

Effects of feedback on time production errors in aging participants

Brian C. Rakitin^{a,b,*}, Chara Malapani^{c,d,1}

^a *Cognitive Neuroscience Division, Taub Institute, Columbia University, New York, NY 10032, USA*

^b *Department of Neurology, Columbia University, New York, NY 10032, USA*

^c *Department of Psychiatry, Columbia University, New York, NY 10032, USA*

^d *Department of Cognitive Neuroscience, New York State Psychiatric Institute, New York, NY 10032, USA*

Received 19 April 2006; received in revised form 9 July 2007; accepted 11 July 2007

Available online 3 August 2007

Abstract

In two experiments, healthy participants ages 60 years and older provided peak-interval time production data for two target intervals (6 and 17 s) over 2 days (baseline and retest sessions). In Experiment 1, three groups of participants were provided with two types of feedback during the baseline session that assisted either decision criteria setting or memory updating. During the retest session, run after a 24-h delay, each group received either one of the two types of feedback, or no feedback at all. Experiment 2 varied three additional groups' feedback during the baseline session only. Results indicated that the duration-dependent timing errors previously associated with aging did not occur during the retest session with the decision-criteria feedback regimen, or during the baseline session even in the complete absence of feedback. Thus, testing following the delay and without decision-criteria feedback are the necessary and sufficient conditions for the expression of the timing errors in aging. The efficacy of memory updating feedback could not be established. The discussion contrasts these results with the conditions that produce abnormal timing in Parkinson's disease patients in a similar procedure.

© 2007 Published by Elsevier Inc.

Keywords: Time production; Interval timing; Aging; Feedback

1. Introduction

In a previous report we presented evidence for a dysfunction of temporal memory associated with aging [28]. Healthy aging individuals (>60 years of age) were tested with the peak-interval timing procedure [26] in a 2-day experiment that omitted feedback 24 h after initial learning and production of the intervals [18]. Prior to the delay and with feedback, performance was comparable to young control participants, but after the delay and without feedback they over-produced a 6 s interval but accurately produced a 17 s interval. Age-associated, magnitude-dependent free-recall errors did not occur in two different line-length production tasks with designs and demands similar to the timing task. On this basis, we favored the conclusion that the timing errors indicated an age-related problem with temporal mem-

ory, although problems with other timing processes could not be ruled out. The two experiments with healthy aging participants reported here directly continue this line of research by addressing the following issues. Is the long-term retention of temporal memories sufficient for the expression of the aging deficit? Does the retention interval necessarily give rise to the aging deficit? How do the elderly use feedback to adjust their timing errors? The methods we use to address these questions are an increase in feedback during the retest session (Experiment 1), a reduction of feedback during the baseline (pre-retention interval) session (Experiment 2) and comparison of two different feedback regimens (Experiments 1 and 2), respectively.

Several reports have noted overproduction of intervals by aging individuals [2,3,24,30,34,35]. Though the methods used in these reports varied greatly, a consensus has emerged, reinforced by studies of time using other methods, e.g., [13,14], that aging impairs temporal attention. In the standard information-processing models of temporal cognition [33], such as Scalar Expectancy Theory (SET) [8,22] and the Attention Gate Model (AGM) [38], attention to time closes a "switch" or "gate" [12] allowing pulses from an internal pacemaker to increment the value of an accumulator. The accumulator

* Corresponding author at: 630 W 168th Street, P&S Box 16, New York, NY 10032, USA. Tel.: +1 212 305 7476; fax: +1 212 342 1838.

E-mail addresses: cm302@columbia.edu, br130@columbia.edu (B.C. Rakitin).

¹ Both authors contributed equally to this report. Order of authorship was determined by a coin-toss.

value serves as the subjective estimate of the currently elapsing interval's duration. Mismatches in the relation between real time and accumulator values, such as those caused by changes in attention to time, lead to timing errors and changes to timing variability that depend on when in the task (during acquisition or testing) distraction occurs or changes [27].

We rejected attention as a mechanism for the age-related errors in our previous study because the direction and magnitude of the errors depended on the target interval duration [28,29]. The effects of attention on timing can be expressed either as a constant delay to start timing at the beginning of an interval [22,27] or sporadic interruptions occurring regularly throughout the elapsing interval [4,12,23]. In either case, the effect on inattention to time on the longer of two intervals should equal or exceed that effect on the shorter, in contradiction to the observed effects. On the other hand, our two previous reports [17,28] showed an overall increase in timing variability in aging participants compared to younger adults like that reported in several other studies and attributed to inattention to time [2,13,21,30,35]. However, the age-related increases in timing variability occurred in both baseline and retest sessions, whereas the timing errors were restricted to the retest session [28]. Thus, if attention was the mechanism for the effect of aging on timing variability, it was unlikely to also be the source of age-related timing errors.

It is notable that *no* aspect of the standard internal clock model is capable of explaining the age-related error pattern from our previous report [28]. In SET, the best specified of the timing models, model parameters are set independent of the duration of the target time interval in order to account for the scalar property of timing variability, or the tendency for timing of different intervals to have, for example, a constant ratio between the standard deviation and mean (i.e., the coefficient of variation, or CV) of response latencies in time production tasks, e.g., [26,37]. Duration-independent parameters also help explain the fact that most experimental manipulations produce constant or monotonic changes across a range of target intervals [19,31]. This restriction applies to all the major sources of timing variability in the model [8,9] including the rate of the pacemaker, the operation of reference memory that stores accumulator values, and the decision process that compares accumulator values to reference memory in order to generate a timing response. This inability of SET to accommodate duration-dependent errors with different signs came to the fore in our attempts to understand the *migration effect* observed in Parkinson's disease patients [17,18].

When PD patients' peak-interval timing is tested in the "off" state (following 24 h of levodopa deprivation) they overproduce the shorter of two intervals (e.g., 6 or 8 s) but under-produce the longer of two intervals (e.g., 17 or 21 s). This is the migration effect. One model of this effect [19,31] proposes that dopamine deprivation causes the ordinarily near-linear increment of accumulator value with respect to real time to become curvilinear and accelerating when temporal memories are retrieved from memory. Thus, subjective time lags behind real time for short intervals and leads real time at longer values—producing the characteristic duration-dependent timing errors.

Attempts to explain the age-related duration-dependent timing errors using the curvilinear accumulator (or other) models of the PD migration effect are currently premature due to several important differences in the apparent causes of the aging and PD effects. First, the migration effect occurs following a retention period only in dopamine-depleted patients even if given ample feedback [17]. Second, in dopamine-depleted patients the migration effect occurs during the baseline session, before the retention period [18]. That is, the retention period is neither necessary nor sufficient to produce the migration effect in PD, while dopamine depletion is both necessary and sufficient. With these differences in mind, and as a predicate to future efforts at modeling the age-related timing errors, the goal of the current experiments is to determine whether delayed free-recall conditions, that is, the 24 h retention period and the omission of feedback, are in fact both necessary and sufficient to produce duration-dependent errors in aging participants. We restricted the present experiments to participants 60 years and older for this reason, and because timing errors in young participants are much smaller and less reliable [28].

