

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/7861713>

The effects of aging on time reproduction in delayed free-recall

Article in *Brain and Cognition* · July 2005

DOI: 10.1016/j.bandc.2004.09.006 · Source: PubMed

CITATIONS

43

READS

89

3 authors, including:



Brian C Rakitin

Columbia University

51 PUBLICATIONS 2,587 CITATIONS

[SEE PROFILE](#)



Yaakov Stern

Columbia University Vagelos College of Physicians and Surgeons

1,023 PUBLICATIONS 83,244 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Predictive Modeling in Psychiatry and Neurology [View project](#)



Cognitive Reserve [View project](#)

The effects of aging on time reproduction in delayed free-recall[☆]

Brian C. Rakitin^{a,*}, Yaakov Stern^a, Chariklia Malapani^b

^a *Division of Cognitive Neuroscience, Taub Institute, Columbia University, United States*

^b *New York State Psychiatric Institute, United States*

Accepted 16 September 2004

Available online 10 December 2004

Abstract

The experiments presented here demonstrate that normal aging amplifies differences in time production occurring in delayed free-recall testing. Experiment 1 compared the time production ability of two healthy aged groups as well as college-aged participants. During the test session, which followed a 24-h delay and omitted all feedback and examples of the two target intervals, the two groups of aged participants' over-produced a 6 s interval. This effect is similar in form to errors shown by young participants, but twice the magnitude. Productions of a 17 s interval were generally accurate overall. However, further analysis indicated that the majority of aged participants over-produced the 17 s interval while a minority greatly under-produced the 17 s interval. Furthermore, aged participants showed violations of the scalar property of timing variability in the training session, while in the test session, only those who under-produced the 17 s interval showed this tendency. In contrast, training session performance was good for all participants. Experiments 2 and 3 investigated the ability of the participants in Experiment 1 to reproduce the length of a line from memory, under conditions analogous to those of the time production experiments. These experiments provided tests of the specificity of the errors observed in Experiment 1. Performance in the older participants was accurate, if more variable, compared to the young participants, in contrast to the time production results, indicating that production inaccuracy in free-recall is specific to interval timing in the current context.

© 2004 Elsevier Inc. All rights reserved.

1. Introduction

Interval timing requires several cognitive processes such as an internal time basis, as well as attention, memory, and decision processes (Gibbon, 1977; Gibbon, Church, & Meck, 1984; Zakay & Block, 1997). Some of these processes are also mental functions that decline

with age, including attention, working memory, and reaction time (see Raz, 2000 for review). Likewise, the ability to time short intervals in the seconds to minutes range is served by the same brain structures and neurotransmitter systems that underlie higher cognitive functions, such as attention and working memory, and are most sensitive to age (Meck, 1996; Meck & Benson, 2002; Raz, 2000; Rao, Mayer, & Harrington, 2001). These factors may make timing tasks very sensitive measures of age-related changes in cognition.

Information-processing models, such as scalar expectancy theory (SET; Gibbon, Church, & Meck, 1984; see also Treisman, 1963), attempt to describe the cognitive processes involved in interval timing. In SET, as well as related models like the Attention Gate Model (AGM; Zakay & Block, 1997), attention to time allows pulses from a pacemaker to pass into an accumulator, and the accumulation of such pulses marks the passage of time.

[☆] Funding for this research was provided through NIMH Grant No. 2-R01-MH54793, awarded to Chariklia Malapani and John Gibbon. The authors thank Steven Fairhurst for providing data collection and analysis software, and Bonnie Taylor, Jula Veerapong, Courtney Wustoff, Katherine Dube, and Sharon Lobo for data collection. Thanks to Sylvie Lavesque and Tina Li for assistance in manuscript preparation.

* Corresponding author. Present address: Columbia Physicians and Surgeons, 630 West 168th Street, Box 16, New York City, NY 10032, United States.

E-mail address: br130@columbia.edu (B.C. Rakitin).

