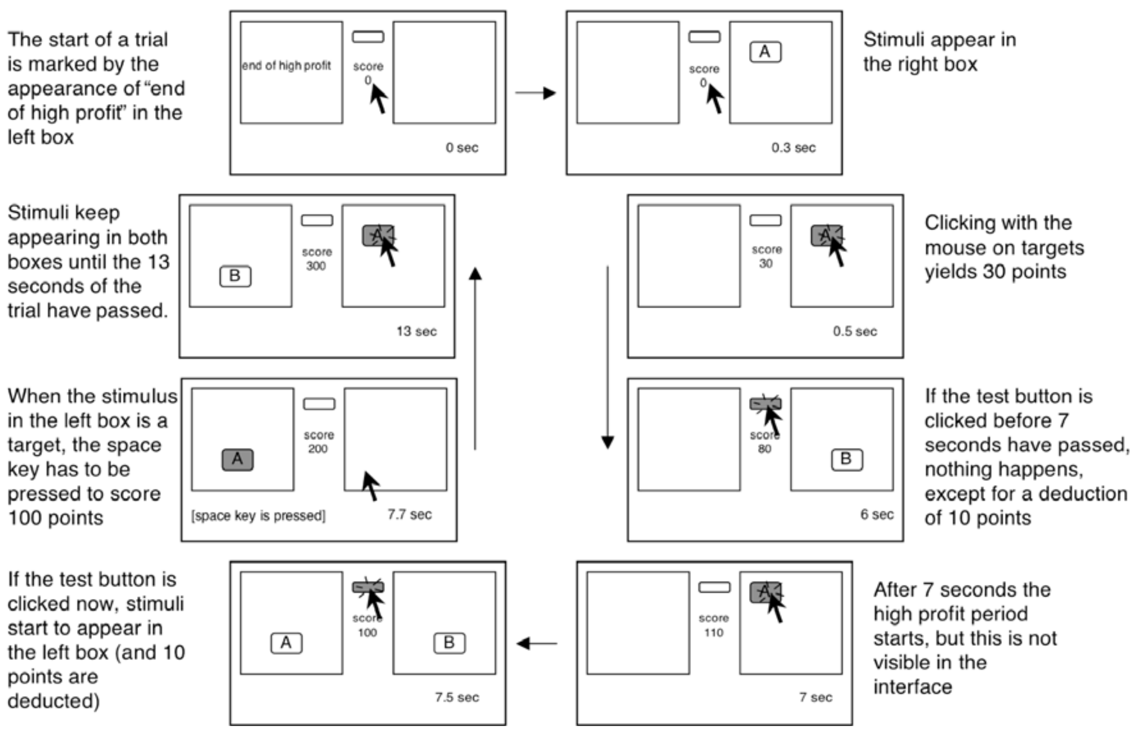

2-8 sec discrimination
Proportion long
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
3
4
5
6
7
8
Time (s)
Model
Data
3-6 sec discrimination
Proportion long
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
3
3.5
4
4.5
5
5.5
6
Time (s)
Model
Data
4-12 sec discrimination
Proportion long
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4
6
8
10
12
Time (s)
Model
Data

Figure 5

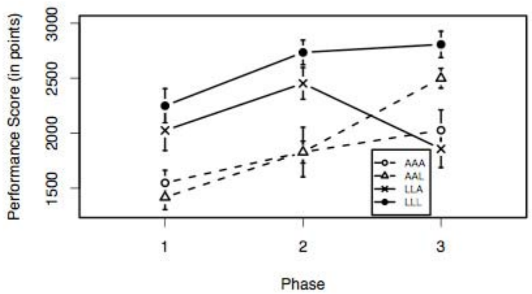

High profit
Low profit
Test High
8+1=9
Score
0.0
4+3=7

(a)