There have been a few reports of the effects of feedback on timing changes in aging participants [2,35]. In these cases feedback attenuated problems with timing previously learned, verbally cued stimuli. The efficacy of feedback in these cases may reflect relearning of intervals learned in the more distant past, reducing the difference in the internal clock's function between acquisition and test [27,28,30]. These studies did not address the question of whether the delay between learning the interval and estimating the interval was related to the age-related changes. In Experiment 1, we test whether the retention interval is sufficient to produce duration-dependent errors in older adults by adding feedback to the retest session. Feedback may help participants compensate for a problem with temporal cognition associated with the delay by emphasizing use of a relatively unimpaired process, or it may serve to reinforce the process that is impaired by the delay. In either case, errors will not always follow the delay and the delay would not be sufficient to produce the errors.

Conditions in Experiment 1 that include feedback during the retest session are similar to those of previous studies [17,26], with the addition that here we separately examine the effects of two feedback regimens previously employed together that differ in the type and number of feedback trials. To be consistent with our prior reports the present design matches the frequency of retest session feedback trials to the frequency of the feedback trials in the baseline session, but not the number of trials between the two regimens. This design therefore allows us to test the relative efficacy of the existing regimens, but not the independent effects of feedback type and frequency.

The first regimen is comprised of the more frequent post-trial histograms, which indicate the placement of the participants' multiple peak-interval production responses within a given trial on a relative-time axis [26]. This feedback allows the participant to gauge whether they began and stopped responding too early or too late, that is whether their decision criteria were appropriately narrow, without informing the participant about the absolute duration of the interval. The second regimen con-

sists of the less frequent fixed time (FT), or “reminder,” trials that demonstrate the target interval [26]. FT trials allow the participant to update their memory for the interval but provide no information regarding their decision criteria. Differences in the efficacy of the two regimens would suggest which process, memory or decision, is the more important for either causing or maintaining the retest session errors. Alternatively, differences in errors may be attributed to the differences in the number of trials provided with each regimen, although this possibility does not exclude the former one.

In Experiment 2, we test whether the retention interval is necessary to produce the duration-dependent errors in aging by reducing feedback in the baseline session, where in previous experiments timing in aging participants has been accurate [28,29]. The three conditions in this experiment apply the feedback regimens from the retest sessions in Experiment 1 to the baseline sessions, and then retest production without feedback following a 24-h delay. If the delay is not necessary for the expression of the timing errors in older adults, we may observe the errors during the baseline session by reducing or removing feedback that is allowing participants to either compensate for or correct an impaired timing process. The results also will be useful in determining whether the baseline session conditions help determine the stability of temporal cognition across the retention interval.

The principle questions in Experiments 1 and 2 regard the conditions that give rise to timing errors, so the principle analyses examine systematic changes in the duration of produced intervals. However, we also provide parallel analyses for measures of intra-individual timing variability. In our previous study [28] aging was associated with increased timing variability, but the delayed free-recall conditions that gave rise to timing errors of interest here did not specifically alter the age-related increase in timing variability. However, it is always possible that timing variability will differentiate the new conditions in Experiments 1 and 2 in an informative way, in addition to providing an opportunity to replicate the relatively poor variability scaling previously observed in aging participants.

2. Experiment 1

The goal of this experiment is to examine the effects of the feedback regimens provided to participants during the baseline sessions on performance during the retest sessions. This manip-

ulation will address two questions. First, does the timing deficit in aging participants resist feedback during the retest session? That is, is retesting timing after a retention interval a condition sufficient to elicit the aging deficit, or is the absence of feedback required as well? Second, will the aging deficit show selective sensitivity to one or both of the feedback regimens?

2.1. Method

2.1.1. Participants

Forty-eight elderly participants were recruited from senior centers in Manhattan, NY. Interested participants were screened for neurological or psychiatric disorders or use of psychoactive medication. All participants were determined to be free of dementia, having obtained above 125 on the Mattis Dementia Rating Scale (DRS) [20] and above 45 on the modified Mini-Mental Status Exam (mMMS) [32]. Table 1 provides demographic details. Participants were randomly assigned to one of three experimental feedback type groups—“no feedback” (NO), “fixed time” (FT), and “feedback” (FB). Data from the NO group was previously reported as aging group 2 in [28] to demonstrate replication of the basic aging effects.

2.1.2. Apparatus

Testing was conducted on an Apple Macintosh Powerbook G3 laptop computer with a 14.1 in. color display. Participants were tested in senior citizen centers located around Manhattan. A custom-designed computer application was used to deliver the stimuli and record the responses for the time production task. Responses were recorded through the keyboard.

2.1.3. Procedure

On Day 1 of the experiment, participants provided informed consent and were administered the neuropsychological tests and the baseline session of the experiment. On Day 2 participants were administered the retest session of the experiment and were compensated for their participation upon completion. The New York State Psychiatric Institute IRB and the Columbia University Medical Center IRB approved all procedures.

2.1.4. Tasks

The time production task was a variant of the human peak-interval task [26]. Production of the 6 and 17 s target intervals occurred in separate blocks. The order in which the

Table 1
Participant demographics by group and experiment

Group	<i>N</i>	Age range	Mean age	% Female	Education	mMMS	DRS
Experiment 1							
NO	16	62–83	70.2 ± 6.2	43.75	13.2 ± 3.5	49.1 ± 7.5	137.1 ± 8.1
FT	16	66–89	75.9 ± 7.0	56.25	13.3 ± 2.7	47.6 ± 3.8	137.5 ± 5.8
FB	16	60–88	74.6 ± 8.3	56.25	12.8 ± 2.6	48.1 ± 6.4	136.1 ± 6.8
Experiment 2							
NO	16	61–87	70.9 ± 7.7	75.0	12.3 ± 3.2	51.4 ± 4.2	139.9 ± 4.6
FT	16	62–81	71.4 ± 6.3	75.0	12.3 ± 4.5	51.0 ± 4.4	138.9 ± 6.4
FB	16	63–86	74.1 ± 6.6	93.75	12.3 ± 2.0	51.2 ± 4.3	139.5 ± 4.1

Note: Values for age, education, mMMS, and DRS are the mean ± S.E. Education is in years.

intervals were tested was counterbalanced across participants within group, and the order was the same for the baseline and retest sessions.

The baseline session for each interval consisted of 80 trials with a composition consistent with our prior reports [17,18,26]. The first 10 “fixed time” (FT) trials demonstrated the target interval. On each trial a blue square was presented until the target interval had elapsed, at which time it turned magenta for 1 s and then terminated. Participants were instructed to remember the duration of the blue square. The next 10 trials were “peak-interval” (PI) trials, where participants produce the target interval. Participants were instructed to respond on each trial with a bout of at least four space bar presses centered in time on the end of the target interval, although trials were included in the analyses if at least two responses were made. All responses were recorded. Following the responses, the participants were shown feedback in the form of a graph indicating their response times relative to the target interval, whether all responses were within 10% of the target interval, and whether sufficient responses were emitted. These trials were considered practice and were not included in the data analysis. The remaining 60 trials consisted of a random sequence of PI trials without feedback, PI trials with post-trial graph feedback (i.e., “FB” trials), and FT trials (see Table 2). Prior to each trial a message was presented indicating whether it is a “test” trial (peak interval), in which participants needed to respond, or a “reminder” trial (fixed time), in which participants did not respond, but were reminded of the given interval of that block.

