Traces of times past: Representations of temporal intervals in memory

Niels Taatgen · Hedderik van Rijn

Abstract Theories of time perception typically assume that some sort of memory represents time intervals. This memory component is typically underdeveloped in theories of time perception. Following earlier work that suggested that representations of different time intervals contaminate each other (Grondin, 2005; Jazayeri & Shadlen, 2010; Jones & Wearden, 2004), an experiment was conducted in which subjects had to alternate in reproducing two intervals. In two 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 the adjustment of one interval carried over to the other interval, indicating that subjects were 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

Although there are many theories and models of how people estimate time intervals, they typically contain three components: a clock component, a memory component,

N. Taatgen (✉)
Department of Artificial Intelligence, University of Groningen,
P.O. Box 407, 9700 AK Groningen, The Netherlands
e-mail: n.a.taatgen@rug.nl

H. van Rijn
Department of Psychology, University of Groningen,
Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
e-mail: hedderik@van-rijn.org

and a comparison component (e.g., Church, 1984; Gibbon & Church, 1984; Michon, 1967; Treisman, 1963). A primary topic of debate related to these theories is the nature of the clock component. Some theories assume that the clock represents time linearly, such as the scalar expectancy theory (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 the 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 et al., 2007). Apart from these studies that have focused on the separate components themselves, 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) and also clinical studies and pharmacological manipulations have shown that the clock and memory components have independent biological substrates (for a review, see Buhusi & Meck, 2005), many studies have shown that these systems are intimately tied together to produce accurate time estimations. In such studies, it is typically assumed that the memory system contains information that reflects earlier temporal experiences, which are then 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. Two paradigms are used in this research: experiments in which a presented interval has to be compared to an explicitly

or implicitly learned standard, and experiments in which subjects have to reproduce an earlier-presented duration.

Jones and colleagues used the first of these paradigms for various purposes: to test whether multiple presentations of an interval improve temporal reproduction (Jones & Wearden, 2003; Ogden & Jones, 2009), whether performance is degraded when multiple durations have to be learned and kept active simultaneously (Jones & Wearden, 2004), whether memory traces are modality independent (Ogden & Jones, 2009; Ogden, Wearden, & Jones, 2010), and whether the temporal structure of the task itself influences temporal judgments (Ogden, Wearden, & Jones, 2008). We will discuss two of these studies in more detail below.

In the experiments reported in Jones and Wearden (2003), subjects were presented with one, three, or five 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 (2003) constructed a number of computational models that varied in the ways 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 a 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 because (1) it is difficult to 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 (2004) themselves. In their experiments, subjects had to judge whether presented intervals were the same as or different from a previously learned interval. The subjects had to learn two intervals and were presented with a test stimulus that they had to match up with one of the intervals. The results, in particular those of Experiment 2, showed that short intervals tended to be judged as longer on average, and long intervals tended to be judged as shorter (see Fig. 5 later in this article, where a model fit is compared to the results of this experiment).

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, 250 and 750 ms, and were told before each trial whether to base their judgment on the shorter or the longer trial. The results indicated that the subjects tended to shift their

representations of the 250-ms and 750-ms intervals toward each other, as if the representations contaminated—instead of replaced—each other in memory (see Fig. 6 below).

In the second paradigm used to study memory in time perception, subjects are asked to reproduce intervals presented earlier. In an earlier study of our own in which subjects learned and had to reproduce 2- and 3-s intervals, we noticed that estimates of the 2-s interval tended to be long, and estimates of the 3-s interval tended to be short (Van Rijn & Taatgen, 2008). In a more recent study, Jazayeri and Shadlen (2010) also used a temporal reproduction paradigm. In their experiment, subjects were repeatedly presented with time intervals that they had to reproduce immediately afterward. Depending on the condition, the presented intervals were drawn from a particular range—for example, 671–1,023 ms. The results showed that the reproduced intervals tended to regress to the middle, so that the shorter intervals were reproduced as longer and the longer intervals as shorter, again supporting the idea that earlier experiences (or "context," as Jazayeri and Shadlen called it) affect the current representation.

All of these studies support a view of the temporal reference memory system in which individual representations are not completely separable. Here, it is important to note that Jones and Wearden (2004) themselves argued that their perturbation model, as sketched here and in their study, is probably too simple, and that instead of completely replacing the old value, a more gradual change might fit the data better.

The 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. A number of studies have provided evidence that there is an intimate link between the temporal reference memory system and more general memory phenomena such as working memory. 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. Fortin, Champagne, and Poirier (2007) have shown that when a concurrent memory task is performed during time estimation, the temporal estimates are more strongly influenced if the concurrent task requires order judgments. Similarly, Baudouin, Vanneste, Pouthas, and Isingrini (2006) have shown in a study with older adults that temporal reproduction is correlated to working memory capacity. These studies are just examples of many that have shown evidence for the notion that we cannot treat temporal reference memory as if it were independent from other memory functions.

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 the present article, we 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 of how solid the memory representations of time intervals are. One explanation is that over the course of an experiment involving multiple intervals, solid representations of each of the intervals are formed. In that explanation, the formation of these representations can be influenced by the fact that another interval has to be learned at the same time, resulting in interference. Another explanation is that solid memories never form.[1] This distinction is analogous to one in the discussion of 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 "pool of experiences," 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 of how this single retrieved representation is constructed.

One of the computational models explored by Jones and Wearden (2003) is the *sampling* strategy (they abbreviated it as SAM, but we will refer to this model as SAMP to avoid confusion with the Raaijmakers & Shiffrin, 1981, theory with the same name). The SAMP 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 observed, this makes

it difficult to account for the flexibility observed when learners are confronted with a changed standard time. Because, 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 (2003) 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 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 a 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.