Judgments about the current duration versus previously relevant durations are made by comparing current accumulator values to samples from a reference memory distribution of accumulator values that were previously associated with the relevant event (e.g., in previous trials, how many pulses were accumulated before reinforcement occurred?). The distinguishing feature of SET is that in its computational form, variability in clock and memory processes is used to explain the common finding that standard deviations in timing performance are a constant proportion of (i.e., “scaled to”) the timed duration (Gibbon, Malapani, Dale, & Gallistel, 1997). This is referred to as the scalar property of timing variability, and is considered a domain specific example of Weber’s law, which describes the psychophysical relationship between stimulus magnitude and thresholds in a variety of stimulus domains (Gibbon, 1977). Experimental manipulations that lead to violations of the scalar property typically are thought to affect non-temporal aspects of performance (e.g., decision processes, motor output; Gibbon et al., 1997; Ivry & Hazeltine, 1995). Thus, measures of variability as well as accuracy are essential for understanding effects on interval timing.

Several previous studies of older adults’ interval timing accuracy have focused on age effects on attention’s role in SET and related models (Block, Zakay, & Hancock, 1998; Craik & Hay, 1999; Lustig & Meck, 2001; Vanneste & Pouthas, 1999). According to a recent meta-analysis (Block et al., 1998), verbal estimations grew shorter with age, verbal productions grew longer with age, and temporal production (most akin to the task used here) was unaffected by age. These effects were independent of target interval duration. AGM and SET explain attention effects in general by means of the switch or gate between the pacemaker and accumulator. When attention is directed away from time some proportion of pacemaker pulses do not increment the accumulator. As a result, a mismatch occurs between the estimate of current subjective time and the value of subjective remembered time because the direction of the effect depends on whether distraction occurred during acquisition or testing.

Although less heavily researched, there is also a potential link between the age-associated changes in interval timing and memory. A recent study (Perbal, Droit-Volet, Isingrini, & Pouthas, 2002) found time reproductions under distraction shortened with age, independent of interval duration, and were associated with poorer working memory performance. Other studies have found progressively noisy encoding of temporal values with increasing age (McCormack, Brown, Maylor, Darby, & Green, 1999; Wearden, Wearden, & Rabbitt, 1997), also largely independent of target interval duration.

The common findings that older participants’ timing errors, due to either memory or attention effects, are

independent of target interval duration are of particular interest because existing timing theories such as SET and AGM predict that if two durations were tested within the same session, older adults’ errors for both durations would be in the same direction. In the case of attention effects predicted by AGM (Zakay & Block, 1997) and SET (Meck & Church, 1983), the magnitude of the effects of attention stem from the degree to which a particular manipulation causes the switch between the pacemaker and accumulator to open. The direction of the attention effects is a function of when in the task distraction occurs or changes, and whether the timing task is prospective or retrospective (Block et al., 1998; Zakay & Block, 1997). As a net result of these factors, any given attention manipulation included during a given phase of a timing experiment should produce effects of uniform direction and magnitude regardless of the target interval duration. Similarly, studies conducted within the SET framework have identified a number of parameters that determine the function of the various temporal memory systems (Gibbon et al., 1984), all of which operate independently of the target interval duration. For example, “K*” multiplies all temporal values prior to storage in memory, resulting in a mismatch between currently elapsing and remembered time values (Meck, 1983, 2002). While the direction of both attention and memory effects predicted by SET is independent of the target interval duration, the magnitude of memory effects can grow with the target interval duration.