The retest session on Day 2 differed for each of the three experimental groups. There were a total of 60 trials, and the number of either FB or FT trials was the same as during the baseline session (see Table 2). This design favors a balance between sessions over equal probability of feedback trials between the FB and FT groups. The “NO” feedback group received all PI trials without any feedback. The “FT” group received a random sequence of PI trials without feedback and FT trials. The “FB” group received a random sequence of PI trials with and without post-trial graph feedback.

For the entire procedure, participants were instructed to refrain from counting or tapping. To further discourage counting, random digits were intermittently superimposed over the

blue square [26]. Inter-digit times were determined randomly using two uniform distributions, 200–700 and 1300–1800 ms. All trials included the digit distracters, and participants were instructed to name the digits aloud. Performance was qualitatively monitored by the experimenters, but not recorded. The digit distracters have been used in some previous studies using the human peak-interval procedure, e.g., [16–18,26,28], and may add attention demands to timing [28] in addition to their role in reducing chronometric counting [10,11,26].

2.1.5. Analysis

All trials with two or more responses were included in the analysis, and all responses from included trials contributed to the summary variables. Excluded trials constituted less than 1% of all trials in each group.

The mean, or “middle,” of the earliest (i.e., “start”) and latest (i.e., “stop”) response latencies summarized time production performance for each PI trial (with or without feedback). The median (MD) and coefficient of variation (CV or the semi-interquartile range divided by the MD) of the trial middles summarized time production within each subject \times duration \times session condition. As noted previously [29], the CV is typically calculated as the standard deviation divided by the mean and hence would be more correct to refer to our current measure as the “non-parametric” CV, or NPCV. However, we decided to keep the CV nomenclature in this report to be consistent with its precedents, e.g., [17,26,28]. The MD was used to test hypotheses concerning production accuracy and the CV was used to test hypotheses concerning intra-individual variability, including the scalar property of timing variability, which predicts that target interval duration should not affect the CV.

Mixed-model ANOVAs tested hypotheses concerning the group mean values of the condition MD and CV. Each model included feedback-types group as a three-level between-participant factor, and two within-participant factors with two levels, target interval and session, in a full-factorial design. Planned contrasts within these models compared the FT and NO groups, and the FB and NO groups, and tested the session \times duration interaction within each of the three groups.

2.2. Results and discussion

Table 3 presents the group means and standard errors for the MD (minus the target interval duration) and CV of middle responses. Fig. 1 presents the group average response distributions. Panels A, C, and E present the data in absolute time, and are useful for examining the effects of feedback on time production accuracy. Panel A shows the results for the NO group that received no feedback during the retest session. While accuracy is generally good during the baseline session (indicated by the fact that the peak of the open-symbol functions lie close to the target intervals), participants’ productions of the short interval were too long during the retest session. This effect constitutes a replication of a previously reported effect on aging in delayed free-recall time production [28]. The results from the FT group (panel C) are much the same, except this group’s productions of

Table 2
No. of trials of each type per session, by experiment and group

Group	Day 1 (baseline)			Day 2 (retest)		
	FT	PI w/FB	PI w/o FB	FT	PI w/FB	PI w/o FB
Experiment 1						
NO	15	30	15	0	0	60
FT	15	30	15	15	0	45
FB	15	30	15	0	30	30
Experiment 2						
NO	0	0	60	0	0	60
FT	15	0	45	0	0	60
FB	0	30	30	0	0	60

Note: All groups in both experiments received 10 FT trials and 10 PI trials with feedback preceding the listed mixture of 60 trials for the baseline session.

Table 3
Group mean accuracy and variability for all conditions and experiments

Group	Target	MD – target		CV	
		Baseline	Retest	Baseline	Retest
Experiment 1					
NO	6	0.28 ± 0.15	2.00 ± 0.32	0.116 ± 0.009	0.126 ± 0.020
	17	−0.47 ± 0.40	−0.78 ± 0.70	0.097 ± 0.006	0.109 ± 0.013
FT	6	0.58 ± 0.21	2.50 ± 0.59	0.154 ± 0.013	0.127 ± 0.008
	17	1.36 ± 0.48	0.84 ± 0.48	0.113 ± 0.013	0.147 ± 0.012
FB	6	0.23 ± 0.20	0.39 ± 0.23	0.130 ± 0.013	0.129 ± 0.011
	17	−0.84 ± 0.38	−0.45 ± 0.31	0.105 ± 0.013	0.105 ± 0.004
Experiment 2					
NO	6	0.73 ± 0.25	1.99 ± 0.51	0.132 ± 0.010	0.112 ± 0.008
	17	0.51 ± 1.28	−1.55 ± 0.96	0.119 ± 0.011	0.105 ± 0.008
FT	6	0.70 ± 0.39	2.89 ± 0.60	0.123 ± 0.015	0.112 ± 0.010
	17	0.27 ± 0.99	−0.03 ± 1.00	0.129 ± 0.016	0.118 ± 0.013
FB	6	0.24 ± 0.11	1.99 ± 0.44	0.136 ± 0.011	0.136 ± 0.014
	17	−0.14 ± 0.38	−1.48 ± 0.84	0.124 ± 0.013	0.114 ± 0.013

Note: Values are the group mean ± S.E. We have subtracted the target value from the MD in order to emphasize the direction of the errors. Statistical analyses were carried out on the uncorrected values.

the long interval were also too long. However, this tendency is present during both sessions, and is therefore the likely result of random variability rather than an effect of feedback. In contrast to the other two groups, the FB group showed no appreciable changes in timing accuracy between sessions (panel E).

A significant session × duration × group interaction on the MD ($F(2, 45) = 4.11, p < 0.05$) indicates that these observations are reliable. Moreover, planned contrasts within the three-way interaction indicated that the session × duration interaction on the MD was significantly bigger for the NO group than the FB group ($F(1, 45) = 5.08, p < 0.05$). This result primarily reflects the FB group's smaller 0.16 s difference between retest and baseline in production of the 6 s target compared to the NO group's 1.72 s shift, and the relatively small retest–baseline differences in production of the 17 s interval of −0.31 and 0.39 s for the NO and FB group, respectively. In contrast the session × duration effect was approximately equal for the FT and NO groups ($F(1, 45) = 0.17, n.s.$), both having large positive retest–baseline differences for the 6 s target (1.92 s and 1.72 s, respectively) and small differences for the 17 s target (−0.52 and −0.31 s, respectively). Simple-effects tests indicated that the session × duration interaction was significant for the FT ($F(1, 15) = 9.82, p < 0.05$) and NO groups ($F(1, 15) = 6.07, p < 0.05$), but not for the FB group ($F(1, 15) = 0.23, n.s.$), indicating an absence of duration-dependent errors in the retest session only for the FB group.