A different approach to modeling temporal representations is the Bayesian modeling approach by Jazayeri and Shadlen (2010). The assumption of their model is that humans take two factors into account when estimating the duration of an interval. One of these factors is the temporal context: What is the range of possible durations for an interval? The second factor is the (self-)knowledge about the imprecision involved in estimating this interval. These two factors together can explain the regression to the mean that Jazayeri and Shadlen have found in their experiment. The drawback of this model is that it does not update the temporal context during the experiment, and is therefore not suitable to predict or explain what happens if a temporal standard changes.

All three accounts assume a relatively perfect memory system, in the sense that experiences that are stored are not subjected to decay, or other influences that have been identified in the memory literature. Thus, although the SAMP, AVE, and perturbation models can probably be extended to account for (some or most) new phenomena, another approach would be to 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; see, e.g., Pothos & Chater, 2002; Tenenbaum & Griffiths, 2001), in which recent experiences have a much stronger weight than older experiences. This pool of experiences may be polluted by

---

[1] This could occur either because of the assumption that traces never merge, or because of the assumption that they cannot merge, due to the nonsymbolic nature of internal time. According to the latter assumption, 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.

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 the intervals were represented. The first was to analyze the reproductions of the two intervals on a trial-by-trial basis. This allowed us to inspect the changes of temporal estimations in much more detail than if one were just to compare average performance, as had been done in earlier work. We also 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 only have 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 would expect the intercept to be rather small, with larger influences for recent experiences.

For our 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 was given on the basis of the changed duration, subjects would have to adjust their internal representations to remain at a reasonable level of performance. However, because the experimental manipulations were only introduced after a number of trials in the experiment, there was 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 are more the result of a set of experiences, we would expect that a change in one of the intervals would spill over into the other interval, something that one would not expect if the intervals were solidly represented.

In the rest of this article, we first elaborate on the experiment and the analysis of the results of that experiment. We then proceed with a model of these results that is based on an implementation of the pool model in 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

Subjects

A total of 70 students from the university of Groningen participated in this study for course credit. Of these

subjects, 12 were removed from the pool because more than 3% of their responses were shorter than 1.25 s or longer than 4.25 s. Of the remainder, 10 were not able to distinguish between the two intervals after training, and were therefore also removed from the data set. We will reconsider these 10 subjects in the discussion. The remaining 48 subjects (16 per condition detailed below) were an average age of 20.3 years and consisted of 10 men and 38 women. The number of excluded subjects was fairly high, but this was expected, given that experiments with multiple time intervals often elicit 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 s and a long interval of 3.1 s—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-s interval, while the circle on the left was associated with the 3.1-s 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 each presentation of the standard interval, the subjects had to reproduce this temporal interval. The onset of the interval was randomly sampled from a uniform distribution ranging from 500 to 1,000 ms after presentation of the standard and was indicated by the gray circle turning blue or green again. 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 these reproductions 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–estimate–feedback trials, alternating between the two durations.

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

The main experimental manipulation was a shift in the criterion for the long interval in two of the three between-subjects conditions. In the FF (flat–flat) condition, the criterion remained the same for the rest of the experiment (185 estimates for each interval). In the DR (dike–river, referred to as such because the graphical depiction of the standard follows the typical outline of a dike next to a riverbed; see the dotted lines in Fig. 1c) condition, the
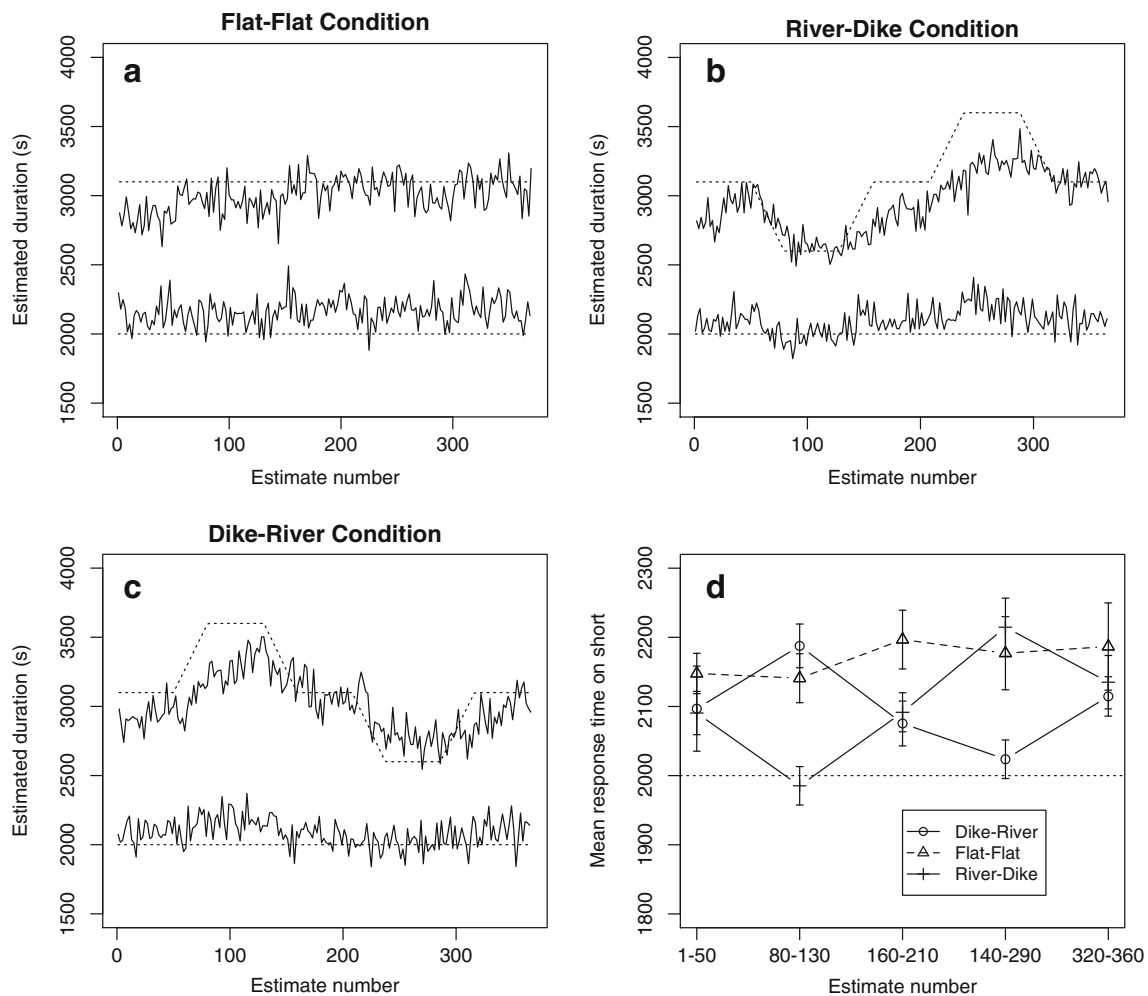
Fig. 1 (**a–c**) Mean estimates for the three conditions in the experiment. The lower curve in each graph represents the estimates for the short interval, while the upper curve represents the long interval. The dotted lines indicate the criterion, which is always constant at 2 s for the short interval, but which changes for the long interval in the river–dike and dike–river conditions. (**d**) Mean durations for the short interval in the parts of the experiment where the long interval does not change. Error bars represent ±1 standard error

