# Traces of Times Past: Representations of Temporal Intervals in Memory

Niels Taatgen and Hedderik van Rijn

University of Groningen, the Netherlands

Please address correspondence to:

Niels Taatgen

University of Groningen

Department of Artificial Intelligence

P.O. Box 407

9700 AK Groningen

Netherlands

Telephone: +31 50 363 6435

Email: n.a.taatgen@rug.nl

Running head: Traces of Times Past

Word count:

| | |
|---|---|
| Abstract | 181 words |
| Main body | 7604 words |

Abstract

Theories of time perception typically assume some sort of memory to represent time intervals. This memory component is typically underdeveloped in theories of time perception. Following earlier work that suggests that representations of different time intervals contaminate each other (Grondin, 2005, Jones & Wearden, 2004) an experiment was conducted in which subjects had to alternate in reproducing two intervals after which they were given feedback. In two of the conditions of the experiment, the duration of one of the intervals changed over the experiment, forcing subjects to adjust their representation of that interval, while keeping the other constant. The results show that adjustment of one interval carries over to the other interval, indicating that subjects are not able to completely separate the two representations. We propose a temporal reference memory that is based on existing memory models (Anderson, 1990). Our model assumes that the representation of an interval is based on a pool of recent experiences. In a series of simulations, we show that our pool model fits the data, while two alternative models that have previously been proposed do not.

Keywords: Time perception, Memory, Cognitive model, Temporal Reference Memory, Feedback

# Traces of Times Past: Representations of Temporal Intervals in Memory

Although there are many theories and models of how people estimate time intervals, they typically contain three components: a clock component, a memory component, and a comparison component (e.g., Church, 1984; Gibbon & Church, 1984; Michon, 1967; Treisman, 1963). A primary topic of debate among these theories is the nature of the clock component. Some theories assume that the clock represents time linearly, like the Scalar Expectancy Theory (SET, Gibbon, 1977, 1991), whereas others propose a nonlinear representation (e.g., Staddon & Higa, 1999; Taatgen, Van Rijn & Anderson, 2007). Another controversy related to the clock is whether attention is necessary for accurate performance of the clock: according to the attentional-gate models (for a description, see Zakay & Block, 1996), attention is necessary for the clock itself to function, while other models explain the effects of divided attention by other means (Lejeune, 1998; Taatgen, Van Rijn, & Anderson, 2007). Apart from these studies that focus on separate components, a large body of research has been devoted to unraveling the boundaries and relations between the different components. For example, although imaging studies (e.g., Lewis & Miall, 2006) but also clinical studies and pharmacological manipulations have shown that the clock and the memory component have independent biological substrates (for a review, see Buhusi & Meck, 2005), many studies have shown that both systems are intimately tied together to produce accurate time estimations. In these studies, it is typically assumed that the memory system contains information that reflects earlier temporal experiences that are matched to the current state of the clock. This memory system is therefore often referred to as temporal reference memory. When the reference reflecting a previous experience matches the current state, the system knows that the same amount of time has passed.

The last years have seen an increased interest in the nature of the memory component. One line of research that can be identified is focused on the memory representations themselves. A prime example of this work is a series of studies by Jones and Wearden.

Jones and colleagues tested whether multiple presentations of an interval improved temporal reproduction (Jones & Wearden, 2003; Ogden & Jones, 2009), whether performance was degraded when multiple durations had to be learned and kept active simultaneously (Jones & Wearden, 2004), and whether memory traces are modality independent (Ogden & Jones, 2009; Ogden, Wearden & Jones, 2010). We will discuss two of these studies in more detail.

In the studies reported in Jones and Wearden (2003), participants were presented 1, 3 or 5 examples of a standard duration, and had to judge whether later presented stimuli were equal in duration to the presented standard. The number of presentations of the standard never affected performance, yielding the surprising conclusion that an increased number of presentations of an item does not improve or affect later performance. Jones and Wearden constructed a number of computational models, which varied in the way the presented standards were stored and retrieved. Based on these simulations, they concluded that the best model was one in which a single memory trace represents the standard. If, on later trials, another estimation of the standard is obtained that deviates more than a certain preset value, then there is perturbation of the trace and this new estimation replaces the old. Although this perturbation model explains the original data quite well, the generality of this model can be questioned as (1) it is difficult generalize this model to other phenomena related to temporal reference memory and (2) this model assumes an atypical memory system.

One set of results that is difficult to explain with the simplest form of the perturbation model is presented by Jones and Wearden themselves. In their experiment (Jones and Wearden, 2004) subjects had to judge whether presented intervals were the same or different from a previously learned interval. In the standard condition, they just learned a single interval, but in the double condition they learned two intervals, and were only told after the presentation of the test interval with which of the two it had to be compared. Performance in the double condition was poorer, probably because the irrelevant interval interfered with the relevant interval. However, according to the perturbation model as presented in Jones and Wearden (2003), the only possible explanation of this behavior is that there is a complete confusion of the two intervals in that instead of using the relevant interval, the irrelevant interval was used. The behavioral results are not in support of this but instead suggest that the presentation results in a more continuous blend of the two standards.

Grondin (2005) carried out a similar experiment, in which subjects had to judge whether presented intervals were longer or shorter than a previously learned interval. In one condition of the experiment, subjects had to learn two intervals, 250ms and 750ms, and were told before each trial whether to base their judgment on the shorter or the longer trial. Results indicated that subjects tended to shift their representations of the 250ms and the 750ms interval towards each other, as if they contaminated – instead of replaced – each other in memory. In our own earlier study in which subjects performed mental calculations on 2- and 3-second intervals we had noticed a similar effect: estimates of the 2-second interval tended to be long, and estimates on the 3-second interval tended to be short (Van Rijn & Taatgen, 2008).

All these studies support a more continuous view of the temporal reference memory system. Here, it is important to note that Jones and Wearden themselves have argued that their perturbation model as sketched here and in their 2004 paper is probably too simple, and that instead of completely replacing the old value, a more gradual change might fit the data better.

However, question is whether designing a temporal reference memory system from scratch is to be preferred over using existing memory models to explain memory phenomena in time perception. Especially since a number of studies have provided evidence that there is an intimate link between the temporal reference memory system and more general memory phenomena. For example, Brown (1997) has shown that if a secondary task is presented during temporal reproductions, performance on the secondary task is negatively affected if this task requires working memory and Fortin, Champagne and Poirier (2007) have shown that when a concurrent memory task is performed during time estimation, the temporal estimates are stronger influenced if the concurrent task requires order judgments. Similarly, Baudouin, Vanneste, Isingrini and Pouthas (2006) have shown in a study with older adults that temporal reproduction is correlated to working memory capacity. These studies are exemplary for many studies that have shown evidence for the notion that a single memory system is involved in both memory tasks and time estimations.

In our view, it is therefore desirable to use general models of memory as the memory component for models of time perception. This is what we proposed in our own model of time perception (Taatgen et al., 2007): instead of using specialized mechanisms for attention, memory and comparison, we used mechanisms from the more general ACT-R cognitive architecture (Anderson, 2007). In the experimental work that supported our model, memory failure was one of the mechanisms to explain a breakdown of time perception in complex situations. In this article we want to further develop the memory component of time perception by focusing on the issue of how representations of time intervals are learned and represented, and, in the case of multiple intervals, how representations influence each other.