The following effects were also significant, but secondary to the aforementioned session × duration × group interaction: group ($F(1, 45) = 9.15, p < 0.05$), session ($F(1, 45) = 6.80, p < 0.05$), duration ($F(1, 45) = 25.35, p < 0.05$), duration × group ($F(2, 45) = 3.42, p < 0.05$), and session × duration ($F(1, 45) = 12.08, p < 0.05$). The session × group interaction was not significant ($F(2, 45) = 0.44, n.s.$).

Based on these results we conclude that aging individuals can compensate for their intrinsic tendency towards duration-dependent timing errors in delayed free recall by adjusting their decision criteria based on post-trial feedback designed to provide

just such an opportunity. In contrast, providing aging individuals an opportunity to directly update their memory of the target time interval seems to have no effect on performance. It is unclear, however, whether the apparent difference between the FT and FB groups indicates that in aging participants, problems in free-recall timing are resistant to memory updating, or if the number of trials provided were insufficient. We revisit this issue in the general discussion.

The effects of feedback during the retest session on the expression of the scalar property of timing variability can be observed in panels B, D, and F in Fig. 1. When the scalar property is expressed, the normalized response distributions should *superpose*, or lie over each other exactly, e.g., [6,18,26]. Overall, superposition is generally present. However, the short-interval distributions (circles) are slightly wider than the long-interval distributions (triangles) for most groups and conditions (with the exception of the FT group's retest session, see below), indicating violations of the scalar property in the form of greater relative variability in productions of the short interval. The tendency for short-interval production to be more variable than long interval production is also evident in the CVs presented in Table 3, and the reliability of this tendency is confirmed by a significant effect of duration on the CVs ($F(1, 45) = 11.69, p < 0.05$).

In contrast to the other groups and conditions the FT group's retest session (Fig. 1, panel D) productions of the long interval (closed triangles) are relatively more variable (i.e., wider) than productions of the short interval (closed circles). A similar pattern can be observed in the CVs presented in Table 3. The FT group's unique pattern of CVs in the retest session is reflected in a significant overall session × duration × group interaction ($F(2, 45) = 3.34, p < 0.05$), and a significant planned contrast between the FT and NO groups' session × duration effects ($F(1, 45) = 4.91, p < 0.05$). All other tests of effects on the CVs were not significant. While the mean CVs in Table 3 suggest a violation of the scalar property in the FT group's retests session that is opposite of that in the baseline session, post hoc tests

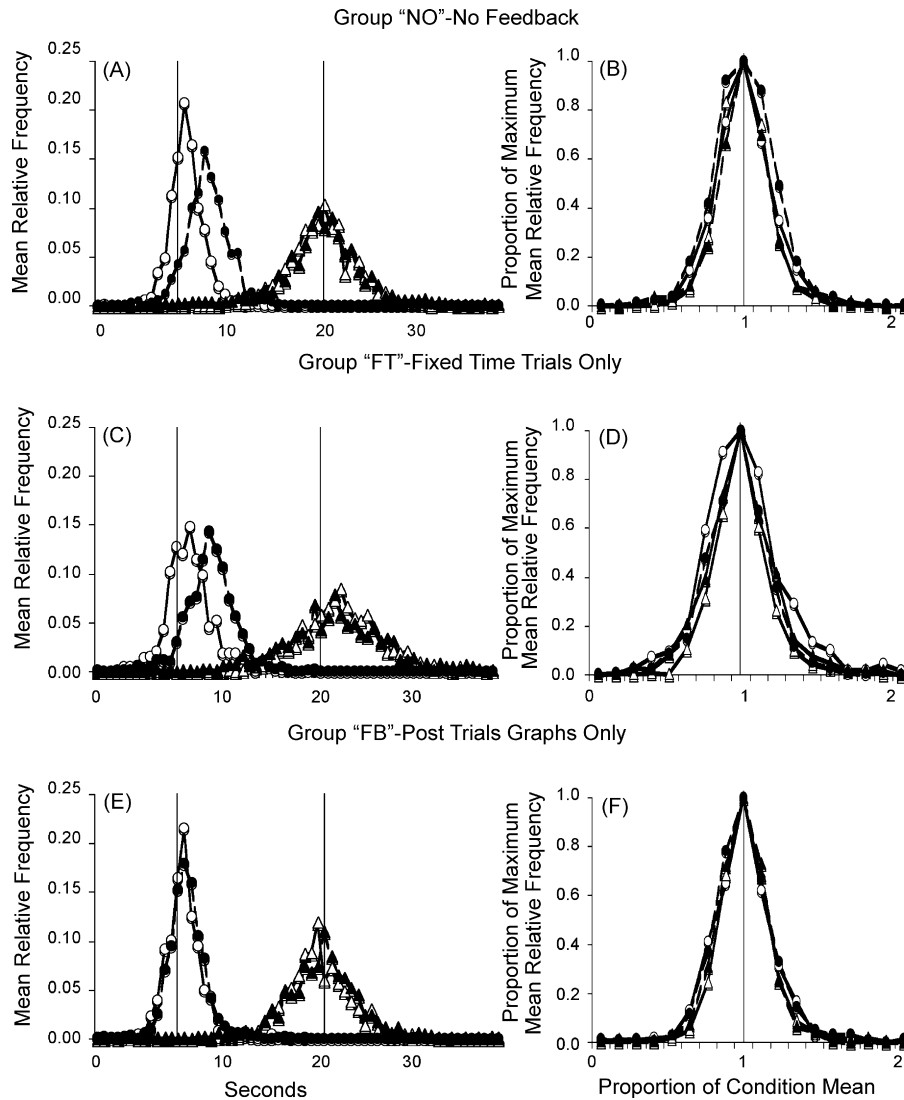


Fig. 1. Results of Experiment 1. This figure shows the group-average absolute response latency frequency histograms in real time (panels A, C, and E) and the group-average relative response latency frequency histograms in relative time (panels B, D, and F). The four functions in each panel show the results for the baseline session, 6 s target (open circles), baseline session, 17 s target (open triangles), retest session, 6 s target (closed circles), and the retest session, 17 s target (closed triangles). The vertical lines indicate the target interval.

indicated a violation of the scalar property in the baseline session ($F(1, 15) = 7.17, p < 0.025$) but not in the retest session ($F(1, 15) = 2.56, n.s.$). That is the FT group shows the typical pattern of greater relative variability for production of the short interval compared to the long interval during the baseline session, but exhibits the scalar property during the retest session.

In sum, the analyses of the CV data indicate that the conditions that change the (poor) expression of the scalar property are different from the conditions that affect the expression of duration-dependent timing errors. The FT and NO groups showed different patterns of variability scaling but similar patterns of timing accuracy across the two sessions. To put it another way, there were group \times session \times duration effects on both the MD and CV, but the sources of interactions differ. This finding suggests that the conditions that gave rise to the violations of the scalar property were independent of the conditions that gave rise

to the changes in timing accuracy. As such, the CV data cannot help us diagnose the information-processing source of the effect of feedback on timing accuracy.