criterion remained at 3.1 s for the first 25 estimates. After that, the criterion was increased linearly to 3.6 s over 15 estimates. This meant that at some point subjects received "too short" feedback for a duration that had previously been correct. After the shift to 3.6 s, the criterion stayed at 3.6 s for 25 estimates, then decreased back linearly to 3.1 s over 15 estimates, stayed there for another 25 estimates, then decreased further to 2.6 s over 15 trials, stayed at 2.6 s for 25 trials, increased back to 3.1 s over 15 trials, and stayed there for the remaining 25 estimates. Meanwhile, the criterion for the short interval (remember that the short and long intervals were alternated) remained constant at 2 s. 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 s, leading to this sequence: 25 trials at 3.1 s, 15 decreasing to 2.6 s, 25 at 2.6 s, 15 increasing to 3.1 s, 25 at 3.1 s, 15 increasing to

3.6 s, 25 at 3.6 s, 15 decreasing to 3.1 s, 25 at 3.1 s. A graphical depiction of all three conditions is presented with the dotted lines in Figure 1a–c.

Results

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 was estimated as longer and the long interval as shorter, consistent with earlier findings (e.g., Grondin, 2005), and suggesting that both estimates influence each other. More pronounced were the results in the other two conditions, because the estimates of the short interval were influenced by changes in the duration of the long interval, because the short interval's estimations resemble a dampened pattern of

the long interval. Figure 1d shows this influence in more detail.

An analysis of variance on the mean durations for the short interval in the five estimate ranges in which the long interval was not changing and in the FF condition (the estimates plotted in Fig. 1d) showed a clear interaction between condition and range, $F(8, 179) = 4.95$, $p < .001$, with no main effect of range, $F < 1$, and no main condition effect, $F(2, 44) = 2.03$, $p = .14$.

The overall accuracy (i.e., the proportion of estimates within 12.5% of the target interval) was 55% for short intervals and 63% for long intervals, with only small differences between the conditions (for short, FF 52%, RD 59%, and DR 55%; for long, FF 63%, RD 66%, and DR 61%). Apparently, the criterion manipulation for the long interval was slow enough not to affect accuracy.

In order to analyze the influence of one interval on the other in more detail, we had to acknowledge that two factors determined the next estimate that a subject would make: 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 and feedback for the other interval could influence the estimate. Figure 2 illustrates these factors: At the right side of the figure, a short interval had to be estimated by the subject, which could be influenced by the previous estimates and previous feedback of both intervals.

To assess all of these factors, we have fit linear mixed-effect models to the each of the two interval durations (Baayen, Davidson, & Bates, 2008). The estimations were entered as the dependent variable, and previous (up to $n - 10$) estimates and feedback were entered as predictors, both for the current duration and for the other durations, while allowing a random effect for subjects. Linear mixed-effect models provided information about the contributions 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 fewer estimates/feedback) using the Akaike information criterion (AIC; Akaike, 1974) and the maximum likelihood criterion, as discussed in Burnham and Anderson (2002). We first compared the contribution of the previous estimations of the target interval, then the contribution of feedback on these previous estimations of the target interval, and then the estimations and feedback on previous estimations of the other interval. The comparisons were first performed on the RD data set, and then the same models were fit to the DR and FF data sets. We report the preferred models, meaning that increasing or decreasing the number of predictors from the reported models 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 1,217 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 s, so something on the order of 360 ms), and 0.079 times the short estimate before that. The difference between these fractions (0.18 and 0.079) indicates that the influence of the $n - 2$ estimation contributes 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 because adding $long_{n-2}$ did not result in an improved RD (not FF) 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 was "too long," it is decreased the current estimate by 106 ms. Finally, feedback on the previous long intervals also impacts the predicted estimate on the short interval: 130 ms is