*Representations of Time*

In order to explain how representations of time intervals affect each other, we have to ask ourselves the question how solid the memory representations of time intervals are. One explanation is that over the course of an experiment solid representations of the two intervals are formed, but that the formation of these representations is influenced by the fact that another interval has to be learned at the same time. Another explanation is that solid memories never form[1]. This distinction is analogous to the discussion in memory theories, in which some theories hold that each presentation is stored and retrieved separately (e.g., Landauer, 1975) whereas others propose that additional presentations strengthen a single, more general memory trace, which will later be retrieved (e.g., Bower, 1961, Raaijmakers & Shiffrin, 1981).

The distinction between a solid memory representation for both durations on the one hand, and a more distributed approach on the other, does not need to be problematic for the comparison process, if one assumes that both approaches eventually result in the retrieval of a single representation that can be compared to the current clock value. This brings us to the question how this single retrieved representation is constructed.

One of the computational models explored by Jones and Wearden (2003) is the sampling strategy (abbreviated to SAM). The SAM mechanism assumes that all representations in memory are equally likely to be retrieved, and that just one of these is sampled and used in subsequent timing processes. However, as Jones and Wearden (2003) themselves argue, this makes it difficult to account for the flexibility observed when learners are confronted with a changed standard time. As, according to their theory, all representations are defined as a constant value plus a Gaussian distributed noise with a fixed standard deviation and all representations have an equal chance of being retrieved, sampling a single value from memory is very similar to just using the last perceived value. On the basis of this reasoning, Jones and Wearden proposed the perturbation model of temporal reference memory. Although this perturbation model can potentially account for a broad range of phenomena (see also Ogden & Jones, 2009), it is difficult to envision how the simpler version of this model could explain the contamination phenomena as discussed earlier.

The alternative strategy, referred to as the averaging (AVE) strategy, presented by Jones and Wearden (2004) is similar to a single, solid memory representation account. AVE assumes that all previous experiences with certain duration are averaged to a single value that is used in subsequent comparison processes. A difficulty of this strategy is that a lot of parameters are underspecified. For example, how many past experiences are involved in the averaging process? It cannot be all experiences, since if one assumes that *all* experiences with a certain interval are averaged, it is again difficult to account for the flexibility associated with sudden changes in standard times.

All three accounts assume a relative perfect memory system. Experiences that are stored are not subjected to decay, interference, or any other influence that has been identified in memory literature. Thus, although the SAM, AVE and perturbation models can probably be extended to account for (some or most) new phenomena, another approach would be to let go of the assumption of a perfect memory system, and instead rely on well established memory concepts to explain temporal behavior. We propose that the retrieved representation is constructed on the basis of a pool of previous experiences (an idea that is central in many fields of cognitive science, e.g., Tenenbaum & Griffiths, 2001; Pothos & Chater, 2002) in which recent experiences have a much stronger weight than older experiences. This pool of experiences may be polluted by experiences with other intervals, creating the biases found in experiments dealing with multiple intervals.

To test our hypothesis we designed an experiment in which subjects had to learn and reproduce two intervals. While reproducing the intervals, they received accuracy feedback. We used two methods to assess how intervals are represented. The first is to analyze the reproductions of the two intervals on a trial-by-trial basis. This allows us to inspect the changes of temporal estimations in much more detail than if one just compares average performance, as was done in earlier work. Hereto, we calculated a regression equation to predict a duration for each reproduction ($t$) on the basis of a fixed intercept, the durations of recent reproductions ($t-n$) and feedback on those reproductions ($f_{t-n}$). If solid memories are formed, the durations and associated feedback should have only a very limited influence on the prediction of the next reproduction. Therefore, the regression equation's intercept should be close to the interval to be estimated. However, if the reproduction is based on a pool of experiences, we expect the intercept to be rather small, with larger influences for recent experiences.

For the second method to assess the nature of the memories for time intervals we introduced an experimental manipulation that forced subjects to gradually change the representation of one of the intervals (the longer interval in our experiment). Given that the feedback is given on the basis of the changed duration, participants will have to adjust their internal representations to remain at a reasonable level of performance. Although the solid memories and pool of memories explanations do predict some differences in behavior (e.g., more sudden transitions from one type of performance to another as described in Van Rijn, Van Someren and Van der Maas, 2005), these differences are so subtle that their signatures are most likely lost in the inherent noise associated with temporal reproductions (Wagenmakers, Farrell & Ratcliff, 2004). However, as the experimental manipulations are only introduced after a number of trials in the experiment, there is no reason why changing the baseline of the long interval should influence the short interval if one assumes solid and independent representations for the two intervals. On the other hand, if representations of time intervals were more the result of a set of experiences, we expect that a change in one of the intervals will spill over in the other interval, something that one would not expect if the intervals were solidly represented.

In the rest of this article we will first elaborate on the experiment and the analysis of the results of that experiments. We will then proceed with a model of these results that is based on the declarative memory theory of the ACT-R architecture (Anderson, 1990, 2007) with a modification that allows it to deal with real values (Lebiere, Gonzalez, & Martin, 2007).

# Method

*Participants*

Seventy students from the university of Groningen participated in this study for course credit. Twelve subjects were removed from the pool because more than 3% of their responses were shorter than 1.25s or longer than 4.25s. Of the remainder, ten subjects were not able to distinguish between the two intervals after training, and were therefore also removed from the dataset. We will reconsider these 10 subjects in the discussion. The remainder, 48 subjects (16 per condition detailed below), had an average age of 20.3 and consisted of 10 men and 38 women. The number of subjects dropping out was fairly high, but this is to be expected given that experiments with multiple time intervals often observe fairly high error rates (e.g., Brown & West, 1990, Meijering & Van Rijn, 2009, Wearden, 2002).

*Design and Procedure*

In the experiment, subjects learned two intervals, a short interval of 2 seconds, and a long interval of 3.1 seconds, which they had to reproduce repeatedly, always alternating between the short and the long. Subjects were presented with two circles on the screen, which were gray when they were not active. The circle on the right of the screen was associated with the 2-second interval, while the circle on the left was associated with the 3.1-second interval. During training, one of the circles would change color (blue for the short interval and green for the long interval) for a specific duration, and would then turn back to gray. After the presentation of the standard interval, participants had to reproduce the temporal interval. The onset of the interval was machine paced, indicated by the gray circle turning blue of green again, and subjects had to press a key to indicate the end of the interval ("f" for the long interval, and "j" for the short interval). Subjects received feedback on the accuracy of their produced intervals (we will refer to them as *estimates* from here on): "too short" if they responded earlier than 87.5% of the interval, "too long" if they responded later than 112.5% of the interval, or "correct" otherwise. Training consisted of 10 presentation-reproduction-feedback trials, alternating between the two durations.

The presentation phase was removed from each trial in the experimental block, but all other aspects were kept the same. Subjects received 15 warm-up trials of each duration, alternating between them, followed by the experiment proper.