While studies of normal aging have produced evidence of duration-independent changes in timing accuracy, studies of Parkinson’s disease (PD), a neurodegenerative disease of aging (Fahn, 1995; Rakitin & Stern, 2002), have produced evidence of duration-dependent temporal memory effects. When taken off dopamine-replacement therapy, PD patients’ productions of the shorter of two intervals (e.g., 6 or 8 s) were longer than the standard intervals. Also, productions of the longer of two intervals (e.g., 17 or 21 s) were shorter than the standard intervals (Malapani, Deweer, & Gibbon, 2002; Malapani, Rakitin, Fairhurst, & Gibbon, 2002; Malapani et al., 1998). This “migration effect” strongly contrasted with performance by PD patients on medication, which was generally superior to that of aged controls (Malapani et al., 1998, 2002), and was determined to stem from a dopamine-dependent deficit in encoding temporal memory (Malapani et al., 2002). Thus, while current theories emphasize the likelihood that the direction of age-related changes in timing accuracy would be independent of target interval duration, there are known duration-dependent timing changes in pathological aging that must be considered as an alternative hypothesis.

In contrast to the many findings regarding timing inaccuracy in aging, the findings regarding timing variability are more mixed. Studies using diverse methods

such as repetitive tapping (Krampe, Engbert, & Kliegl, 2002; Vanneste, Pouthas, & Wearden, 2001), time production (Block et al., 1998; Wearden et al., 1997), and temporal bisection (McCormack et al., 1999; Wearden et al., 1997), found little if any increase in timing variability associated with aging. In contrast, temporal generalization gradients (McCormack et al., 1999; Wearden et al., 1997), temporal decision thresholds (Wearden et al., 1997), verbal estimation, and verbal production (Block et al., 1998) do evidence age-related increases in temporal variability. In addition, age-related changes in bisection sensitivity do occur under some circumstances (Lustig & Meck, 2001). Where available, data has supported the preservation of the scalar property of timing variability in older participants (Block et al., 1998; McCormack et al., 1999; Wearden et al., 1997). Specific components that account for these between-task differences in age-related increases in timing variability include IQ, availability of feedback, and examples of the temporal standards (Wearden et al., 1997), the availability of attention resources (Lustig & Meck, 2001; Vanneste & Pouthas, 1999), working memory integrity (Vanneste & Pouthas, 1999), stimulus modality (Lustig & Meck, 2001), and circadian periodicity (Lustig & Meck, 2001).

Another issue in the analysis of age-related changes to timing abilities concerns the degree to which observed effects can be specifically ascribed to the internal clock, as opposed to the many more general cognitive operations such as memory and attention that decline with aging (Raz, 2000). Some studies have explicitly attempted to generalize timing effects beyond the internal clock by associating observed effects with age-related changes in other working memory tasks (Craig & Hay, 1999; Perbal et al., 2002), or global IQ measures (Wearden et al., 1997). Another tactic taken in studies of brain-damaged patients (Ivry & Keele, 1989) or in studies of brain imaging (Rao et al., 2001) is to compare closely matched studies of time perception and perception in other domains such as auditory pitch. Of particular concern to us is the lack of methodology for comparing production in the domain of time and in other domains—a problem we seek to solve in two of the current studies.

Our current studies of the effects of aging on timing use the methods previously used to assess the time production abilities of PD patients. This choice was motivated by a spate of findings suggesting that not only is dopamine depleted as a function of aging in both humans (Agid, Ruberg, Dubois, & Pilon, 1987; Agid et al., 1989) and non-human primates (Goldman-Rakic & Brown, 1981; Wenk, Pierce, Struble, Price, & Cork, 1989), but age-related dopamine depletion may be specifically linked to the age-related cognitive deficits related to frontal-lobe function (Arnsten, Cai, Steere, & Goldman-Rakic, 1995). These methods entail use of the peak-interval procedure designed for use with human

participants (Rakitin et al., 1998; see also Catania, 1970; Roberts, 1981), known to produce especially stable estimates of timing accuracy and variability, as well as free-recall testing following a 24 h delay (Malapani et al., 2002). Combined with the use of two target time intervals, this procedure will allow us to detect mild “migration-like” (i.e., duration-dependent in direction) temporal memory effects that could potentially arise from mild age-related dopamine depletion, as well as more common age-related memory or attention effects that create temporal production errors in the same direction for both target time intervals.