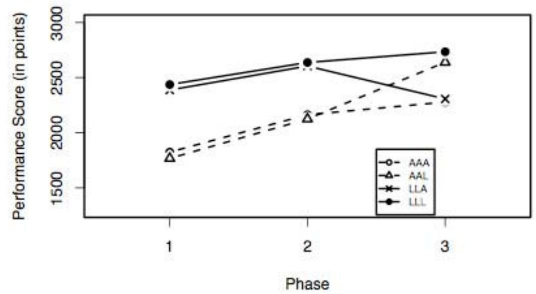

The start of a trial is marked by the appearance of "end of high profit" in the left box
end of high profit
score
0
0 sec
Stimuli appear in the right box
A
0.3 sec
Clicking with the mouse on targets yields 30 points
score
30
0.5 sec
If the test button is clicked before 7 seconds have passed, nothing happens, except for a deduction of 10 points
score
80
B
6 sec
After 7 seconds the high profit period starts, but this is not visible in the interface
score
110
7 sec
If the test button is clicked now, stimuli start to appear in the left box (and 10 points are deducted)
score
100
A
B
7.5 sec
When the stimulus in the left box is a target, the space key has to be pressed to score 100 points
A
score
200
[space key is pressed]
7.7 sec
Stimuli keep appearing in both boxes until the 13 seconds of the trial have passed.
B
score
300
13 sec

(b)

Figure 6.

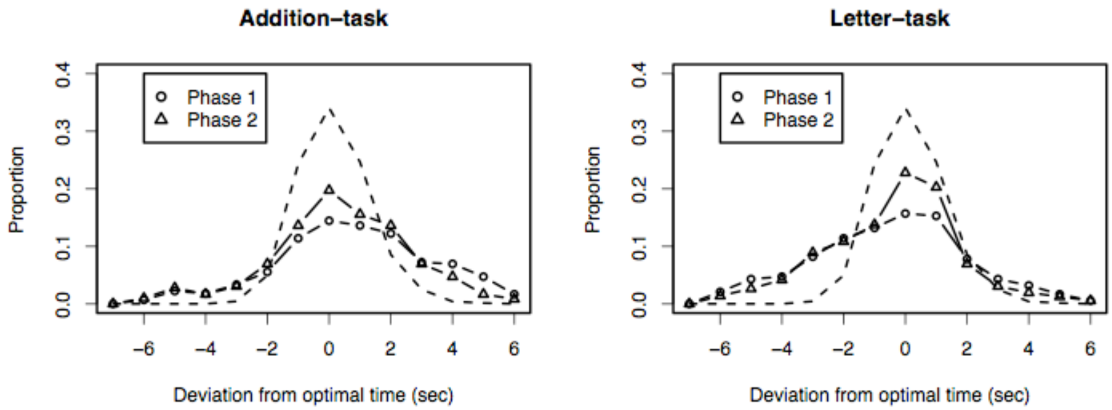


(a) Empirical data.

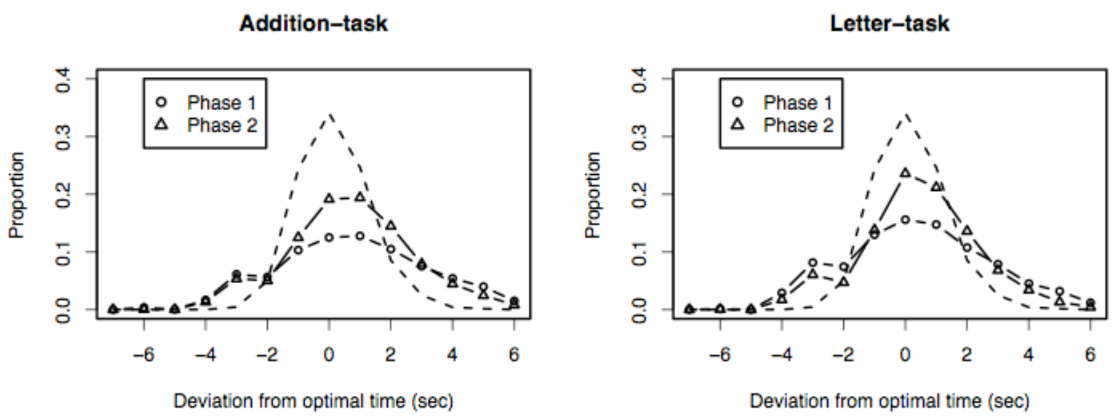


(b) ACT-R model fit

Figure 7.