The main experimental manipulation is that the criterion for the long interval shifts in two of the three between-subject conditions. In the FF (flat-flat) condition, the criterion remains the same for the rest of the experiment (185 estimates for each interval). In the DR (dike-river, referred to as such since the graphical depiction of the standard follows a typical riverbed with dike outline, see the dotted lines in Figure 1) condition, the criterion remains at 3.1 seconds for the first 25 estimates. After that, the criterion is increased linearly to 3.6 seconds over 15 estimates. This means that at some point subjects received “too short” feedback for a duration that previously was correct. After the shift to 3.6 seconds, the criterion stayed at 3.6 seconds for 25 estimates, then decreased back linearly to 3.1 seconds over 15 estimates, stayed there for another 25 estimates, then decreased further to 2.6 seconds over 15 trials, stayed at 2.6 seconds for 25 trials, increased back to 3.1 seconds over 15 trials and stayed there for the remaining 25 estimates. Meanwhile, the criterion for the short interval (remember that short intervals and long intervals were alternated) remained constant at 2 seconds. The RD (river-dike) condition was the exact opposite of the DR condition: instead of increasing the interval after 25 estimates, the criterion would first decrease to 2.6 seconds, leading to the sequence of: 25 at 3.1s – 15 decreasing to 2.6s – 25 at 2.6s – 15 increasing to 3.1s – 25 at 3.1s – 15 increasing to 3.6s – 25 at

3.6s – 15 decreasing to 3.1s – 25 at 3.6s. Graphical depicts of all three conditions are presented as the dotted lines in Figure 1.

## Results

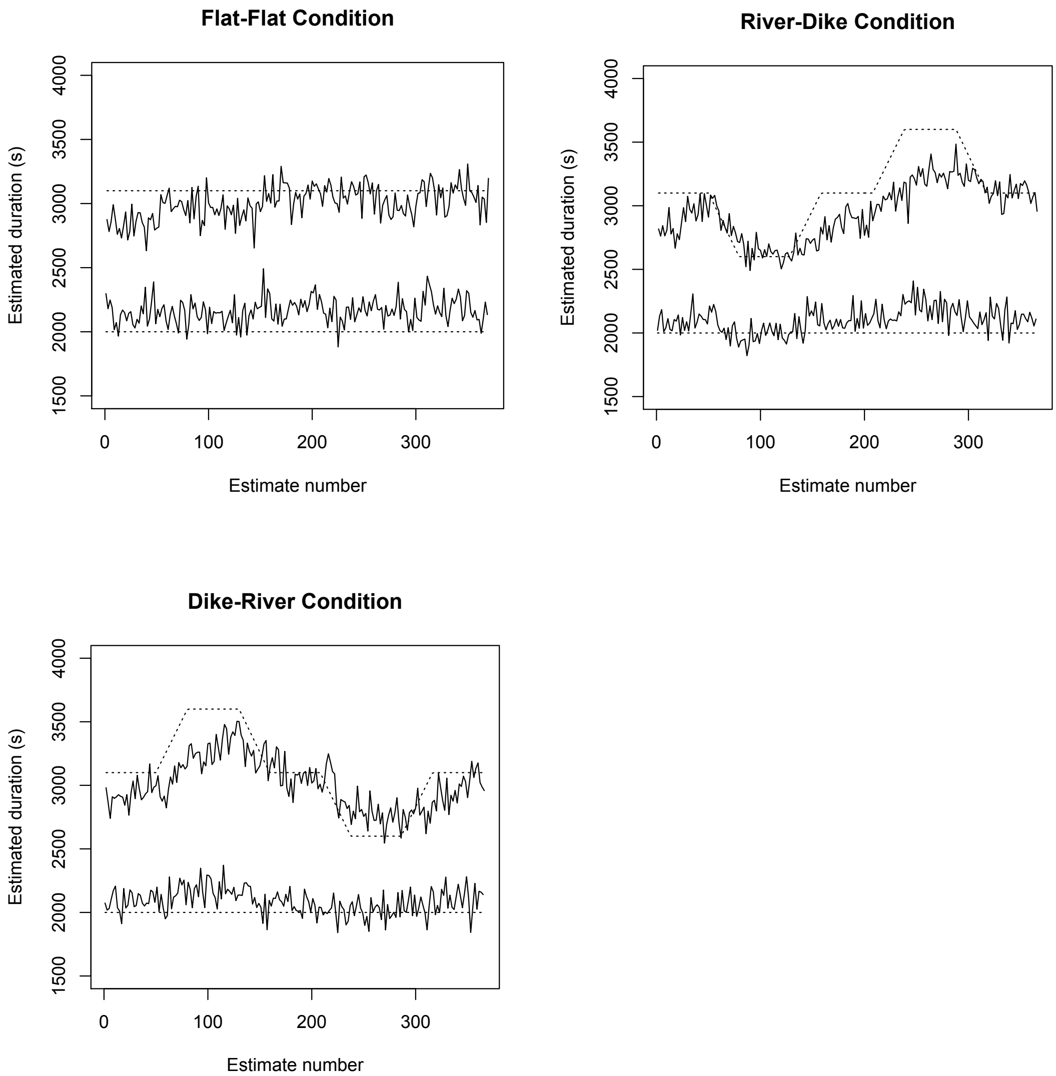


Figure 1. Mean estimates for the three conditions in the experiment. The lower line in each graph represents the estimates for the short interval, while the upper line represents the long interval. The dotted lines indicate the criterion, which is always constant at 2s for the short interval, but which changes for the long interval in the River-Dike and Dike-River conditions.

Figure 1 shows the mean estimates over the course of the experiment. Visual inspection of the results suggests indeed that in the FF condition the short interval is estimated longer, and the long interval shorter, consistent with earlier findings (e.g., Grondin, 2005), and suggesting both estimates influence each other. More pronounced are the results in the other two conditions, because there is a clear suggestion that the estimates of the short interval are influenced by changes in the duration of the long interval as the short interval's estimations resemble a dampened pattern of the long interval. The overall accuracy (i.e., the proportion of estimates within 12.5% of the target interval) is 55% for short intervals and 63% for long intervals, with only small differences between conditions (short: FF 52%, RD 59%, DR 55%; long FF 63%, RD 66%, DR 61%). Apparently, the criterion manipulation for the long interval is slow enough to not affect accuracy.

In order to analyze the influence of one interval on the other, we have to acknowledge that two factors determine the next estimate that a subject makes: the representation of that interval in memory, and the feedback given by the experiment (too short, correct or too long). On top of that, both the representation of the other interval, as well as the feedback for the other interval can influence the estimate. Figure 2 illustrates these factors: at the right side of the figure a short interval has to be estimated by the participant, which may be influenced by previous estimates and previous feedback of both intervals.

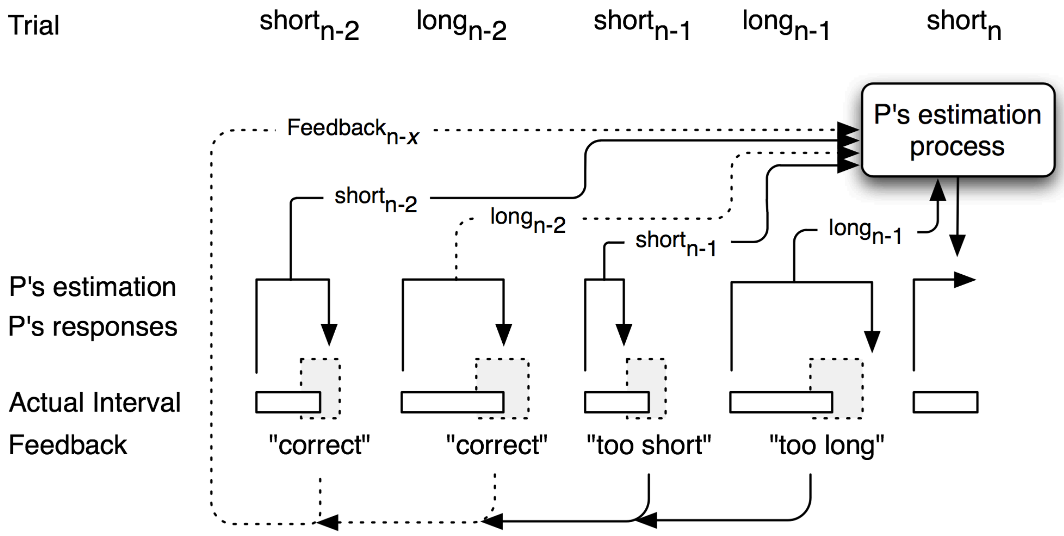


Figure 2. Factors that may impact the estimation process. In this example, estimation of a short interval is shown. P stands for "participant", and the gray areas indicate the intervals in which the participant received "correct" as feedback. The factors with a solid arrow will turn out to have a significant impact, the dashed arrows not.