Finally, we offer for the first time two different experiments that examine the ability of aged participants to produce lines of varying length. These tasks were carefully constructed to control for several domain-independent aspects of memory, attention, motor activity and executive function.

2. Experiment 1: Time production

The experiment was originally conceived and run as a comparison between 16 young participants and 16 healthy aged individuals, referred to here as Aged Group 1. Upon completing and fully analyzing the results of that study, we recognized the need to replicate the unusual findings. We therefore collected data from a second set of 16 aged individuals, referred to as Aged Group 2. The analyses were complicated by the differences in the demographic and neuropsychological features of the two aged groups stemming from the different methods used to recruit them. In addition, the second aged group produced more variable results that reduced the reliability of our statistical tests. To overcome this problem we used the results of the initial comparisons between Aged Group 1 and the young participants to generate forward hypotheses regarding outcomes of the three group analyses.

2.1. Method

2.1.1. Apparatus

Data from the young participants were collected on a PC DOS system equipped with a 15 in. CRT color monitor. Data from Aged Group 1 were collected using a Macintosh CI computer with a 15 in. CRT color monitor. The young and Aged Group 1 participants were tested in a 9 ft. × 9 ft. windowless room dedicated to running experiments in the New York State Psychiatric Institute. Data from Aged Group 2 were collected in senior centers located throughout Manhattan, using a Macintosh G3 Powerbook equipped with a 13 in. LCD monitor.

Stimulus sizes were matched across the systems, taking into consideration seating distance from the visual

displays. In all three systems, responses were input to the computers via the keyboard. Custom DOS and Macintosh data collection programs with identical algorithms were used to present the stimuli in the appropriate time intervals and record responses. Accuracy was limited to approximately ± 15 ms as a result of both keyboard imprecision and software keyboard polling rate.

2.1.2. Participants

Sixteen young participants (eight males and eight females) were recruited from Columbia University. These participants ranged in age from 18 to 28 years of age, and had an average of 20 years of education. None had a history of neurological or psychiatric disorders, or were currently taking psychoactive medication.

Aged Group 1 consisted of 16 aged individuals (four males and 12 females) recruited from the New York State Psychiatric Institute's Comparison Groups Project (Schechter, Singer, Kuperman, & Endicott, 1998) and from ads placed in local newspapers. Aged Group 2 consisted of 16 aged individuals (nine males and seven females) recruited from senior day centers located in Manhattan, New York. All participants were screened for past and current medical, neurological, and psychiatric disorders, and were not currently being treated with psychoactive drugs. Aged Group 2 was also administered the Mattis Dementia Rating Scale (DRS) prior to all other testing. Those who obtained scores lower than 125 were paid for their time, but not tested further. The obtained mean DRS score for those included in the analyses was 137.1 (± 8.1). The normative DRS score for individuals aged 62–79 is 134 (± 6.3).

Neuropsychological tests were administered to characterize participants' general and specific cognitive functioning. For Aged Group 1, these tests included the full Wechsler Adult Intelligence Scale, Third Edition (WAIS-III), or Revised Edition (WAIS-R), the Trail-

Making Test A and Trail-Making Test B, the Stroop test, and the Wisconsin Card Sort Test (WCST). For Aged Group 2 the American New Adult Reading Test (AMNART) was substituted for the full-scale WAIS.

Demographic and neuropsychological test scores for the aged participants are listed in Table 1. Where more than one normative test-score is given, the score for the younger normative group applied to the aged participants in this experiment. The normative scores are associated with "all" ages when the test scores are individually aged scaled.

Demographic and neuropsychological test scores for the aged participants are listed in Table 1. As can be seen, the method of recruiting Aged Group 2 yielded an older, less educated population with overall lower cognitive performance. The differences in neuropsychological profile between Aged Group 1 and Aged Group 2 precluded an analysis approach that included pooling all aged participants. Moreover, these differences suggested a need for direct comparison of the two aged groups. The choice of planned contrasts, discussed below, reflects these considerations.