We previously reported that in aging participants the CV of production of the short (6 s) interval was greater than the CV of production of the long (17 s) interval [28]. That previous report included the present NO group, so the current FB group replicated that previous report during both the baseline and retest sessions, while the FT group replicated the previous report only for the baseline session.

3. Experiment 2

This experiment varies the feedback available to participants during the main phase of the baseline session. This manipulation will allow us to address two outstanding issues. First, is the retention interval a necessary condition for the expression

of the timing deficit in aging participants? That is, will aging participants show the duration-dependent errors seen in delayed free recall in immediate recall if feedback is removed from the baseline session, making it more similar to the retest session? Second, will changing the feedback regimen during the baseline session affect the timing errors in the retest session? This issue can also be viewed as asking whether feedback during the baseline session affects the stability of temporal cognition across the retention interval.

3.1. Method

3.1.1. Participants

Recruitment of the 48 aging participants used the same methods described in Experiment 1 (see Table 1 for demographic details). Participants were randomly assigned to one of three experimental feedback type groups—“NO”, “FT”, and “FB”.

3.1.2. Apparatus

The apparatus was identical to the one used in Experiment 1.

3.1.3. Procedure

The procedure was identical to the one used in Experiment 1.

3.1.4. Tasks

The task was identical to the one used in the previous experiment, but here the groups received different feedback regimens during the baseline session, while feedback was omitted entirely during the retest session (see Table 2). In the baseline session, all groups first received 10 FT trials and 10 peak trials with post-trial graph feedback. Of the remaining 60 trials, the “NO” group received 60 peak trials without feedback, the “FT” group received a random sequence of 15 FT trials and 45 peak trials without feedback, and the “FB” group received a random

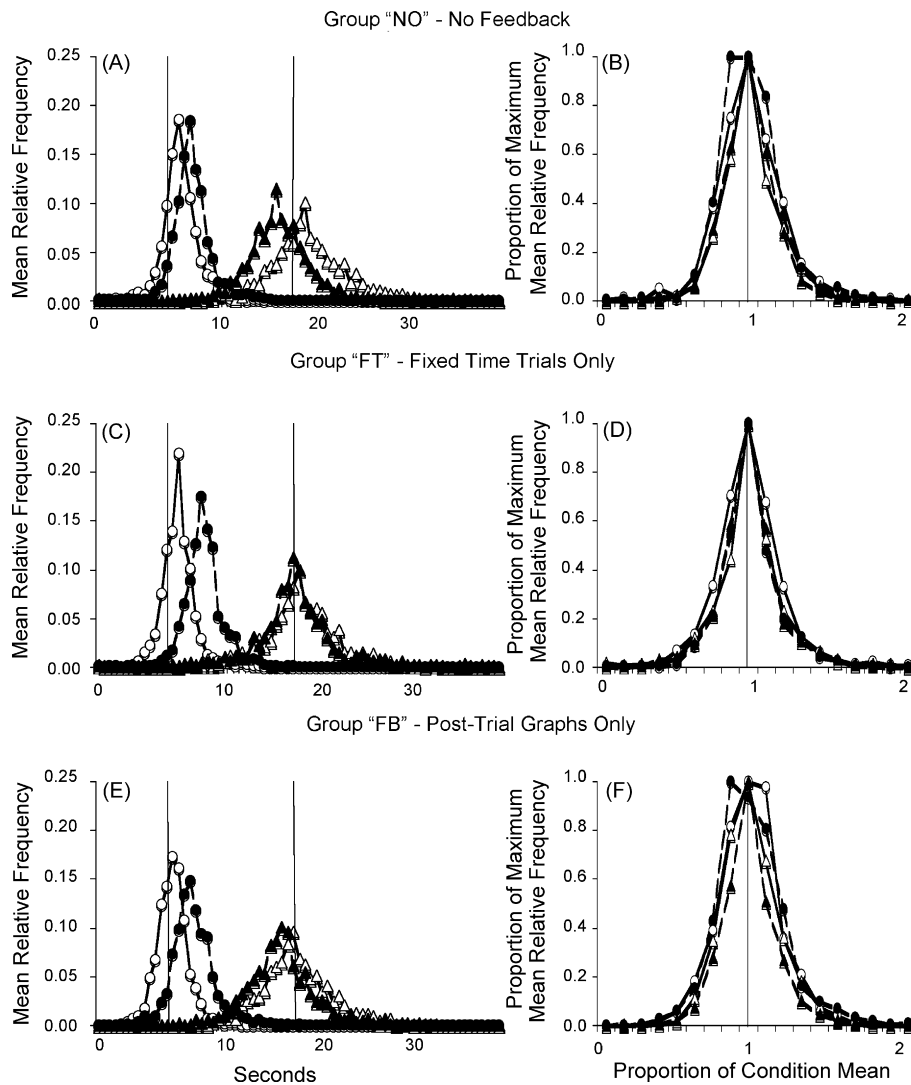


Fig. 2. Results of Experiment 2. This figure shows the group-average absolute response latency frequency histograms in real time (panels A, C, and E) and the group-average relative response latency frequency histograms in relative time (panels B, D, and F). The four functions in each panel show the results for the baseline session, 6 s target (open circles), baseline session, 17 s target (open triangles), retest session, 6 s target (closed circles), and the retest session, 17 s target (closed triangles). The vertical lines indicate the target interval.

sequence of 30 trials with post-trial graph feedback and 30 trials without feedback. The retest session on Day 2 was identical for all three groups and consisted of 60 peak trials without feedback.

3.1.5. Analysis

The initial analysis was identical to the one used in Experiment 1. Additionally, tests of the group \times duration interaction within session were conducted as planned contrasts of simple effects within the session \times duration \times group interaction. Post hoc comparisons employed two additional ANOVA models. The first compared the baseline session data among the six groups of Experiments 1 and 2 in order to determine whether the decreased feedback affected performance. The second model included both the baseline and retest session data from the “NO” group in Experiment 1 along with the data from the three Experiment 2 groups in order to compare the effects of the retention interval on retest session performance following the full and reduced baseline session feedback regimens.

3.2. Results and discussion

Table 3 and Fig. 2 present the data from this experiment. Beginning with the baseline session, we note little difference in accuracy between the groups (Fig. 2, panels A, C, and E, open symbols). The total average error (or the MD – the target interval duration) in production across the three groups is 0.38 s, compared to an average of 0.19 s in Experiment 1. This finding is reflected in non-significant tests of the effect of group ($F(2, 45) = 0.30$, n.s.), duration ($F(1, 45) = 0.42$, n.s.), and the duration \times group interaction ($F(2, 45) = 0.30$, n.s.) on the baseline session MD errors. Moreover, accuracy in the current three groups is similar to the baseline session data from Experiment 1 as indicated by non-significant tests of the effect of group ($F(5, 90) = 1.30$, n.s.), duration ($F(1, 90) = 1.42$, n.s.), and the duration \times group interaction ($F(5, 90) = 0.80$, n.s.) on the MD when all six groups are included. Apparently, the FT and PI trials with post-trial feedback that precede the baseline session are sufficient to produce stable and accurate performance in the immediately following PI trials of the baseline session. In other words, aging individuals show no impairment of temporal cognition in immediate recall.