Table 1. Results of fitting mixed-effect models to the estimates of the short interval in the three conditions. *p<0.05, **p<0.01, ***p<0.001.

| | FF condition | | RD condition | | DR condition | |
|---|---|---|---|---|---|---|
| Factor | Beta (SE) | t | Beta (SE) | t | Beta (SE) | t |
| Intercept (ms) | 1217 (120) | 10.1*** | 962 (93) | 10.3*** | 826 (92) | 9.0*** |
| **Impact of previous short intervals on current short interval** | | | | | | |
| short$_{n-1}$ | 0.18 (0.031) | 5.7*** | 0.17 (0.03) | 4.9*** | 0.28 (0.03) | 8.6*** |
| short$_{n-2}$ | 0.079 (0.018) | 4.3*** | 0.087 (0.018) | 4.8*** | 0.11 (0.02) | 6.3*** |
| short-fb$_{n-1}$ = "too long" (ms) | -106 (25.8) | -4.1*** | -96.8 (22.9) | -4.2*** | -153 (23) | -6.7*** |
| short-fb$_{n-1}$ = "too short" (ms) | 92.0 (27.6) | 3.3*** | 48.9 (22.7) | 2.2* | 68.2 (23.9) | 2.9** |
| **Impact of previous long interval on current short interval** | | | | | | |
| long$_{n-1}$ | 0.14 (0.03) | 4.3*** | 0.21 (0.02) | 11.0*** | 0.15 (0.02) | 8.0*** |
| long-fb$_{n-1}$ = | -130 (32) | -4.0*** | -97.5 | -4.2*** | -83.8 | -3.8*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| "too long" (ms) | | | (23.1) | | (22.1) | |
| long-$fb_{n-1}$ = "too short" (ms) | 39.1 (28.5) | 1.4 | 67.2 (17.2) | 3.9*** | 69.1 (19.1) | 3.6*** |

To assess all of these factors, we have fit linear mixed-effect models to the each of the two interval durations (Baayen, Davidson, & Bates, 2006). Estimations were entered as dependent variable, previous (up to *n-10*) estimates and feedback were entered as predictors, both for the current duration and for the other duration, while allowing a random effect for participant. Linear mixed-effect models provide information about the contribution of individual factors to a dependent variable and about the reliability of the estimates. We compared more complex models (i.e., models including estimates or feedback of trials longer ago) with simpler models (i.e., models with less estimates/feedback) using the Akaike Information Criterion (AIC; Akaike, 1974) and maximum likelihood criterion as discussed in Burnham and Anderson (2002). We first compared the contribution of previous estimations of the target interval, then the contribution of feedback on previous estimations of the target interval, and then the estimations and feedback on previous estimations of the other interval. The comparisons were performed on the RD dataset, and then the same models were fit to the DR and FF datasets. We report the preferred models, meaning that increasing or decreasing the number of predictors resulted in less optimal AIC/maximum likelihood criterion scores.

Table 1 shows the results for the short interval for the best-fitting model. Let us examine the factors in the FF condition to get an idea of what this analysis means. The predicted response time for trial $short_n$ consists of a fixed intercept of 1217 ms. Added to this intercept are fractions of the previous short intervals (the betas in Table

1), so 0.18 times the previous short estimate (of approximately 2 seconds, so something in the order of 360 ms), and 0.079 times the short estimate before that. The difference between these factions (0.18 and 0.079) indicates that the influence of the *n-2* estimation attributes less to the current estimation. Added to this is also a fraction of the previous long interval: 0.14 times the previous long interval. Note that the estimation of the $long_{n-2}$ interval is not incorporated in the model since $long_{n-2}$ did not result in an improved RD fit.

Previous feedback also modifies the interval: if the feedback on the previous short interval was "too short", the estimate is increased by 92 ms, but when it is "too long" it is decreased by 106 ms. Finally feedback on the previous long intervals also impacts the predicted estimate on the short interval: 130 ms is subtracted if the feedback was "too long", and 39 ms was added if the feedback was "too short" (this last adjustment is not significant in FF, although it is in the RD and DR conditions).

The results generally support the hypothesis that the representation of an interval is the result of a pool of recent experiences, and not of a single representation. This is indicated by the relatively small intercepts in the regression formula and the susceptibility of the estimates to changes in the other interval. It is also interesting to see that the different factors are roughly the same between the three conditions, indicating that the same underlying processes might be affecting them.

Table 2 shows the analysis of the long interval. The results are similar to those of the short interval. We can see that both the estimate of the previous short interval and the feedback on that interval have an impact on the estimates of the long interval. Here the factors differ a bit more between conditions. The $long_{n-1}$ and $long_{n-2}$ factors are larger in the DR and RD conditions because they are needed to track the changing interval criterion.

Table 2. Results of fitting mixed-effect models to the estimates of the long interval in the three conditions. *p<0.05, **p<0.01, ***p<0.001.

| | FF condition | | RD condition | | DR condition | |
|---|---|---|---|---|---|---|
| Factor | Beta (SE) | t | Beta (SE) | t | Beta (SE) | t |
| Intercept (ms) | 1245 (131) | 9.5*** | 568 (98) | 5.8*** | 906 (107) | 8.5*** |
| **Impact of previous long intervals on current long interval** | | | | | | |
| $long_{n-1}$ | 0.28 (0.03) | 8.4*** | 0.50 (0.02) | 21.5*** | 0.46 (0.02) | 19.2*** |
| $long_{n-2}$ | 0.12 (0.02) | 6.5*** | 0.17 (0.02) | 10.0*** | 0.17 (0.02) | 9.4*** |
| long-$fb_{n-1}$ = "too long" (ms) | -108 (35) | -3.1** | -237 (26) | -9.2*** | -170 (26) | -6.6*** |
| long-$fb_{n-1}$ = "too short" (ms) | 114 (30) | 3.8*** | 267 (19) | 13.8*** | 220 (22.3) | 9.8*** |
| **Impact of previous short interval on current long interval** | | | | | | |
| $short_{n-1}$ | 0.25 (0.03) | 7.7*** | 0.16 (0.04) | 4.3*** | 0.10 (0.04) | 2.6* |
| short-$fb_{n-1}$ = "too long" (ms) | -87.2 (27) | -3.2** | -56.9 (25.3) | -2.2* | -74.8 (26.6) | -2.8** |
| short-$fb_{n-1}$ = "too short" (ms) | 136.3 (29.1) | 4.7*** | 80.9 (25.2) | 3.2** | 38.0 (27.8) | 1.4 |

## Cognitive Model

Although the results support the idea that the representation of time intervals involves a pool of experiences, they do not show that such an account can actually produce the behavior found in the experiment. This is why we developed a computational model, referred to as the pool model. This pool model should also show that all the factors identified in the statistical analysis can be attributed to properties of a single memory system. As indicated in the introduction, we will use the declarative memory system of the ACT-R architecture (Anderson, 1990, 2007), which has been proved to accurate in modeling many different aspects of human cognition. Instead of using the full ACT-R architecture, we have used only the declarative memory theory and the time estimation theory of ACT-R, simplifying both the model and the necessary explanation.