Fig. 2 Factors that might impact on the estimation process. In this example, estimation of a short interval is shown. S stands for "subject," and the gray areas indicate the intervals in which the participant received "correct" as feedback. The factors with solid arrows turned out to have significant impacts, the dashed arrows did not
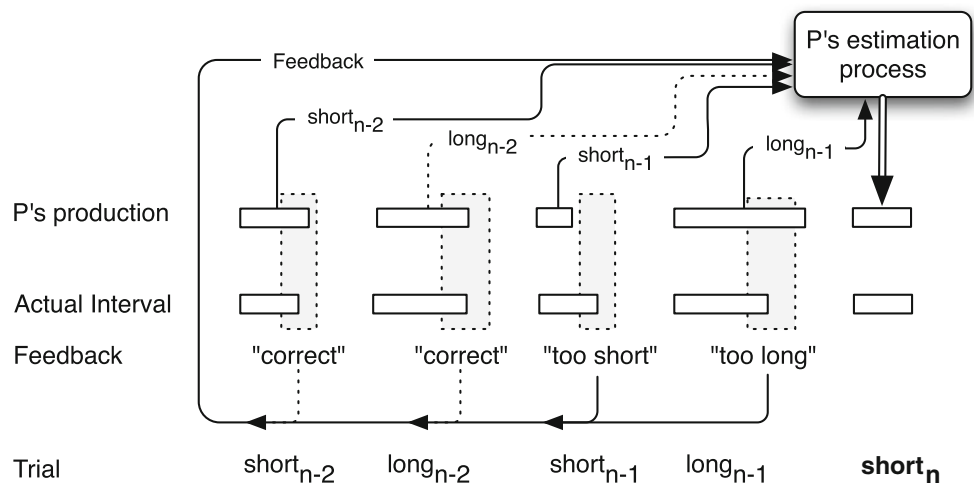
Table 1 Results of fitting mixed-effect models to the estimates of the short interval in the three conditions

	FF Condition		RD Condition		DR Condition	
Factor	Beta (*SE*)	*t*	Beta (*SE*)	*t*	Beta (*SE*)	*t*
Intercept (ms)	1,217 (120)	10.1***	962 (93)	10.3***	826 (92)	9.0***
Impact of the previous short intervals on the 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***
$short\text{-}fb_{n-1}$ = "too long" (ms)	−106 (25.8)	−4.1***	−96.8 (22.9)	−4.2***	−153 (23)	−6.7***
$short\text{-}fb_{n-1}$ = "too short" (ms)	92.0 (27.6)	3.3***	48.9 (22.7)	2.2*	68.2 (23.9)	2.9**
Impact of the previous long interval on the current short interval						
$long_{n-1}$	0.14 (0.03)	4.3***	0.21 (0.02)	11.0***	0.15 (0.02)	8.0***
$long\text{-}fb_{n-1}$ = "too long" (ms)	−130 (32)	−4.0***	−97.5 (23.1)	−4.2***	−83.8 (22.1)	−3.8***
$long\text{-}fb_{n-1}$ = "too short" (ms)	39.1 (28.5)	1.4	67.2 (17.2)	3.9***	69.1 (19.1)	3.6***

FF, flat–flat; RD, river–dike; DR, dike–river. See the text for further details. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

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 the FF condition, 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 of 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 affect all of them.

Table 2 shows the analysis for the long interval. The results are similar to those for 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 the conditions. The $long_{n-1}$ and $long_{n-2}$ factors are larger in

the RD and DR conditions because they are needed to track the changing interval criterion.

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 of the factors identified in the statistical analysis can be attributed to the properties of a single memory system. As indicated in the introduction, we will base our model on the declarative memory system of the ACT-R architecture (Anderson, 1990, 2007), which has proved accurate in modeling many different aspects of human cognition. Instead of using the full ACT-R architecture, we have used only the time

Table 2 Results of fitting mixed-effect models to the estimates of the long interval in the three conditions

	FF Condition		RD Condition		DR Condition	
Factor	Beta (*SE*)	*t*	Beta (*SE*)	*t*	Beta (*SE*)	*t*
Intercept (ms)	1,245 (131)	9.5***	568 (98)	5.8***	906 (107)	8.5***
Impact of the previous long intervals on the 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***
$long\text{-}fb_{n-1}$ = "too long" (ms)	−108 (35)	−3.1**	−237 (26)	−9.2***	−170 (26)	−6.6***
$long\text{-}fb_{n-1}$ = "too short" (ms)	114 (30)	3.8***	267 (19)	13.8***	220 (22.3)	9.8***
Impact of the previous short interval on the current long interval						
$short_{n-1}$	0.25 (0.03)	7.7***	0.16 (0.04)	4.3***	0.10 (0.04)	2.6*
$short\text{-}fb_{n-1}$ = "too long" (ms)	−87.2 (27)	−3.2**	−56.9 (25.3)	−2.2*	−74.8 (26.6)	−2.8**
$short\text{-}fb_{n-1}$ = "too short" (ms)	136.3 (29.1)	4.7***	80.9 (25.2)	3.2**	38.0 (27.8)	1.4

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

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

Experience	Pulses	t (s)	A_j	P_j	P_jV_j	Feedback	t FB	A_j FB	P_j FB	P_jV_j FB	Grand Sum
long$_{n-1}$	28	4.5	−1.67	0.029	0.8	−2 (late)	0.7	−0.74	0.37	−0.75	
short$_{n-1}$	22	7.5	−1.01	0.797	17.5	0 (correct)	3.7	−0.65	0.58	0	
long$_{n-2}$	27	11.1	−2.12	0.003	0.08	+2 (early)	7.3	−1.91	0.001	0.002	
short$_{n-2}$	20	13.8	−1.31	0.171	3.43	+2 (early)	10	−1.15	0.05	0.10	
Sum					21.8					−0.65	21.15

estimation and declarative memory theories of ACT-R, thus 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 that found in scalar expectancy theory (Gibbon, 1977, 1991). Indeed, even the pacemaker–accumulator setup is not critical to the model's performance, and could probably be replaced by other systems, as long as they produce some sort of representation of the passage of time (e.g., a measurement of memory decay, in the case of Staddon & Higa's, 1999, model). 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 s corresponds to approximately 17 pulses, and the long interval of 3.1 s 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 a 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.