Informed consent was obtained from all participants in accordance with the procedures of both Columbia University and the New York State Psychiatric Institute.

2.1.3. Procedure

The time production task is a variant of the human peak-interval task (Rakitin et al., 1998). This task was in turn adapted from the operant peak-interval procedure used in studies of animal behavior (Catania, 1970; Roberts, 1981). The task was carried out over two consecutive days. The first day of the procedure is the "training" session. The second day is the "testing" session. The sessions differ in terms of the types of task trials, a design referred to as the encode–decode design (Malapani et al., 2002). There were two target time intervals, 6 and 17 s.

Table 1
Demographic and neuropsychological description of aged participants

	Mean (SD)			
	Experiment 1 Aged Group 1	Experiment 1 Aged Group 2	Experiment 3 Aged Group	Test norms by Age
Age	63.5 (9.2)	70.2 (6.2)	75.1 (7.5)	—
Years of education	15.7 (2.1)	13.2 (3.5)	14.3 (3.0)	—
Full-scale IQ (WAIS-III)	113.5 (17)	—	—	All, 100 (15)
AMNART estimated IQ	—	105.05 (11.3)	112.2 (10)	All, 100 (15)
WAIS-III colored matrices	12.3 (3.0)	8.57 (2.8)	11 (2.6)	All, 10 (3)
Trails A: time to complete	40.1 (20.1)	57.8 (33.1)	58.3 (26.3)	60–69, 35.8 (11.9) 70–74, 41.3 (15)
Trails B: time to complete	82.8 (28.6)	145.5 (81.6)	135.7 (85.2)	60–69, 81.2 (38.5) 70–74, 111.4 (72.2)
Trails B–A	50.0 (22.9)	87.6 (65.3)	89.4 (75)	60–69, 45.6 (26.6) 70–74, 111.4 (72.2)
Stroop word reading	102.7 (9.9)	97.1 (16.0)	96.21 (13.4)	All, 108 (20)
Stroop color naming	73.9 (9.9)	67.1 (16.4)	67.92 (12.7)	All, 80 (15)
Stroop word color naming	45.5 (9.8)	38.2 (9.9)	42 (6.8)	All, 45 (10)
WCST categories attained	4.9 (1.8)	4.1 (2.3)	4.8 (1.8)	60–64, 4.46 (1.7) 70–74, 3.97 (1.6)

The intervals were the same for both sessions, and the tasks for each interval were carried out in a blocked fashion.

2.1.3.1. Training session. Time production training occurred during the first experimental session. Production of 6 and 17 s intervals occurred in separate sets of trials. Half of the participants were trained on the 6 s interval first and half were trained on the 17 s interval first. Training for each interval consisted of a total of 80 trials divided into two sets of 10 and one set of 60 trials. The first two sets of trials served to acquaint participants with the interval to be tested in the third set of trials, the training block. Data presented below are from responses collected in the third set of trials.

The first set of trials consisted of 10 “fixed-time trials.” On each trial a blue square (2.1 in. × 2.1 in.) was presented on the monitor. After the target interval had elapsed, the blue square turned magenta. Participants were instructed to try to remember the time interval defined by the appearance of the blue square on the screen and the subsequent color change. Participants were informed that they should refrain from counting or tapping so as to avoid extraneous timekeeping. To further discourage counting, random digits were intermittently superimposed over the rectangle, and participants were instructed to read the digits aloud (Rakitin et al., 1998). Inter-digit times were determined randomly using two uniform distributions. The first had a mean of 0.45 s and a range of 0.5 s. The second had a mean of 1.55 s and a range of 0.5 s.