*Time Estimation*

The time perception component is a classical pacemaker-accumulator system in which a pacemaker generates pulses that are counted by an accumulator (Taatgen et al., 2007). The system can be given a start signal that resets the accumulator and starts the pacemaker. The accumulator therefore represents the amount of time that has passed since the start signal. Time is measured in units that start at 100 ms, but become gradually longer, creating a nonlinear representation of time. For the purposes of the present model, this nonlinearity is not very important, and qualitatively similar results could be obtained with a linear clock such as found in SET (Gibbon, 1977, 1991). The temporal module can be given a start signal, which resets the clock, after which an accumulator starts collecting pulses. The short interval of 2 seconds corresponds to approximately 17 pulses, and the long interval of 3.1 seconds to approximately 26 pulses. Noise is added to each pulse, which means that estimates are always approximate. For the purposes of the model, the important aspect of the time estimation module is that it can estimate a particular time interval by translating it into number of pulses, and that it can reproduce a time interval by waiting until a particular number of pulses has been accumulated. The noise produces variability in the estimates that corresponds to the variability in human time estimation.

*Declarative Memory*

The assumption of the model is that when a particular time interval has to be produced, the number of pulses representing that interval is retrieved from memory. There is no single representation of a particular interval in memory, but rather a pool or collection of past experiences. Each past experience is represented by a memory chunk, which contains the type of interval (long or short), and a number of pulses. When an interval is retrieved from memory at time *t*, each chunk receives an activation value on the basis of its age (how old is the experience), and whether it matches the current request:

$$A(t) = \log(t - t_{creation})^{-d} + mismatchpenalty \quad (1)$$

In this equation, $t_{\text{creation}}$ is the time the chunk is created, so the activation of a chunk decreases with time. The *mismatchpenalty* of a chunk is 0 if the request matches the chunk (e.g., we are retrieving a short interval and the chunk represents the short interval), but a negative value in the case of a mismatch (e.g., we try to retrieve a short interval but the chunk represents a long interval).

In default ACT-R, the activation determines the probability of retrieval of that chunk. This means, if one assumes that each trial is reflected in a separate chunk in memory, that more recent experiences that match the request have the highest probability to be retrieved. The following equation estimates these probabilities (where *t* is a noise parameter, and the summation is over all candidate chunks):

$$P_i = \frac{e^{A_i/t}}{\sum_j e^{A_j/t}} \quad (2)$$

With the blending mechanism (Lebiere et al., 2007), however, a weighed average of all candidate chunks is retrieved. If we try to retrieve the duration of the short interval, the results will be a blend of all intervals in the memory pool, with the more recent intervals having a higher impact, and the intervals that match the request (short) having a higher impact than the mismatching long intervals. The resulting value can simply be calculated by multiplying the number of pulses in a chunk ($V_i$) by the probability of retrieval:

$$\text{Result value} = \sum_j P_j V_j \tag{3}$$

In order to determine how many pulses to wait for an interval, the model not only retrieves the representation of the interval, but also the feedback received for that interval. For this we use exactly the same mechanism as for the retrieval of the interval. Whenever feedback is received, the model stores this in memory. If the feedback was "correct" it stores the value of 0, if it was "too long" it stores a negative value, and when it is "too short" it stores a positive value (this value is referred to as the feedbackshift, which is a free parameter in the model). Retrieval of the feedback is performed in the same way as the retrieval of the interval itself. This means that the feedback of previous trial for the same duration has the highest impact, but that earlier feedback and feedback for the other duration can also weigh in.

Table 3. Example of how the model calculates the number of pulses

| Experience | Pulses | $t$ (sec) | $A_j$ | $P_j$ | $P_jV_j$ | Feedback | $P_jV_j$ Feed-back | |
|---|---|---|---|---|---|---|---|---|
| $\text{long}_{n-1}$ | 28 | 4.5 | -1.67 | 0.029 | 0.8 | -2 (too late) | -0.76 | |
| $\text{short}_{n-1}$ | 22 | 7.5 | -1.01 | 0.797 | 17.5 | 0 (correct) | 0 | |
| $\text{long}_{n-2}$ | 27 | 11.1 | -2.12 | 0.003 | 0.08 | +2 (too early) | 0.003 | |
| $\text{short}_{n-2}$ | 20 | 13.8 | -1.31 | 0.171 | 3.43 | +2 (too early) | 0.15 | |
| Sum | | | | | 21.8 | | -0.61 | 21.19 |

Table 3 shows an example in the hypothetical case in which the number of pulses for the next short interval is calculated on the basis of the last four experiences (the actual pool model uses all previous experiences, but older experiences have less impact due to decay). Each line in the table shows an experience in memory, starting with the type (long or short), how many pulses were used as estimate in that experience and how long ago the experience was. On that basis, using Equation 1, an activation is calculated, in which the long experiences are penalized because they do not match the current request (i.e., a short interval). Equation 2 is then used to calculate the probability of retrieval of that experience, which multiplied with the numbers of pulses, gives the contribution of that experience to the blend in the sixth column of the table. These contributions are added up (Equation 3) to produce the result of the blended retrieval, 21.8. This process is repeated for the feedback, summarized in the next two columns, and leading to a contribution of -0.61. The sum of the two retrievals is 21.19, which, rounded down to 21 means that the estimate for the next short interval will be slightly shorter than the previous, even though it received positive feedback.

To summarize: if the pool model has to produce a certain interval, it determines the number of pulses by retrieving a blend of memory representations for that interval. It then retrieves previous feedback for that interval, which is also a blend of earlier feedback. It adds the two together, and waits for that many pulses to produce the interval.

The pool model's behavior is partly determined by a number of parameters, some of which are derived from earlier work, whereas others are new. The time estimation module parameters were left at their defaults (from Van Rijn & Taatgen, 2008), but had no great impact of the results ($t_0$ = 100 ms, $a$ = 1.02, $b$ = 0.015). ACT-R's memory decay parameter $d$ was also left at its default value of 0.5. We used the remaining three parameters, listed in Table 4, to produce a good fit between the model and the data in de DR (dike-river) condition[2].