Table 4 Free parameters in the pool model

Parameter	
Noise parameter t	0.2
Mismatch penalty between short and long	0.92
Feedback shift: How many pulses to add or subtract on the basis of feedback	1.8

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:

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

In this equation, t_{creation} is the time when 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 that, if one assumes that each trial is reflected in a separate chunk in
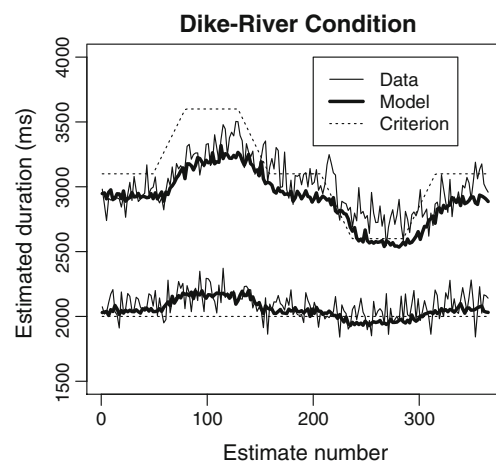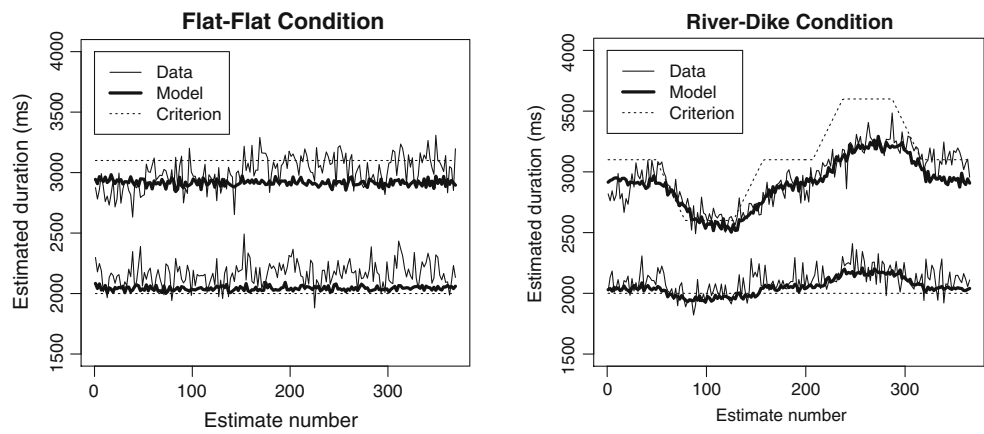

Fig. 3 Model fit of the DR condition

Flat-Flat Condition

River-Dike Condition

memory, more recent experiences that match the request
have the highest probability to be retrieved. This also
effectively implements a mechanism for forgetting: Even
though memory traces are technically not removed, their
influence can become so small over time that they are, for
all practical purposes, forgotten. The following equation
estimates these probabilities (where t is a noise parameter
and the summation is over all candidate chunks):

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

With the blending mechanism (Lebiere et al., 2007),
however, a weighted 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:

$$\text{Result value} = \sum_j P_j V_j. \qquad (3)$$

The consequence of this memory model is that it does not
make much of a difference whether the memory is based on one
or a great number of experiences, even though the latter model
might be slightly more accurate. This fits well with Jones and
Wearden's (2003) finding that the number of presentations of
a standard does not impact performance strongly.

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 process, 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 the feedback was "too

Table 5 Results of comparing the factors of mixed-effect models fitted to the data and of the cognitive model to 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)	1,217	1,296	962	992	826	1,000
Impact of the previous short intervals on the current short interval						
short_{n-1}	0.18	0.27	0.17	0.30	0.28	0.30
short_{n-2}	0.079	0.024	0.087	0.043	0.11	0.037
short-fb_{n-1} = "too long" (ms)	−106	−148	−96.8	−156	−153	−150
short-fb_{n-1} = "too short" (ms)	92.0	138	48.9	131	68.2	132
Impact of the previous long interval on the current short interval						
long_{n-1}	0.14	0.06	0.21	0.14	0.15	0.13
long-fb_{n-1} = "too long" (ms)	−130	−116	−97.5	−131	−83.8	−123
long-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 of the cognitive model to 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)	1,245	2,290	568	657	906	675
Impact of the previous long intervals on the current long interval						
$long_{n-1}$	0.28	0.14	0.50	0.50	0.46	0.51
$long_{n-2}$	0.12	0.003	0.17	0.12	0.17	0.11
$long-fb_{n-1}$ = "too long" (ms)	−108	−271	−237	−461	−170	−439
$long-fb_{n-1}$ = "too short" (ms)	114	153	267	315	220	309
Impact of the previous short interval on the current long interval						
$short_{n-1}$	0.25	0.13	0.16	0.18	0.10	0.18
$short-fb_{n-1}$ = "too long" (ms)	−87.2	−127	−56.9	−117	−74.8	−113
$short-fb_{n-1}$ = "too short" (ms)	136.3	8.9	80.9	46.5	38.0	17.9