The second set of trials consisted of 10 “peak trials” in which participants were asked to reproduce the time interval that they had just observed in the preceding fixed-time trials. In these peak trials, the blue square did not turn magenta, but instead remained blue. Participants were instructed to respond by pressing the spacebar just before the expected end of the target interval and releasing it just after the target time had elapsed. That is, participants were to “bracket” the remembered duration of the target interval by pressing, holding and the releasing the space bar. The time when the key is pressed is referred to as the “start” time, and the time when the key is released is referred to as the “stop” time. Digit distracters were employed, as described above. Participants ended the trial by pressing the “enter” key. Alternately, the trial ended when a period equal to three times the target interval had elapsed. During the inter-trial interval (ITI), participants were shown a histogram that provided feedback about their response on the previous trial in relative time. Additional messages informed the participants if they had responded too short or too long with respect to the target time (see also Rakitin et al., 1998).

For the third set of trials of the training session, young participants were randomly assigned to one of

two groups. Group A received 20 fixed-time trials, 30 peak trials with feedback during the ITI, and 10 peak trials without ITI feedback. Group B received 30 fixed-time trials and 30 peak trials without feedback. This design was incorporated to confirm the previously reported observation (Rakitin et al., 1998) that the form of the feedback does not significantly alter performance of young participants. Statistical analysis using the variables described below indicated no main effects or interactions involving feedback group. As a result, young participants were treated as a single group in all subsequent analyses. Aged participants received the same trial mixture as young Group A, just described. This was the same mixture used in a previous study of PD patients (Malapani et al., 1998). Digit distracters were used in all trials for both groups.

2.1.3.2. Testing. In the second session, participants received one set of 60 peak trials without feedback for each of the two target intervals. Distracter digits were used in all trials. The order of testing of the two intervals was counter-balanced so that half of the participants trained first on the 6 s interval were tested first on the 6 s interval, and so on.

2.1.3.3. Analysis. The middle of each trial’s responses, referred to as the “peak time,” was determined by taking the average of the times (i.e., latencies from signal onset) at which the extended responses started and stopped. Each participant contributed four distributions of peak times, representing productions of the long and short intervals during training and testing.

To illustrate these results, individuals’ distributions of peak times were converted to absolute response–frequency histograms with 0.5 s bins. Relative response–frequency distributions were computed from the absolute response–frequency distributions by dividing each bin’s count of responses by the total number of responses. Relative response–frequency functions representing the between-participants averages were created by centering the peak of each participant’s response–frequency distribution of median peak times at the group mean. This alignment eliminates contributions to the average distributions’ total variability by variability in participants’ mean peak times. The average function is then computed by taking the arithmetic mean of each participant’s corresponding time bins.

Initial omnibus statistical analyses compared performance between young and Aged Group 1 participants (group), between 6 and 17 s intervals (duration), and between training and testing sessions (session). Subsequent analysis included Aged Group 2 as a third between-subjects level of the group factor. Only data from this second analysis are presented, but the results should be assumed to be same, except where explicitly noted in the text. Performance was indexed by three

dependent variables; the median, the semi-interquartile range (SIQR: defined as half the difference between the 75th and 25th percentiles), and coefficient of variation (CV: defined as the SIQR divided by the median) of individuals' distributions of peak times. All mean dependent variable values are presented in Table 2. Separate repeated-measures factorial ANOVA, with two within-participants effects (duration and session) and one between-participants effect (group), were conducted for each of the three dependent variables. All probabilities reported with ANOVA effects are computed for 1 and 45 degrees of freedom. The results of these three analyses are reported in full in Table 3.

Planned contrasts were used to elaborate on the differences among the participant groups, and to maximize statistical power based on hypotheses generated from the initial omnibus analysis (not reported here) of young and Aged Group 1 participants. The first contrast compared young and old participants, so the means of

the young, the Aged Group 1, and Aged Group 2, were weighted 1, -0.5 , and -0.5 , respectively. This asks whether the interaction of the effects of session and duration on mean production latency, for example, found in young participants differs from the average effect of the interaction evident in the two aged groups. The second contrast compared the two old groups, with equally weighted means. These planned contrasts are defined for all ANOVA terms incorporating the effect of groups. In order to reflect the planned nature of the test and the directionality of the hypothesis, the F statistic and associated probability value were converted to a t statistic with an associated one-tailed probability value. The expected result implicit in the use of the one-tailed t test for the critical interaction of session, duration, and group, is that the shift in production of the 6 s standard between the train and test sessions will be larger than the shift in production of the 17 s interval between train and test session, and also larger than that seen in the young