Table 4. Free parameters in the pool model.

| Parameter | |
|---|---|
| Noise parameter $t$ | 0.2 |
| Mismatch penalty between short and long | 0.92 |
| Feedbackshift: how many pulses to add or subtract on the basis of feedback | 1.8 |

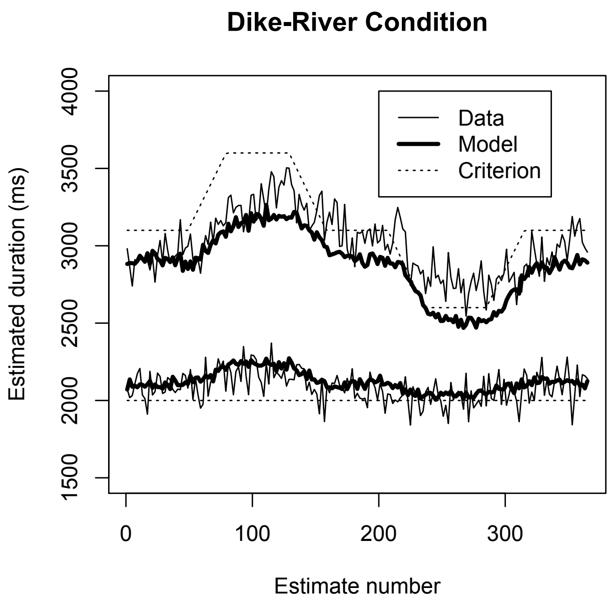


Figure 3. Model fit of the DR condition.

Figure 3 depicts the results of the model, and shows how it fits the data. In the model results there is an impact of the long on the short interval that is similar to the data, and otherwise it tracks the estimates of the subjects rather well, with the exception of the "river" part of the long interval where its estimates are slightly shorter than the subjects'. Although Figure 3 is useful to evaluate the qualitative aspects of the fit, it is only an approximate source of support for the memory model used to produce it. In order to have a better assessment of the model, we applied two strategies: we used the model to predict the outcomes of the other two conditions, and we used the same mixed-model analysis that we used to analyze the data to check whether the same factors that drove the estimates in the data also do so in the model (see also Taatgen & Van Rijn, 2010).

Figure 4 shows the model predictions for the FF and RD condition. The FF condition shows that the model also predicts a shortening of the long interval and a lengthening of the short interval. The RD condition produces a surprisingly good fit that surpasses the quality of the fit in the DR condition.

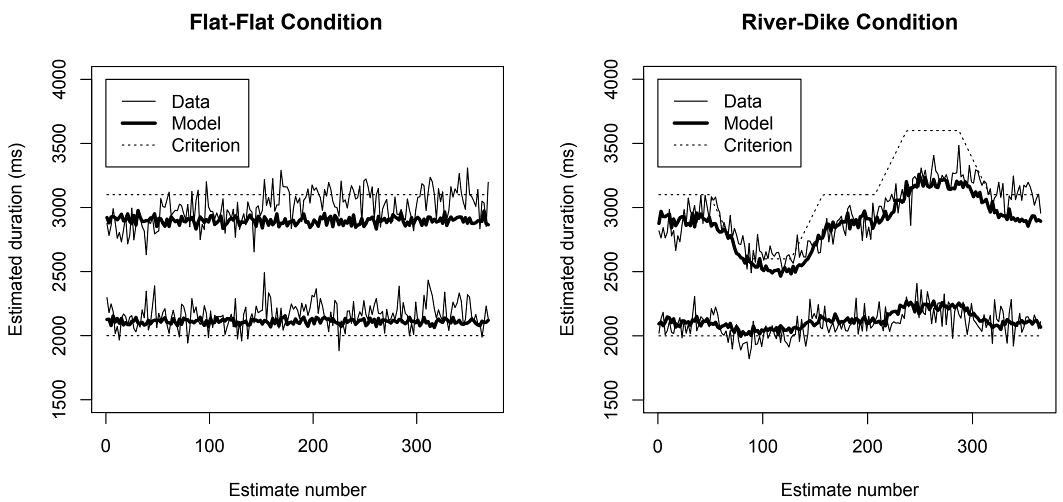


Figure 4. Model predictions for the FF and RD conditions.

To better assess the quantitative aspects of the fit, we applied the same regression analysis to the model outcomes as we used for to analyze the data. In doing so, we can check whether the same factors that played a role in producing the estimate for subjects also play the same role in the model's estimates. Table 5 and Table 6 show this comparison for the short and long interval, respectively.

Table 5. Results of comparing the factors of mixed-effect models fitted to the data and the cognitive model to the estimates of the short interval in the three conditions.

| | FF condition | | RD condition | | DR condition | |
|---|---|---|---|---|---|---|
| Factor | Beta Data | Beta Pool Model | Beta Data | Beta Pool Model | Beta Data | Beta Pool Model |
| Intercept (ms) | 1217 | 1296 | 962 | 992 | 826 | 1000 |
| **Impact of previous short intervals on current short interval** | | | | | | |
| $\text{short}_{n-1}$ | 0.18 | 0.27 | 0.17 | 0.30 | 0.28 | 0.30 |
| $\text{short}_{n-2}$ | 0.079 | 0.024 | 0.087 | 0.043 | 0.11 | 0.037 |
| short-$\text{fb}_{n-1}$ = "too long" (ms) | -106 | -148 | -96.8 | -156 | -153 | -150 |
| short-$\text{fb}_{n-1}$ = "too short" (ms) | 92.0 | 138 | 48.9 | 131 | 68.2 | 132 |
| **Impact of previous long interval on current short interval** | | | | | | |
| $\text{long}_{n-1}$ | 0.14 | 0.06 | 0.21 | 0.14 | 0.15 | 0.13 |
| long-$\text{fb}_{n-1}$ = "too long" (ms) | -130 | -116 | -97.5 | -131 | -83.8 | -123 |
| long-$\text{fb}_{n-1}$ = "too short" (ms) | 39.1 | 120 | 67.2 | 156 | 69.1 | 157 |

Table 6. Results of comparing the factors of mixed-effect models fitted to the data and the cognitive model to the estimates of the long interval in the three conditions.

| | FF condition | | RD condition | | DR condition | |
|---|---|---|---|---|---|---|
| Factor | Beta Data | Beta Pool Model | Beta Data | Beta Pool Model | Beta Data | Beta Pool Model |
| Intercept (ms) | 1245 | 2290 | 568 | 657 | 906 | 675 |
| **Impact of previous long intervals on current long interval** | | | | | | |
| $\text{long}_{n-1}$ | 0.28 | 0.14 | 0.50 | 0.50 | 0.46 | 0.51 |
| $\text{long}_{n-2}$ | 0.12 | 0.003 | 0.17 | 0.12 | 0.17 | 0.11 |
| long-$\text{fb}_{n-1}$ = "too long" (ms) | -108 | -271 | -237 | -461 | -170 | -439 |
| long-$\text{fb}_{n-1}$ = "too short" (ms) | 114 | 153 | 267 | 315 | 220 | 309 |
| | | | | | | |

| **Impact of previous short interval on current long interval** | | | | | | |
|---|---|---|---|---|---|---|
| $short_{n-1}$ | 0.25 | 0.13 | 0.16 | 0.18 | 0.10 | 0.18 |
| $short\text{-}fb_{n-1}$ = "too long" (ms) | -87.2 | -127 | -56.9 | -117 | -74.8 | -113 |
| $short\text{-}fb_{n-1}$ = "too short" (ms) | 136.3 | 8.9 | 80.9 | 46.5 | 38.0 | 17.9 |

Although these tables do not provide us with a neat summarizing number that tells us the quality of the fit, it is doubtful whether a single number can achieve such a goal (e.g., Navarro, Pitt, & Myung, 2004; Roberts & Pashler, 2000; Schunn & Wallach, 2001). Instead, these tables show that the same factors that were important in predicting the estimates in the data, play a similar role in the estimate made by the model. In many, but not all, cases the model betas are very close to the values found in the data. Only in the fit of the long estimates does the model's predictions diverge from the data, because the model's intercept is higher and the factors for $long_{n-1}$ and $long_{n-2}$ smaller, indicating that the computational model (still) has a too stable representation of the long interval.