long," it stores a negative value; and if the feedback was "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 the 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 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 Eq. 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 by the number of pulses, gives the contribution of that experience to the blend in the sixth column of the table. These contributions are added up (Eq. 3) to produce the result of the blended retrieval, 21.8. This process is repeated for the feedback (summarized in the next five columns), leading to a contribution of −0.65. The sum of the two retrievals is 21.15, which, rounded down to 21, means that the estimate for the next short interval will be slightly shorter than the previous one, 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.
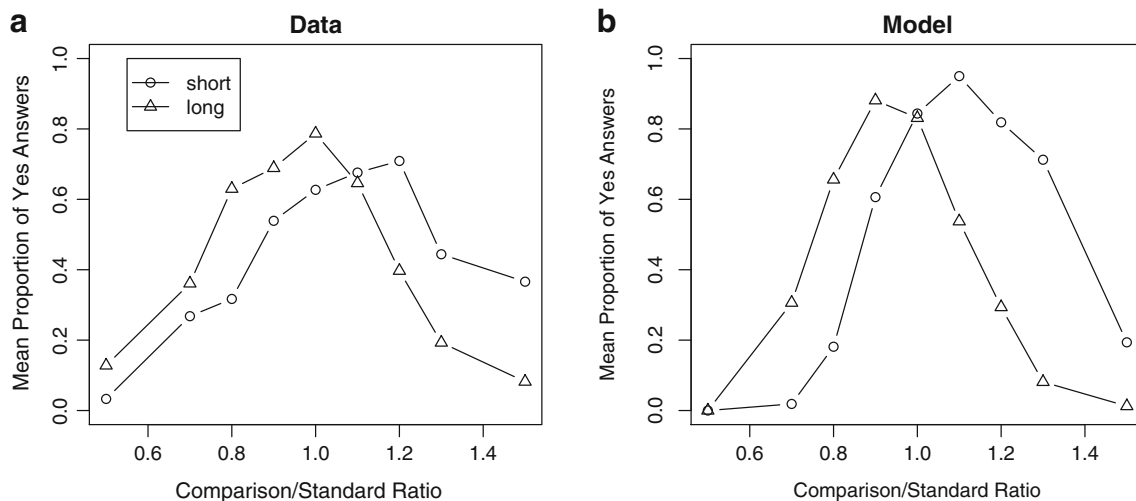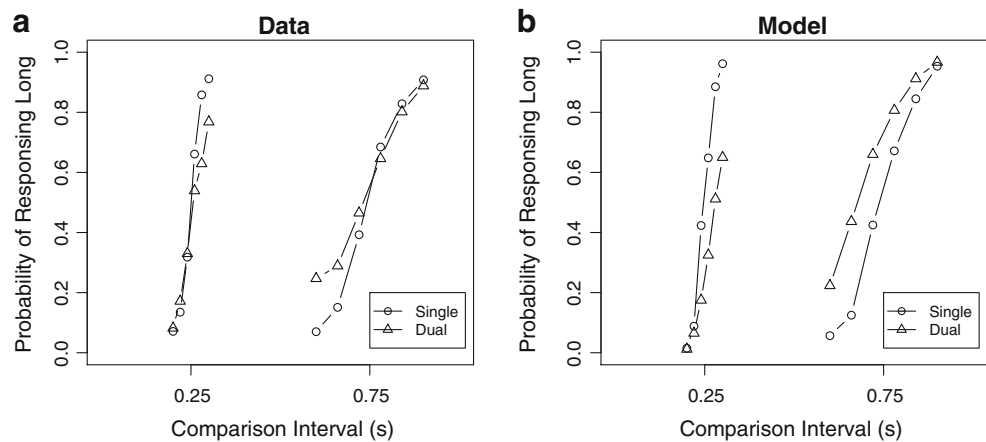

Fig. 5 (**a**) Results from Experiment 2 of Jones and Wearden (2004), and (**b**) the model fit to these data

Fig. 6 (a) Results from the
single condition and the two-
base-durations mixed (*dual*)
condition of Experiment 1 of
Grondin (2005), with visual and
auditory modalities averaged,
and (b) the model fit to these
data

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 the values from Van Rijn and
Taatgen (2008; $t_0 = 100$ 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 the DR (dike–river) condition.[2]

Figure 3 depicts the results of the model and shows how
they fit the data. In the model results, there is an impact of
the long on the short interval that is similar to what was
seen in 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 for
evaluating 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 had used to
analyze the data to check whether the same factors that
drove the estimates in the data also do so in the model.

Figure 4 shows the model predictions for the FF and RD
conditions. 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.

To better assess the quantitative aspects of the fit, we
applied the same regression analysis to the model outcomes
that we had used to analyze the data. In doing so, we could
check whether the same factors that played a role in
producing the estimates for the subjects played the same

role in the model's estimates. Tables 5 and 6 show this
comparison for the short and long intervals, respectively.

The tables do not provide us with a neat summarizing
number that tells us the quality of the fit, but for process
models like ACT-R, as opposed to more mathematical
models, there is no easy way to balance the number of free
parameters with the measure of fit (for discussion on this
topic, see Navarro, Pitt, & Myung, 2004; Roberts &
Pashler, 2000; Schunn & Wallach, 2001). The tables do
show that the same factors that were important in predicting
the estimates in the data play similar roles in the estimates
made by the model. In many, but not all, cases, the model
betas are very close to the values found in analysis of the
data. Only in the fit of the long estimates do 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
too stable a representation of the long interval. This
problem might also be due to the fact that there seems to
be a gradual overall drift in the estimates in the data. This is
visible in the FF condition in Figure 4, where we can see
that the estimates for both intervals tend to increase. We do
not have a clear explanation for this drift.

Can the model fit other data?

As we indicated in the introduction, time perception is
typically studied with either a production or a comparison
paradigm. To show that the model can also fit data from
comparison experiments, we fitted data from both the Jones
and Wearden (2004) and Grondin (2005) experiments.