All three groups show a clear pattern of duration-dependent errors in the retest session (Fig. 2, panels A, C, and E, closed symbols), and there were corresponding significant session \times duration ($F(1, 45) = 27.44$, $p < 0.0001$) and duration ($F(1, 45) = 18.34$, $p < 0.0001$) effects on the MD. The non-significant effect of the session \times duration \times group effect on the MD ($F(2, 45) = 0.30$, n.s.) indicates the equivalent effect of the retention interval on all three groups' duration-dependent errors. In addition, when the MD data from the current groups are compared to the “NO” group from Experiment 1 the test of the session \times duration \times group interaction ($F(3, 60) = 0.38$, n.s.) indicated no difference in the duration-dependent errors between baseline and retest sessions. (No other tests of effects on the MD were significant.) We conclude from these results that feedback during the baseline session has little impact on

duration-dependent errors that emerge after the retention interval during the retest session. That is, the stability of temporal cognition over 24 h seems unaffected by the quantity or type of feedback provided during the baseline session.

Considering together the accurate baseline session and impaired retest session time production accuracy, we further conclude long-term retention of temporal memory is a necessary condition for the expression of the duration-dependent error-pattern in healthy aging adults.

The timing variability data from this experiment (Fig. 2, panels B, D, and F, and Table 3) differed in an important way from our previous experiments with aging participants using the present methodology. There is no overall violation of the scalar property of timing variability, averaged across groups and sessions. That is, there was no significant effect of duration on the CV ($F(1, 45) = 1.62$, n.s.). The only other experimental effect on the CV was a marginal reduction of variability from the baseline to the retest session ($F(1, 45) = 4.05$, $p = 0.0502$).

Closer inspection of the CV values in Table 3 suggests an alternative explanation for the null effect of duration in Experiment 2. While the NO and FB groups' CVs were larger for the short interval than for the long interval in both sessions – following the pattern established in Experiment 1, excluding the FT group's retest session, and in our previous study [28] – the FT group's CVs were larger for the long interval than for the short interval in both sessions. In fact the test of the main effect of duration in Experiment 2 excluding the FT group is significant ($F(1, 30) = 4.21$, $p < 0.05$), while the test of the effect of duration within the FT group is not ($F(1, 15) = 0.41$, n.s.). (In the analysis excluding FT group there were no effects involving group, and for neither analysis was the session \times duration interaction significant.) In the absence of a group \times duration interaction involving all three groups this analysis must be considered exploratory. Nonetheless it suggests a unique role for the FT feedback in suppressing the violation of the scalar property in aging participants. Moreover, the pattern is not simply confounded with the amount of feedback provided because the FT group received a number of feedback trials intermediate between the FB and NO groups. However, it is still possible that there is a more complex (e.g., quadratic) interaction between feedback type and quantity that allows the scalar property to emerge in the FT group.

4. General discussion

The two experiments reported here provide several new pieces of information regarding the factors that control the expression of peak-interval time production errors in healthy aging adults. The results show that the retention interval is a necessary but not sufficient condition for expression of the duration-dependent errors in aging participants. It is necessary because the errors are seen only in the retest session, and not in the baseline session even when all feedback is removed. However, the retention interval is not a sufficient condition because the errors do not occur in the retest session in the presence of the decision-criteria feedback regimen. Therefore, the retention interval and the absence of the decision-

criteria feedback regimen following the retention interval are together the necessary and sufficient conditions for the expression of duration-dependent time production errors in aging adults.

Additional findings include the possibility that decision-criteria feedback is more effective than memory updating feedback in attenuating the retest session timing errors (Experiment 1), although this finding could also reflect the higher frequency of decision feedback. If subsequent experiments determine that some additional memory updating feedback leads to accurate retest session production by aging participants, then the sufficient conditions for the expression of duration-dependent errors in aging participants would be more properly stated as the complete omission of feedback, rather than just the omission of decision-threshold feedback. We intend to address this issue in a future study.

Because we did not include a young control group in this study, we cannot infer that these conditions would or would not determine the much smaller errors occasionally apparent in young participants. Nonetheless, because our previous studies linked the magnitude and reliability of duration-dependent errors to healthy aging we consider the present manipulations, results, and interpretations primarily relevant to healthy aging.

Among the studies of timing and aging, the present studies compare best to Rammsayer's study [30]. In that study, aging participants produced 1 and 15 s target intervals in a mixed-trials block with exemplars of the intervals provided on each trial. As indicated by ratio-transformed errors, aging participants significantly overproduced the short interval but were about equally accurate in producing the long intervals. This effect is reminiscent of the one we reported. However, attribution of Rammsayer's effect to age-related changes in temporal memory are difficult because the experimental design effectively minimizes the influence of reference memory [1,36]. The supplied explanation was that attention affected the short interval, but not the long interval, the timing of which was served by a different mechanism entirely. The basis for the distinction was that the short interval falls within the domain of the "psychological present," which ranges in duration up to 3–5 s [5,25].

While this argument has merit, single mechanism alternatives do exist [19,28,31]. In the current experiments and their immediate precedents, a single manipulation of memory – the introduction of a delay period – produces changes in one or both of a pair of longer intervals (>6 s) depending the duration of the delay period [28,29]. The effect of the retention interval suggests that alterations to temporal memory underlie the observed duration-dependent timing errors, without ruling out the possibility that some other mechanism supports timing and produces timing errors when the target intervals are shorter (<6 s). We also note that it is not currently clear how temporal memory produces duration-dependent errors. That mechanism is the subject of ongoing research, and further speculation about that mechanism is beyond the scope of the current discussion.

Experiment 1 demonstrated that decision-threshold feedback attenuated the retest session timing errors. Given that we propose that temporal memory is mechanism for the timing errors due

to the reliance of the appearance of those errors on the retention interval, we suggest that the decision-criteria feedback is more likely acting to help participants compensate for errors originating in temporal memory, rather than adjusting for problems originating in the decision process itself. That is, the feedback may be allowing participants to adjust for an underlying problem with temporal memory by changing their decision criteria to account for the distorting effects of the retention interval on memory for the duration of the target interval.

This proposal of compensation via the decision criteria for distortions in memory is not undermined by the uncertainty in the current data regarding the efficacy of memory updating feedback in attenuating retest session errors. To the contrary, it is a distinct possibility that with additional memory updating feedback participants would be able to acquire a new criterion for production taking into account distortions to temporal memory induced by the retention interval, and produce intervals without error. Such a demonstration would further indicate that the decision-criteria feedback after the retention interval in Experiment 1 is compensating for a dysfunction in temporal memory rather than the decision process. However, without proof that memory updating feedback can be effective, we must entertain age and delay-dependent dysfunction of the decision-making process as an alternative to the memory dysfunction hypothesis.