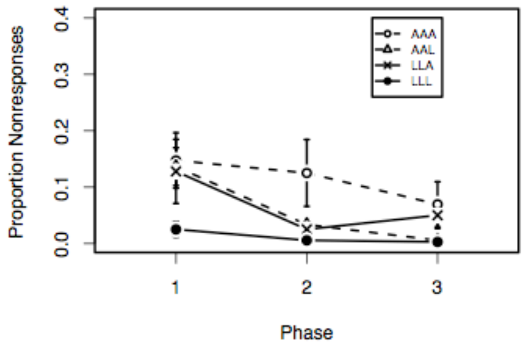


(a) Empirical data

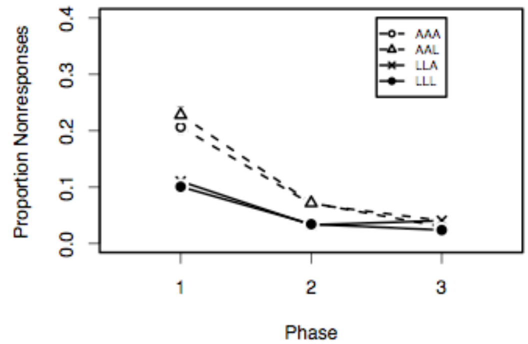


(b) ACT-R model fit

Figure 8.

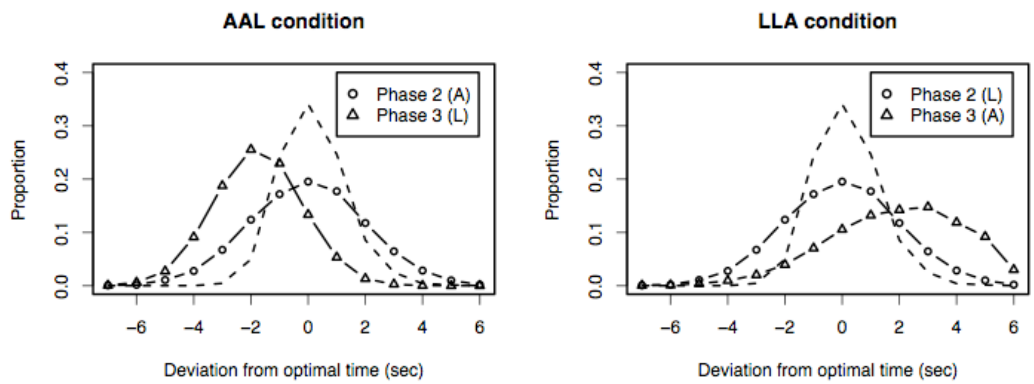


(a) Empirical data

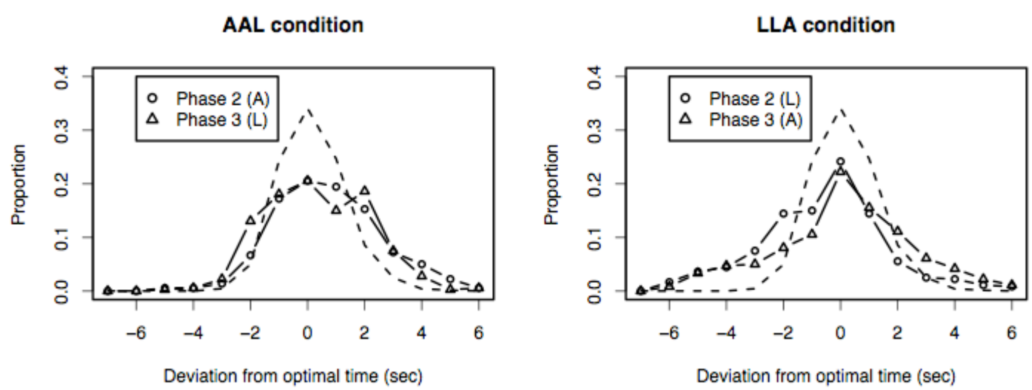


(b) ACT-R model fit

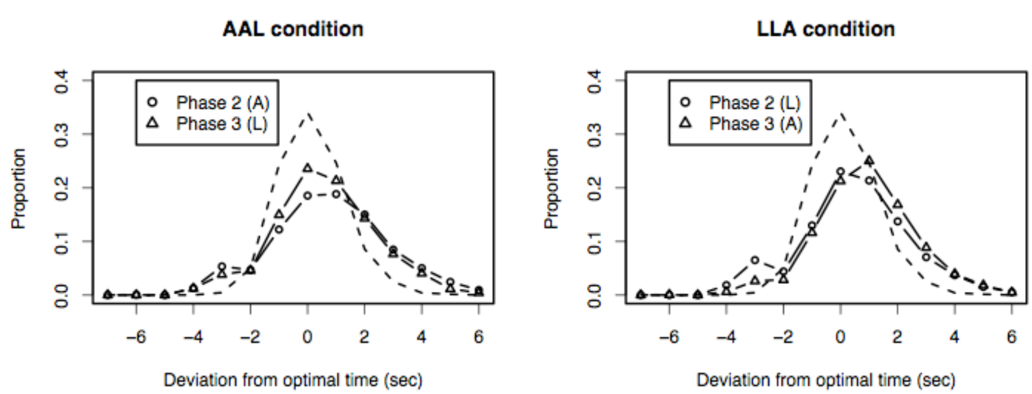
Figure 9.