In the double condition of Experiment 2 of Jones and
Wearden (2004), subjects were presented with three
examples each of a tone of short duration (drawn from
300–500 ms) and a tone of long duration (drawn from 600–
1,000 ms). Subjects were then presented with a series of
tones of which they had to judge whether their duration was
different from or equal to the standard tone that had the

---

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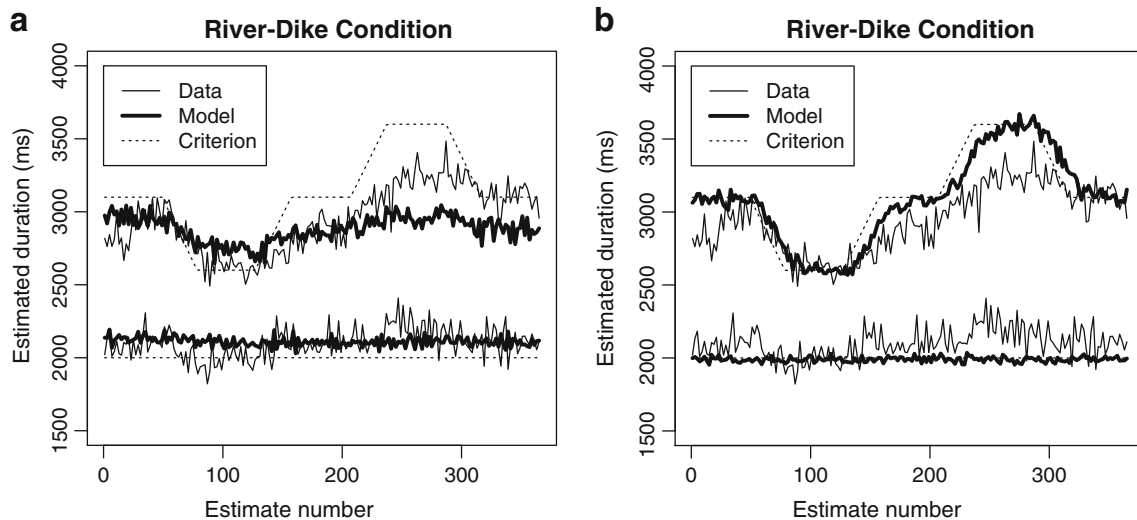data from the other two conditions were analyzed and modeled.

Fig. 7 Alternative model fits for the RD condition. (**a**) Solid-representation model. (**b**) Perturbation model

same frequency. Thus, if the short tone had a low frequency and the long tone a high frequency, a subsequent low-frequency tone had to be compared with the short tone. The comparison tones had a duration of 0.5, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, or 1.5 times the standard tone. Figure 5a shows the results. Clearly, subjects judged an interval to be the same as the standard more as it became closer to the standard. However, there was a bias, in that intervals that were slightly longer than the short standard were judged as the same more often, while the reverse was true for the long standard (the two curves would be on top of each other, otherwise).

To model this, we used the same pool memory principle as with the main experiment. The six examples would all be entered into the memory pool at the appropriate moment of learning. If the model subsequently had to judge whether a newly presented interval was of the same duration as the indicated standard, it would retrieve that standard from memory and compare it to the presented interval. If the difference in pulses between the two was small enough, the model would judge the intervals to be the same, and otherwise to be different.[3] Figure 5b shows the fit of the model, which exhibits a bias in judging short intervals as longer and long intervals as shorter that is similar to the one found in the data.

In a comparable experiment by Grondin (2005, Exp. 1), subjects had to judge whether newly presented intervals were longer or shorter than one of two standards (250 and

750 ms). In the single condition, they based their decisions on only a single standard, but in the double condition, they had to make comparisons with both standards. Figure 6a shows the results for the "single" and the "two base durations mixed" conditions from Grondin's Experiment 1 (the other conditions explored input modality, which is of no particular concern here).

In the dual condition, there is a clear tendency to judge the intervals that are longer than the short standard as long less often, while judging the intervals shorter than the long standard as long more often, indicating that the two representations contaminate each other.

What is different between this experiment and the Jones and Wearden (2004) experiment is that subjects were never explicitly presented with the standard, but instead received examples that they had to judge as longer or shorter than the standard. After their judgment, they received feedback on their correctness, allowing them to construct a representation of the standard on that basis. We modeled this procedure by entering every experience with feedback into the memory pool. When the model then had to judge whether a new interval was long or short, it would retrieve two representations from memory: one for long and one for short. It then decided for the category that was closest to the new interval. The results (Fig. 6b) show that the representations in the dual condition indeed influence each other in the same way as in the data.[4] Summarizing, these fits show that the here-presented model can qualitatively and quantitatively account for a series of experiments in which temporal reference memory drives temporal performance.

---

[3] Because the intervals in this experiment were so short, a t_0 of 100 ms was too coarse. We therefore used parameters from the Taatgen et al. (2007) article to fit these data: $t_0 = 11$ ms, $a = 1.1$, $b = 0.015$. Furthermore, two intervals were judged to be the same if they differed by less than 1.8 pulses, and the mismatch penalty between short and long was set to 0.4. The other parameters were left the same.

[4] The parameters we used in this model were the same as those for the Jones and Wearden (2004) model, except that the mismatch penalty between long and short was 0.8.

Table 7 Comparison between the data and the three models for the short interval in the "river–dike" condition

Factor	Beta, Data	Beta, Pool Model	Beta, Solid-Representation Model	Beta, Perturbation Model
Intercept (ms)	962	992	1,851	1,551
$short_{n-1}$	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

General discussion

Our experiment shows explicitly that representations of both temporal intervals in memory influence each other, supplementing findings by, for example, Jones and Wearden (2004), Grondin (2005), Van Rijn and Taatgen (2008), and Jazayeri and Shadlen (2010). The results not only show that the representations of two intervals tend to shift toward 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 our cognitive model is a simple memory model that is based on Anderson's (1990) rational analysis theory, which has been used to model many memory phenomena (see, e.g., Anderson & Matessa, 1997; Anderson & Reder, 1999; Taatgen & Wallach, 2002; for a more extensive list, see 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 was 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 SAMP and AVE models (Jones & Wearden, 2003), retrieves a single past experience (using Eq. 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. This alternative can account for the impact of one interval on the other because of the possibility that the model may retrieve 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 strong representation of each interval (as the reinforcement results in a "the winner take all" situation), making this model too sluggish to track changes in the long interval, because the new memory trace with the correct new value cannot compete with the established representation. The model also, therefore, cannot explain how such changes impact the short interval.