## General Discussion

The experiment shows very explicitly that representations of both temporal intervals in memory influence each other, supporting suspicions based on findings by, for example, Jones and Wearden (2004), Grondin (2005) and Van Rijn and Taatgen (2008). It not only shows that the representations of two intervals tend to shift towards each other, but also that a change in the duration of one interval not only affects the representation of that interval, but also the representation of the unchanged interval. These findings support a model in which the representation of a time interval is not a single memory trace, but a pool of experiences in which recency and match to the current request determine the impact of single experiences.

The basis of the cognitive model is a simple memory model that is based on Anderson's rational analysis theory (Anderson, 1990), which has been used to model many memory phenomena (e.g., Anderson & Matessa, 1997; Anderson & Reder, 1999; Taatgen & Wallach, 2002, see for a more extensive list: http://act-r.psy.cmu.edu/publications/index.php?topic=2). This model, when combined with a pacemaker-accumulator time perception model, is sufficient to explain the phenomena found in this experiment. The main specific choice we made in this model is to treat every experience with each of the intervals as a separate memory trace. The retrieval process produces a mix of these memory traces through a blending mechanism (Lebiere et al., 2007).

In the introduction we discussed several alternative memory models for representing time intervals. In a first alternative solid representations of an interval are formed and strengthened by experience. This alternative, which is mixture of the SAM and AVG models (Jones & Wearden, 2003), retrieves a single past experience (using Equation 2) and uses that for the next estimate. If it receives positive feedback, it strengthens this experience. On negative feedback it creates a new memory trace that incorporates the feedback. It can account for the impact of one interval on the other because of the possibility that the model retrieves a wrong interval (e.g., a short interval while a long interval was requested). This alternative model, however, is not able to fit the data, because the model quickly establishes a stable representation of each interval (as the reinforcement results in a "the winner takes all" situation), making this model too sluggish to track changes in the long interval, or explain how those changes impact the short interval.

The second alternative, the perturbation model by Jones and Wearden (2003), is also not able to account for the data because the instantiation of this model as presented in Jones and Wearden (2003) lacks the ability to incorporate the influence of other intervals. However, our pool model has an important property in common with that the perturbation model, in the sense that it is mainly driven by recent experiences, but it takes into account some of the more the recent past than the perturbation model.

To demonstrate these differences, we implemented a version of both the solid representation and perturbation models. The resulting estimates for the RD condition are shown in Figure 5. The solid representation model is able to capture some of the general contamination of the two intervals, because it overestimates short intervals and underestimates long intervals. However, it is not able to follow the changes in the long interval very well. The perturbation model follows those changes very well (or one could say, too well), but is not able to capture the contamination of the two intervals. A comparison of the factors in the regression confirms these observations: Table 7 shows those factors for the short interval in the RD condition. The table shows that both alternative models capture most of the factors that refer to earlier experiences with the short interval, but not the long interval. Moreover, both alternative models have a much too high intercept, indicating that previous experiences have less impact than observed in the data.

Of course, the perturbation model could be extended to account for many of the phenomena discussed here. For example, as Jones and Wearden (2003) argue, instead of replacing the old value with the new value, a more gradual change could be proposed. However, we could not come up with a modification of the perturbation system that (1) could produce relative stable performance for both durations, (2) adjust itself to changes in the standard, and (3) show influences of the changed long interval on the short interval. That is, changes necessary to account for these phenomena would make the perturbation model very similar to the pool model of temporal reference memory.

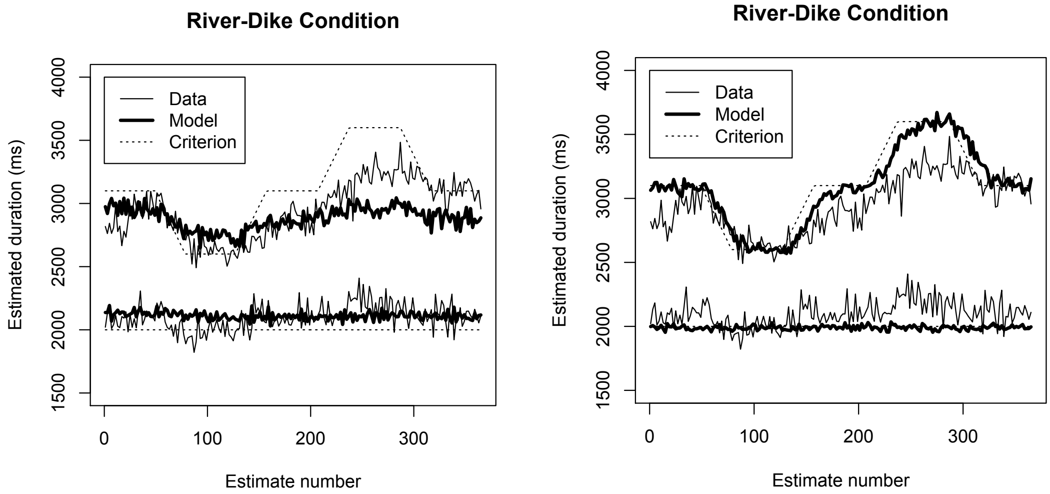


Figure 5. Alternative model fits for the RD condition. Left: Solid representation model, Right: Perturbation model

Table 7. Comparison between data and the three models

| Factor | Beta Data | Beta Pool Model | Beta Solid Representation Model | Beta Perturbation Model |
|---|---|---|---|---|
| Intercept (ms) | 962 | 992 | 1851 | 1551 |
| $short_{n-1}$ | 0.17 | 0.30 | 0.079 | 0.21 |
| $short_{n-2}$ | 0.087 | 0.043 | 0.057 | 0.009 |
| short-$fb_{n-1}$ = "too long" (ms) | -96.8 | -156 | -252 | -138 |
| short-$fb_{n-1}$ = "too short" (ms) | 48.9 | 131 | 250 | 134 |
| $long_{n-1}$ | 0.21 | 0.14 | 0.01 | 0.00 |
| long-$fb_{n-1}$ = "too long" (ms) | -97.5 | -131 | -0.5 | 7.4 |
| long-$fb_{n-1}$ = "too short" (ms) | 67.2 | 156 | 3.2 | -5.9 |

The advantage of a representation of time intervals based on a pool of experiences is that it very flexible (note that this is also true for the perturbation model). A representation can be adapted quickly to changing circumstances. A more practical example of such an adaptation is multitasking during driving. When a driver wants to operate some device in the car, they have to look away from the road for as long as this is safe. This interval is subject to changing circumstances, because one can look away much longer from a quiet straight road than a busy curved road. In an experiment in which addresses has to be typed into a navigation device while driving in a simulator, Salvucci, Taatgen and Kushleyeva (2005) found that subjects adapt the time interval they spend on the navigation device to the difficulty of the driving task. Such an adaptation would be harder to accomplish if intervals are represented as single solid representations.

One of the problems with the experiment, and other experiments involving multiple time intervals, is that many subjects had to be removed from the dataset because they could not keep the representation for the two intervals apart. The mechanism that keeps the model from mixing up two intervals is the mismatchpenalty in Equation 1, so lowering the penalty causes the model to mix up the two intervals. Figure 6 shows a comparison between a model in which the mismatch penalty is lowered to 0.3, and the 10 subjects that were rejected because they mixed up the interval. Even though this only an approximate comparison (the simulation is of the FF condition even though subjects are from all three), and the data are very noisy, it still shows that the model can offer an explanation for this group.