As mentioned in Section 1, the current experiments on timing in healthy aging participants and their precedents began as a comparison to effects found in aged PD participants. How stands this comparison in light of the current results? The current Experiment 2 provided firm evidence that the retention interval is a necessary condition for the expression of the aging deficit. This is decidedly not the case for the PD migration effect, which is quite strongly expressed during the baseline session and in the presence of abundant feedback if dopamine replacement therapy is withheld from the patients. In addition, we have investigated the possibility that the age related timing deficit is dopamine dependent [29] using the same design as our previous PD and aging studies [18,28], but with a 1-h delay between the baseline and retest sessions. The results indicated that levodopa and aging independently affected timing, reducing the likelihood that the cognitive mechanism underlying the PD and aging deficits is the same. Taken together, these findings suggest qualitatively different causes for the PD and age-related accuracy effects. This conclusion makes application of the curvilinear accumulator model of the PD migration effect [31] to the aging data impractical without substantial modifications. These modifications would have to include mechanisms by which a retention interval could affect accumulator characteristics, and explanations of how decision-threshold feedback could compensate for curvilinear accumulation in healthy aging, but not in dopamine-deprived PD patients.

The differences between the timing errors seen in normal aging and PD might indicate that different components of the interval timer ("clock") are affected in each case. This position would be tenable if a future experiment determined that decision-threshold feedback alone was effective in eliminating the retest session errors (complementing the findings

of Experiment 1), and concluded that aging impaired the decision-making process, rather than temporal memory. Such findings would contrast strongly with the altered memory encoding and decoding process underlying the PD effects [18,19,31], and bias against attempts to relate the aging and PD deficits at the level of the internal clock.

Conditions that control the expression of duration-dependent errors have not been completely specified for either healthy aging or PD. For example, it is not clear whether such errors are dependent upon either the absolute duration of the target intervals, or the difference between the two intervals. These two issues could be addressed in a future study by altering the absolute and relative values of the target intervals. If Rammsayer's proposal of different short and long interval timing mechanisms is correct one might expect retention intervals to have qualitatively different effects for pairs of intervals within the short (<1–3 s) range. Also, certain conditions known to give rise to the migration effect in PD, particularly the requirement that two target intervals be processed within a session, have yet to be tested in healthy aging by examining timing performance in sessions with one target interval. Similarly, conditions unique to aging, such as the requirement for a retention interval, have yet to be explored at the limit. Although we know that differences between young and old participants emerge with as little as a 1-h delay [29], future experiments on aging should include manipulations of the duration of the retention interval.

While the present study emphasized timing errors, we did conduct parallel analyses of timing variability. In our previous report [28], we found poorer variability scaling in the aging groups (including the NO group from the current Experiment 1) compared to the younger participants. We replicated and extended these findings of larger CVs for production of a 6 s interval than for production of a 17 s interval in 7 of 10 new combinations of feedback regimens and sessions between Experiments 1 and 2. Only those conditions with FT feedback alone, or without feedback following FT only sessions showed a different pattern.

The most important aspect of the timing variability data in the current experiments was the dissociation between the conditions that gave rise to changes in the generally poor expression of the scalar property and the conditions that alter expression of the duration-dependent timing errors. To the extent that inattention to time is the currently favored mechanism for explaining the failure of the scalar property in aging [2,14,21,30,35] this dissociation further reduces the likelihood that attention is the mechanism underlying the duration-dependent errors in aging, as in our previous report [28].

It is notable that our first report of peak-interval timing with healthy aging individuals [17] indicated excellent variability scaling for three target intervals, 8, 12 and 21 s as well as substantially lower CVs (approximately 0.065). That experiment included blocks with trials and feedback in quantities similar to the baseline sessions in Experiment 1, but the full feedback regimen was available on both testing days. The possibility therefore exists that while feedback initially results in poor scaling in aging individuals, extended exposure to feedback ultimately results in both excellent scaling and superior performance. A future exper-

iment that varies the duration of training (both in terms of the number of trials per block and the number of sessions), and spaces feedback trials evenly [15] in order to allow an analysis of progression of timing performance through blocks could test this hypothesis, as well as provide more specific information regarding the mechanism of both FT and FB feedback.

The variability results from the Experiments 1 and 2 FT groups provide an interesting contrast both to Experiment 1 and our previous report [28] to the extent that the scalar property was better expressed in blocks with FT trials only (Experiment 1 retest and Experiment 2 baseline sessions) or in blocks without feedback that followed FT only blocks (Experiment 2 retest session). These results suggest that FT trials may allow a correction of non-scalar variability that persists across the retention interval. Because this pattern was established using exploratory analysis it requires replication before a unique role for FT feedback in controlling the scalar property in aging participants is reliably established. Nonetheless, the potential importance of FT feedback in controlling the common finding of poor timing variability scaling in aging participants warrants some speculation regarding mechanisms.

A priori, we expect FT feedback to strengthen the representation of the duration of the target interval by increasing the size of the sample of intervals from which a participant estimates target duration [7]. As a consequence the interval timed should vary less from trial to trial, reducing production variability. It is not clear how this mechanism can serve to reverse or overcome a source of non-scalar variability, especially if the source of the non-scalar variability is an accelerating accumulator. Moreover, to explain the present effects, one must make two additional assumptions. First, FB trials can negate the effects of FT trials on non-scalar variability when given concurrently with FT trials. Second, variability patterns persist across blocks with no feedback and retention intervals. The first assumption is necessary to explain why FB in mixed feedback regimens (like all baseline sessions in Experiment 1) does not allow the scalar property to emerge. The second assumption is necessary to explain why the Experiment 2 FT group's retest session does not show the "default" violations of the scalar property. Neither assumption is theoretically grounded, but the previously suggested experiment that tracks timing performance across blocks and following evenly spaced feedback trials could provide the insight into how both FB and FT feedback work necessary to confirm or refute these assumptions.

In conclusion, the current experiments not only shed light on the conditions that give rise to duration-dependent timing errors in healthy aging participants, they provide control data that we hope will advance the study of timing deficits in disorders of aging. These experiments complete our first series of experiments aimed at understanding the performance of healthy aging participants under conditions that give rise to more seriously flawed performance in PD patients. Nonetheless, it is clear that additional experiments are necessary to determine the relative effectiveness of different types of feedback, test the validity of the proposed mechanisms, and to integrate the limited conditions we have examined so far with the broader field of studies of timing in aging.

Acknowledgements

Support for this research was provided by National Institute of Mental Health grant RO1-MH54793 awarded to CM. Additional funding was provided by National Institute on Aging grant K01-AG000991 awarded to BCR. Thanks to Katherine Dubé and Sharon Lobo for assistance with data collection, Stephen Fairhurst for implementing custom software and data analysis methods, and to Tina Li for assistance with preparing the manuscript.