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

To demonstrate these differences, we implemented versions of both the solid-representation and perturbation models. The resulting estimates for the RD condition are shown in Figure 7. 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
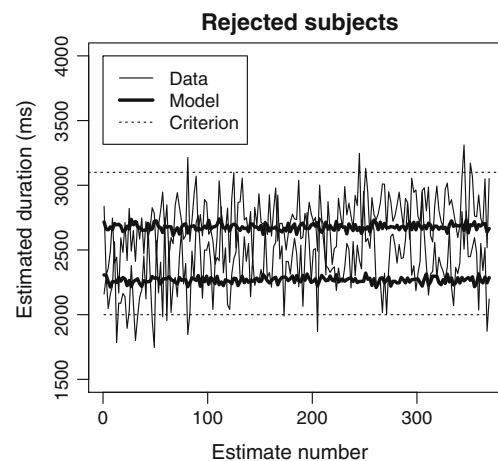

Fig. 8 Comparison between the model with a reduced mismatch penalty and the results of the 10 subjects who were removed from the original analysis

follow the changes in the long interval very well. The perturbation model follows those changes very well (too well), but it is not able to capture the contamination of the two intervals. A comparison of the factors in the regressions confirms these observations: Table 7 shows the 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 with the long interval. Moreover, both alternative models have a much too high intercept, indicating that previous experiences have less impact than is 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) argued, 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 could (1) produce relatively 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.

Even though the memory model we propose is fairly successful in explaining the memory phenomena associated with time perception, we do not claim that it is the only possible model. Indeed, several other memory models share characteristics with it, and might therefore produce similar results. For example, the SIMPLE model (Brown, Neath, & Chater, 2007) also has an exemplar-based approach that mainly differs in that there is no explicit decay, but instead a temporal ratio between experiences. The blending mechanism, however, effectively implements something that is very close to a temporal ratio. Similarly, the temporal-context model of Howard and Kahana (2002) has characteristics that would make it potentially suitable to model the memory aspect of time perception.

The advantage of a representation of time intervals based on a pool of experiences is that it is 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, he or she has 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 from a busy curved road. In an experiment in which addresses have to be typed into a navigation device while driving in a simulator, Salvucci, Taatgen, and Kushleyeva (2006) found that subjects adapted the time interval they spent on the navigation device to the difficulty of the driving task. Such an adaptation would be harder to accomplish if intervals were represented as single, solid representations.

One of the problems with our experiment, and with other experiments involving multiple time intervals, is that many subjects had to be removed from the data set because they could not keep the representations for the two intervals separate. Indeed, Whitaker, Lowe, and Wearden (2003) found that rats can only separate two intervals if they are in a ratio of at least 1:4. The mechanism that keeps the model from mixing up two intervals is the mismatch penalty in Eq. 1, so lowering the penalty would cause the model to mix up the two intervals. Figure 8 shows a comparison between a model in which the mismatch penalty is lowered to 0.3 and the 10 subjects who were rejected because they mixed up the intervals. Even though this is 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.

An important issue is the generality of the proposed models. For the model fits presented in this article, the main mechanism is 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 could model these data. As we discussed, a model with characteristics similar to those of the pool model might also serve the same role. 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. Despite that, we have shown that a combination of a memory system and a time estimation system can explain the data discussed in this study, despite the fact that neither system was specifically designed for these experiments.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

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

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

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,* 390–412. doi:10.1016/j.jml.2007.12.005

Baudouin, A., Vanneste, S., Pouthas, V., & Isingrini, M. (2006). Age-related changes in duration reproduction: Involvement of working memory processes. *Brain and Cognition, 62,* 17–23. doi:10.1016/j.bandc.2006.03.003

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

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576.

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

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

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.* New York: Springer.

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,* 18–33.

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

Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation, 22,* 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,* 869–879.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience, 13,* 1020–1026.

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

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

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

Lebiere, C., Gonzalez, C., & Martin, M. (2007). Instance-based decision making model of repeated binary choice. In R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 67–72). London, U.K.: Taylor & Francis.

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

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

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

Michon, J. A. (1967). *Timing in temporal tracking.* Assen, The Netherlands: Van Gorcum.

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. *Quarterly Journal of Experimental Psychology, 62,* 909–924.

Ogden, R., Wearden, J., & Jones, L. (2008). The remembrance of times past: interference in temporal reference memory. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1524–1544.

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

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

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

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

Salvucci, D. D., Taatgen, N. A., & Kushleyeva, Y. (2006). Learning when to switch tasks in a dynamic multitasking environment. In D. Fum, F. del Missier, & A. Stocco (Eds.), *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 268–273). Trieste, Italy: Edizioni Goliardiche.

Schunn, C. D., & Wallach, D. (2001). Evaluating goodness-of-fit in comparison of models to data. Retrieved November 14, 2005, from 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,* 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 D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 2010 International Conference on Cognitive Modeling* (pp. 247–252). Philadelphia, PA: Drexel.

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,* 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,* 163–204.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. doi:10.1017/S0140525X01000061

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs*, 77(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.

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

Whitaker, S., Lowe, C. F., & Wearden, J. H. (2003). Multiple-interval timing in rats: Performance on two-valued mixed fixed-interval schedules. *Journal of Experimental Psychology: Animal Behavior Processes, 29,* 277–291.

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