(a) Attentional counter theory-based prediction



(b) Empirical data



(c) ACT-R model fit

Figure 10.

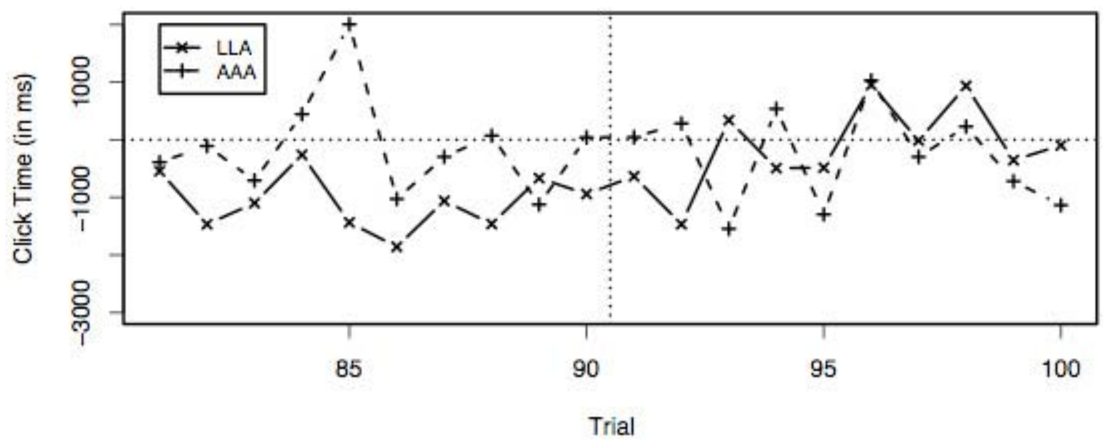
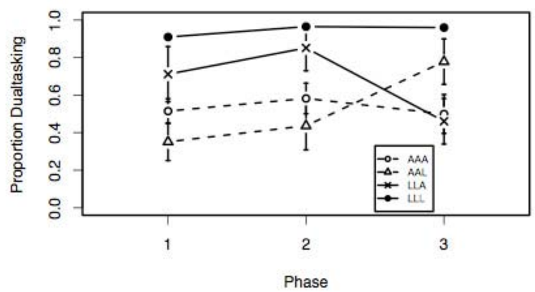
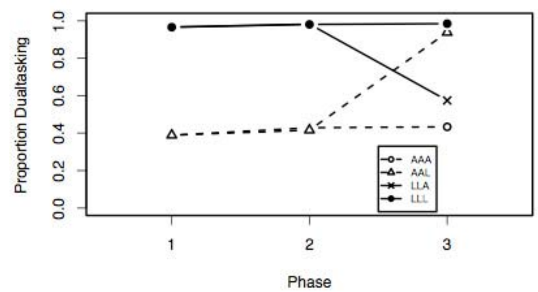


Figure 11.



(a) Empirical data



(b) ACT-R model fit

Figure 12.