References

- [1] L.G. Allan, K. Gerhardt, Temporal bisection with trial referents, *Percept. Psychophys.* 63 (2001) 524–540.
- [2] R.A. Block, D. Zakay, P.A. Hancock, Human aging and duration judgments: a meta-analytic review, *Psychol. Aging* 13 (1998) 584–596.
- [3] F.I.M. Craik, J.F. Hay, Aging and judgments of duration: effects of task complexity and method of estimation, *Percept. Psychophys.* 61 (1999) 549–560.
- [4] C. Fortin, Attentional time-sharing in interval timing, in: W.H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing*, CRC Press, Boca Raton, FL, 2003, pp. 235–260.
- [5] P. Fraisse, Perception and estimation of time, *Annu. Rev. Psychol.* 35 (1984) 1–36.
- [6] C. Gallistel, J. Gibbon, Time, rate, and conditioning, *Psychol. Rev.* 107 (2000) 289–344.
- [7] C.R. Gallistel, *The Organization of Learning*, The MIT Press, Cambridge, USA, 1990.
- [8] J. Gibbon, R.M. Church, Sources of variance in an information processing theory of timing, in: H.L. Roitblat, T.G. Bever, H.S. Terrace (Eds.), *Animal Cognition*, Erlbaum, Hillsdale, NJ, 1984, pp. 465–488.
- [9] J. Gibbon, R.M. Church, W. Meck, Scalar timing in memory, *Ann. N.Y. Acad. Sci.* 423 (1984) 52–77.
- [10] S.C. Hinton, D.L. Harrington, J.R. Binder, S. Durgerian, S.M. Rao, Neural systems supporting timing and chronometric counting: an fMRI study, *cognitive, Brain Res.* 21 (2004) 183–192.
- [11] S.C. Hinton, S.M. Rao, “One-thousand one...one-thousand two...”: chronometric counting violates the scalar property in interval timing, *Psychon. Bull. Rev.* 11 (2004) 24–30.
- [12] H. Lejeune, Switching or gating? The attentional challenge in cognitive models of psychological time, *Behav. Process.* 44 (1998) 127–145.
- [13] C. Lustig, W.H. Meck, Paying attention to time as one gets older, *Psychol. Sci.* 12 (2001) 478–484.
- [14] C. Lustig, Grandfather’s clock: attention and interval timing in older adults, in: W.H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing*, CRC Press, Boca Raton, FL, USA, 2003, pp. 261–293.
- [15] C. Lustig, W.H. Meck, Chronic treatment with haloperidol induces deficits in working memory and feedback effects of interval timing, *Brain Cogn.* 58 (2005) 9–16.
- [16] C. Malapani, B. Dubois, G. Rancurel, J. Gibbon, Cerebellar dysfunctions of temporal processing in the seconds range in humans, *Neurorep. Int. J. Rapid Commun. Res. Neurosci.* 9 (1998) 3907–3912.
- [17] C. Malapani, B.C. Rakitin, R. Levy, W.H. Meck, B. Deweer, B. Dubois, J. Gibbon, Coupled temporal memories in Parkinson’s disease: a dopamine-related dysfunction, *J. Cogn. Neurosci.* 10 (1998) 316–331.
- [18] C. Malapani, B. Deweer, J. Gibbon, Separating storage from retrieval dysfunction of temporal memory in Parkinson’s disease, *J. Cogn. Neurosci.* 14 (2) (2002) 311–322, Feb 2002.
- [19] C. Malapani, B.C. Rakitin, Interval timing in the dopamine-depleted basal ganglia: from empirical data to timing theory, in: W.H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing*, CRC Press, Boca Raton, FL, 2003, pp. 485–514.
- [20] S. Mattis, *Dementia Rating Scale (DRS)*, Psychological Assessment Resources, Odessa, FL, 1988.
- [21] T. McCormack, G.D. Brown, E.A. Maylor, R.J. Darby, D. Green, Developmental changes in time estimation: comparing childhood and old age, *Dev. Psychol.* 35 (1999) 1143–1155.
- [22] W.H. Meck, R.M. Church, A mode control model of counting and timing processes, *J. Exp. Psychol. Anim. Behav. Process.* 9 (1983) 320–334.
- [23] T.B. Penney, Modality differences in interval timing: attention, clock speed, and memory, in: W.H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing*, CRC Press, Boca Raton, FL, 2003, pp. 209–233.
- [24] S. Perbal, S. Droit-Volet, M. Isingrini, V. Pouthas, Relationships between age-related changes in time estimation and age-related changes in processing speed, attention and memory, *Aging Neuropsychol. Cogn.* 9 (2002) 201–216.
- [25] E. Poppel, A hierarchical model of temporal perception, *Trends Cogn. Sci.* 1 (1997) 56–61.
- [26] B.C. Rakitin, J. Gibbon, T.B. Penney, C. Malapani, S.C. Hinton, W.H. Meck, Scalar expectancy theory and peak-interval timing in humans, *J. Exp. Psychol. Anim. Behav. Process.* 24 (1998) 15–33.
- [27] B.C. Rakitin, The effects of spatial stimulus-response compatibility on choice time production accuracy and variability, *J. Exp. Psychol. Hum. Percept. Perform.* 31 (2005) 685–702.
- [28] B.C. Rakitin, Y. Stern, C. Malapani, The effects of aging on time reproduction in delayed free-recall, *Brain Cogn.* 58 (2005) 17–34.
- [29] B.C. Rakitin, N. Scarmeas, T. Li, C. Malapani, Y. Stern, Single-dose Levodopa Administration and Aging Independently Disrupt Time Production, *J. Cogn. Neurosci.* 18 (2006) 376–387.
- [30] T.H. Rammsayer, Ageing and temporal processing of durations within the psychological present, *Eur. J. Cogn. Psychol.* 13 (2001) 549–565.
- [31] E.T. Shea-Brown, J. Rinzel, B.C. Rakitin, C. Malapani, A firing-rate model of Parkinson’s disease deficits in interval timing, *Brain Res.* 1070 (2006) 189–201.
- [32] Y. Stern, M. Sano, J. Paulson, R. Mayeux, Modified mini-mental state examination: validity and reliability, *Neurology* 37 (1987) 179.
- [33] M. Treisman, Temporal discrimination and the indifference interval: implications for a model of the “internal clock”, *Psychol. Monogr.* 77 (1963) 1–13.
- [34] S. Vanneste, V. Pouthas, Timing in aging: the role of attention, *Exp. Aging Res.* 25 (1999) 49–67.
- [35] J. Wearden, A. Wearden, P. Rabbitt, Age and IQ effects on stimulus and response timing, *J. Exp. Psychol. Hum. Percept. Perform.* 23 (1997) 962–979.
- [36] J. Wearden, S. Bray, Scalar timing without reference memory? Episodic temporal generalization and bisection in humans, *Quart. J. Exp. Psychol. B: Comp. Physiol. Psychol.* 54B (2001) 289–309.
- [37] J.H. Wearden, B. McShane, Interval production as an analogue of the peak procedure: evidence for similarity of human and animal timing processes, *Quart. J. Exp. Psychol.* 40B (1988) 363–375.
- [38] D. Zakay, R.A. Block, Temporal cognition, *Curr. Direct. Psychol. Sci.* 6 (1997) 12–16.