Table 2
Experiment 1—Between-participants means and standard errors of median, SIQR, and CV of peak times

Session	Group	Target (s)	Median	SIQR	CV
Training	Young	6	6.13 (0.12)	0.59 (0.05)	0.098 (0.008)
		17	17.45 (0.29)	1.62 (0.17)	0.092 (0.009)
	Aged Group 1	6	6.17 (0.18)	0.72 (0.07)	0.117 (0.011)
		17	16.75 (0.27)	1.56 (0.10)	0.094 (0.006)
	Aged Group 2	6	6.28 (0.15)	0.72 (0.04)	0.116 (0.009)
		17	16.45 (0.39)	1.59 (0.08)	0.092 (0.006)
Testing	Young	6	7.24 (0.42)	0.70 (0.08)	0.093 (0.008)
		17	17.90 (0.75)	1.86 (0.21)	0.102 (0.010)
	Aged Group 1	6	8.24 (0.53)	0.85 (0.08)	0.106 (0.011)
		17	15.74 (0.77)	1.54 (0.12)	0.102 (0.009)
	Aged Group 2	6	8.06 (0.32)	0.96 (0.07)	0.125 (0.012)
		17	16.54 (0.78)	1.74 (0.21)	0.105 (0.011)

Table 3
Experiment 1—ANOVA table by effects (rows) and dependent variables (columns)

Effect	df	Median		SIQR		CV	
		MSE	F	MSE	F	MSE	F
<i>Source—Between subjects</i>							
Group	2	3.62	0.70	0.127	0.27	0.004	1.11
Error (Group)	45	5.14		0.464		0.003	
<i>Source—Within Subjects</i>							
Session	1	27.00	6.67*	0.923	6.33*	0.0004	0.36
Session × Group	2	0.66	0.16	0.095	0.65	0.0003	0.29
Error (session)	45	2.90		0.146		0.001	
Duration	1	4599.64	1588.26*	38.409	202.76*	0.005	6.49*
Duration × Group	2	17.83	17.83*	0.500	2.64	0.002	2.25
Error (duration)	45	2.90		0.189		0.0008	
Session × duration	1	39.20	16.54*	0.017	0.14	0.001	1.90
Session × duration × group	2	5.88	2.48**	0.094	0.75	0.0005	0.66
Error (session × duration)	45	2.37		0.125		0.0007	

Note. ANOVA results reflect independent univariate analysis of peak times. MSE is the mean squared error for the indicated effect. F is the ratio of the MSE for the indicated effect and the MSE for the error listed next below it.

* A significant effect at the $p < .05$ level.

** A marginal effect at the $p < .1$ level.

participants. This was the result of the initial ANOVA, which excluded Aged Group 2.

The median peak time is the primary measure of response accuracy, so the null hypothesis is that median peak times should depend on the duration of the target interval, but not on age group or session. This would indicate that all participants are reproducing the target interval relatively accurately, and that neither age nor feedback affect timing. SIQR of peak times should also increase with interval duration, indicating that timing becomes more error-prone as duration increases. On the other hand, CV of peak times should be unaffected by interval duration if Weber's law applies. That is, absolute variability measured by the SIQR should increase with the magnitude (i.e., duration) of the stimulus, but this variability should be a fixed proportion of the duration produced.

The analysis of subgroups of aged individuals used paired two-tailed t tests to compare each subgroup's changes in median peak times from the train session to the test session. Two-tailed t tests assuming unequal variances compared the subgroups' median peak times within session and target interval duration. The analysis of the expression of the scalar property within each subgroup during the testing session used paired two-tailed t tests to compare the CV obtained from productions of the two target intervals.