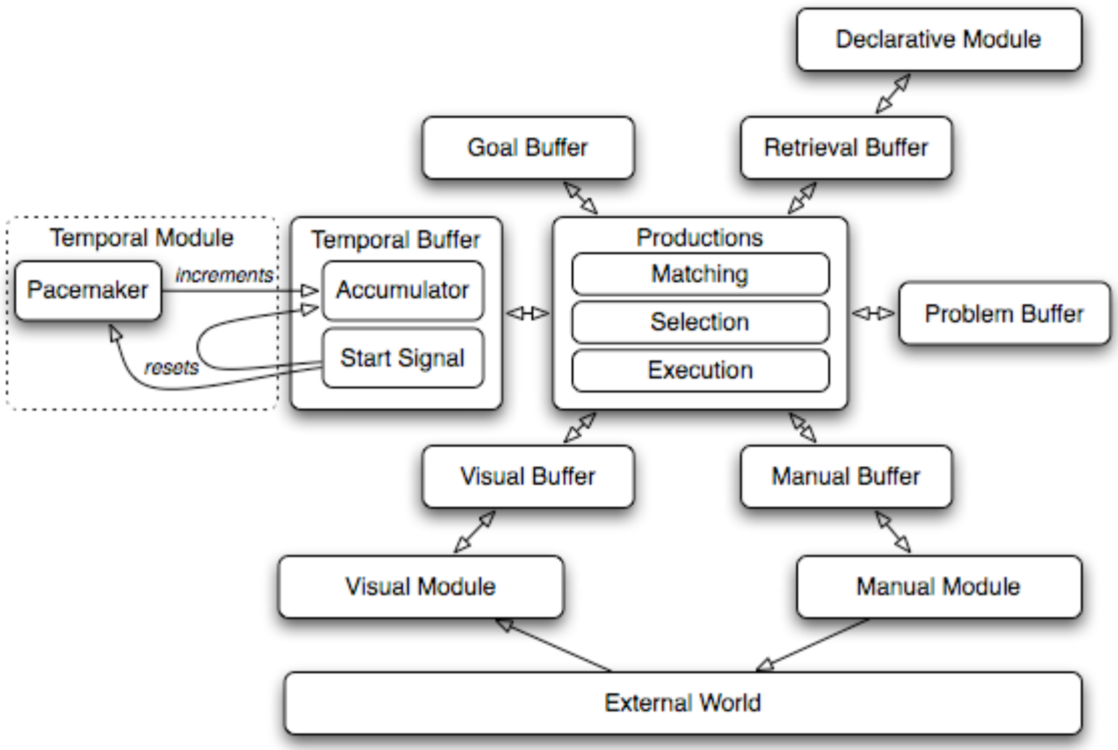
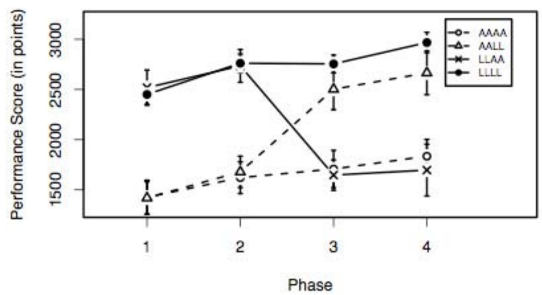
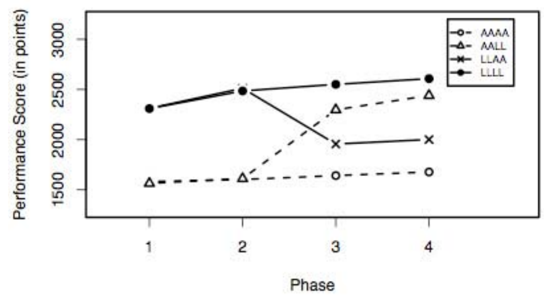

Declarative Module
Goal Buffer
Retrieval Buffer
Temporal Module
Pacemaker
increments
resets
Temporal Buffer
Accumulator
Start Signal
Productions
Matching
Selection
Execution
Problem Buffer
Visual Buffer
Manual Buffer
Visual Module
Manual Module
External World

Figure 13.