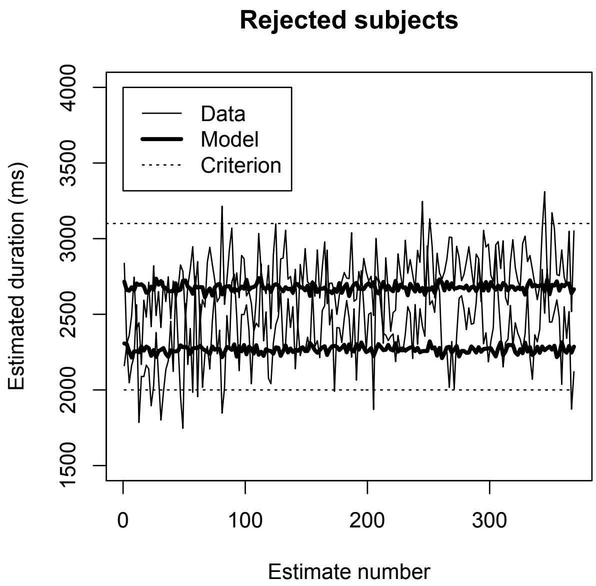


Figure 6. Comparison between the model with a reduced mismatchpenalty and 10 subjects that were removed from the original analysis.

An important issue is the generality of proposed models. For the model fits presented in this paper, the main mechanism is the retrieval from a pool of previous encounters that is driven by well-tested memory mechanisms. However, ACT-R's declarative memory model is probably not the only paradigm that can model these data. Indeed, the approach would be perfectly suitable for a memory model based on neural networks but such a model would adhere to the same idea that a pool of experiences produces the new estimate. And, as already mentioned in the introduction, the fit to the data does not hinge on the linear or nonlinear representation of time in the clock component – as long as a clock component produces temporal information, the pool model would be able to produce new temporal estimates.

# References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control, 19*(6), 716–723.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press, New York.

Anderson, J. R. & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review*, 104, 728-748

Anderson, J. R. & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128, 186-197.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Baudouin, A., Vanneste, S., Pouthas, V., & Isingrini, M. (2006). Age-related changes in duration reproduction: Involvement of working memory processes. *Brain and Cognition, 62*(1), 17–23.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika, 26*, 255–280.

Brown, S. W. & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica, 75*(2), 103–121.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics, 59*(7), 1118–1140.

Buhusi, C. V. & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature reviews. Neuroscience, 6,* 755 –765.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer: New York.

Church, R. (1984). Properties of the internal clock. *Annals of the New York Academy of Sciences, 423*, 566–582.

Fortin, C., Champagne, J., & Poirier, M. (2007). Temporal order in memory and interval timing: An interference analysis. *Acta Psychologica, 126*(1), 18–33.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*(3), 279 – 325.

Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation, 22*(1-2), 3–38.

Gibbon, J. & Church, R. M. (1984). Sources of variance in an information processing theory of timing. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 465–488). Hillsdale, NJ: Erlbaum.

Grondin, S. (2005). Overloading Temporal Memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 869–879.

Jones, L. A. & Wearden, J. H. (2003). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *The Quarterly Journal of Experimental Psychology Section B, 56*(4), 321–343.

Jones, L. A. & Wearden, J. H. (2004). Double standards: Memory loading in temporal reference memory. *The Quarterly Journal of Experimental Psychology Section B, 57*(1), 55–77.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology, 7*(4), 495–531.

Lebiere, C., Gonzalez, C., & Martin, M. (2007). Instance-based decision making model of repeated binary choice. In: P*roceedings of the 8th International Conference on Cognitive Modeling* (pp. 67-72). Oxford, UK: Taylor & Francis.

Lejeune, H. (1998). Switching or gating? The attentional challenge in cognitive models of psychological time. *Behavioural Processes, 44*(2), 127–145.

Lewis, P. A. & Miall, R. C. (2006). A right hemispheric prefrontal system for cognitive time measurement. *Behavioural Processes, 71*(2-3), 226–234.

Meijering, B. & Van Rijn, H. (2009). Experimental and computational analyses of strategy usage in the time-left task. In Taatgen, N. A. and Van Rijn, H. (eds.), Proceedings of the 31th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society. (pp. 1615-1620)

Michon, J. A. (1967). *Timing in Temporal Tracking.* Van Gorcum, Assen, The Netherlands.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49*, 47–84.

Ogden, R. & Jones, L. (2009). More is still not better: Testing the perturbation model of temporal reference memory across different modalities and tasks. *The Quarterly Journal of Experimental Psychology, 62*(5), 909–924.

Ogden, R., Wearden, J., & Jones, L. (2010). Are memories for duration modality specific? *The Quarterly Journal of Experimental Psychology, 63*, 65-80.

Pothos, E. M. & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*(3), 303 – 343.

Raaijmakers, J. G. W. & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134.

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

Schunn, C. D. & Wallach, D. (2001). Evaluating goodness-of-fit in comparison of models to data. Retrieved November 14, 2005, from http://www.lrdc.pitt.edu/schunn/gof.

Staddon, J. E. R. & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior, 71*(2), 215–251.

Taatgen, N. & Van Rijn, H. (2010). Nice graphs, good $R^2$, but still a poor fit? how to be more sure your model explains your data. In Salvucci, D. D. & Gunzelmann, G. (eds.), *Proceedings of the 2010 International Conference on Cognitive Modeling.*

Taatgen, N. A., Van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention and learning. *Psychological Review, 114*(3), 577–598.

Taatgen, N. A. & Wallach, D. (2002). Whether skill acquisition is rule or instance based is determined by the structure of the task. *Cognitive Science Quarterly*, *2*(2), 163-204.

Tenenbaum, J. B. & Griffiths, T. L. (2002). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629–640.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock”. *Psychological Monographs, 77*(13), Whole no. 576.

Van Rijn, H. & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica, 129*, 365–375.

Van Rijn, H., Van Someren, M., & Van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science, 27*(2), 227–257.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/f noise in human cognition. *Psychonomic Bulletin & Review, 11*(4), 579–615.

Wearden, J. H. (2002). Traveling in time: A time-left analogue for humans. *Journal of Experimental Psychology: Animal Behavior Processes, 28*(2), 200–208.

Zakay, D. & Block, R. A. (1996). The role of attention in time estimation processes. *Advances in psychology, 115*, 143–164.

Footnotes

1 This can be either because of the assumption that traces never merge, or because of the assumption the they cannot merge due to the non-symbolic nature of internal time. In this latter assumption, the representation of a particular time interval is represented on a continuous scale (or, in a computationally implemented model, as a real number), and each new temporal experience will be different from earlier experiences.

2 The fit was published in Taatgen & Van Rijn (2010), before the data from the other two conditions were analyzed and modeled.

Author Note

Niels Taatgen, Artificial Intelligence, University of Groningen and Hedderik van Rijn, Psychology, University of Groningen.

We thank Stefan Wierda for collecting the data and the members of the cognitive modeling group for their useful comments.

Correspondence concerning this article should be addressed to Niels Taatgen, Artificial Intelligence, University of Groningen, P.O. Box 407, NL-9700 AK Groningen, Netherlands. E-mail: n.a.taatgen@rug.nl.