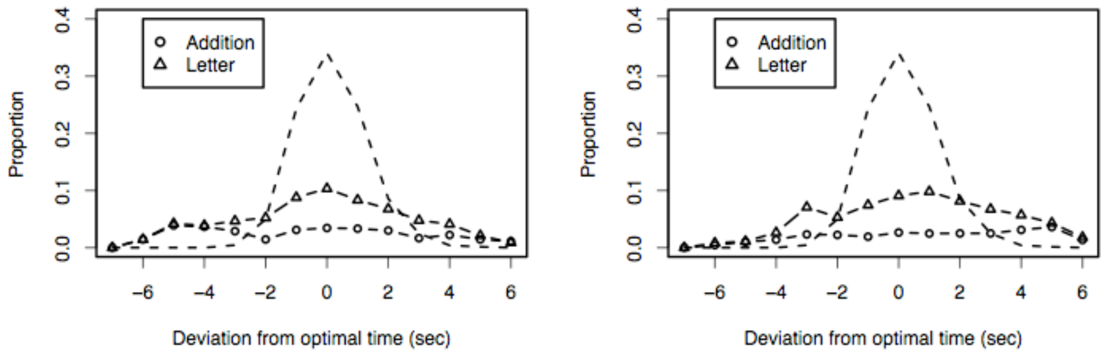


(a) Empirical data



(b) ACT-R model prediction

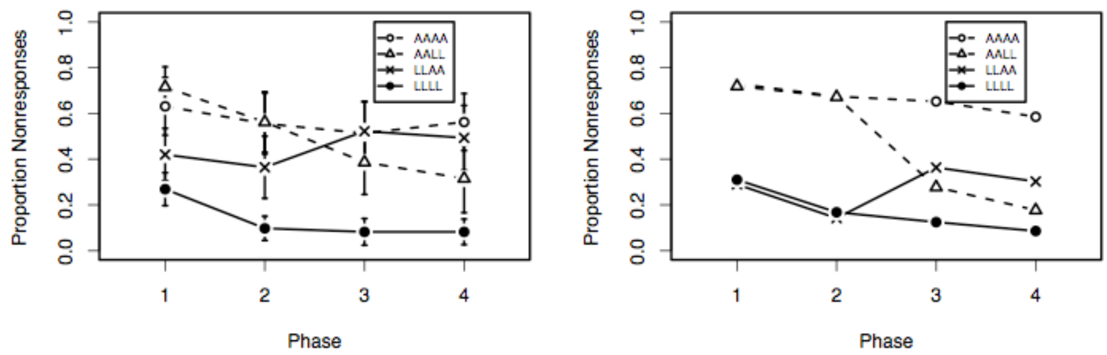
Figure 14.



(a) Empirical data

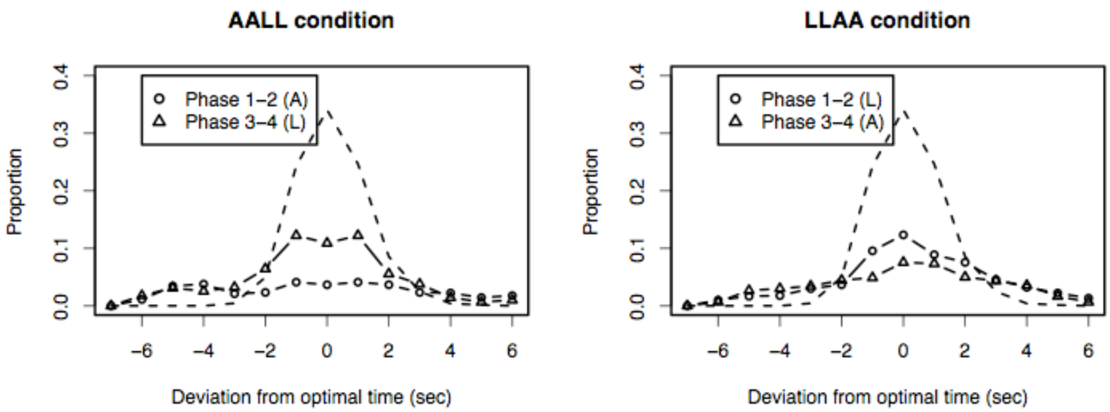
(b) ACT-R model prediction

Figure 15.

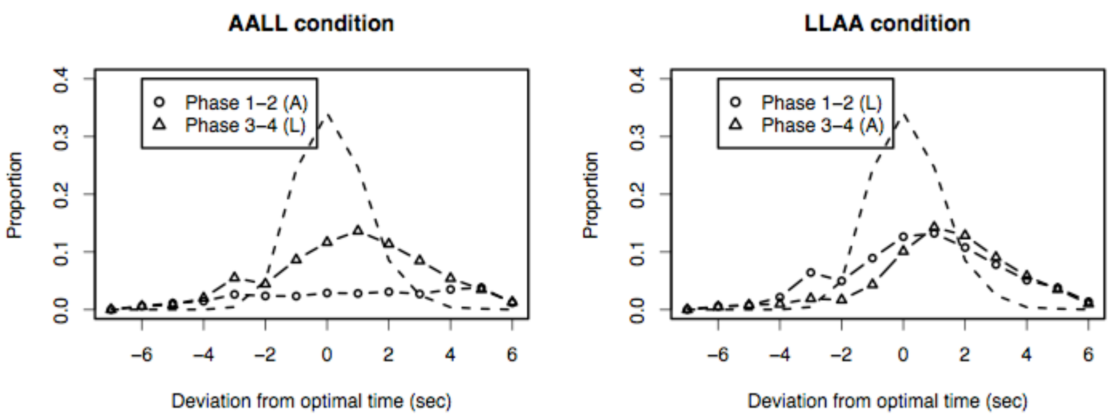


(a) Empirical data (b) ACT-R model prediction

Figure 16.



(a) Empirical data



(b) ACT-R model predictions

Figure 17.