2.2. Results

The results of the timing task are presented in Fig. 1, where each panel illustrates the distributions of individual trials' peak times for productions of 6 and 17 s intervals in the training and test sessions, averaged across participants, for either the young participants (top left), Aged Group 1 (bottom left), or Aged Group 2 (bottom right). The first notable result is that all three groups responded quite accurately in training so that their distributions of responses are centered and peaked at 6 and 17 s (indicated in the figure by vertical black lines). In contrast, during the test session, productions were longer than the criterion time for the 6 s interval, and roughly equal to the criterion time for the 17 s interval. In addition, the errors made by the aged participants in both groups were almost twice as large as those made by the young participants. These effects are also evident in the groups mean peak times reported in Table 2. It is therefore apparent that time production following a 24 h delay and in the absence of feedback is associated with age-related timing errors that are enhanced by aging.

Individuals' data are presented in Fig. 2. The left panel shows the combined data from the two groups of aged participants, while the right panel shows the data from the young participants. In both panels, the

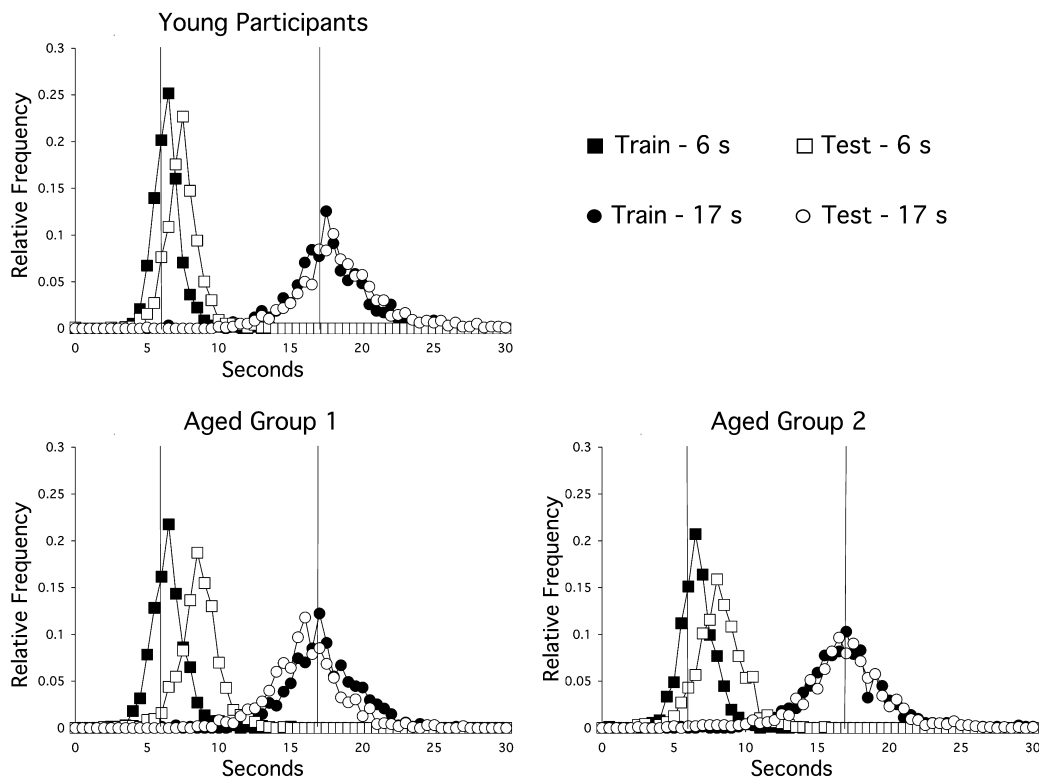


Fig. 1. Average relative frequency histograms showing distributions of individual trials' "peak times" for production of the 6 and 17 s target intervals during training and test sessions for young participants (upper left panel), Aged Group 1 (bottom left panel), and Aged Group 2 (bottom right panel) from Experiment 1. Each function is the within-group average of histograms derived for each individual participant.

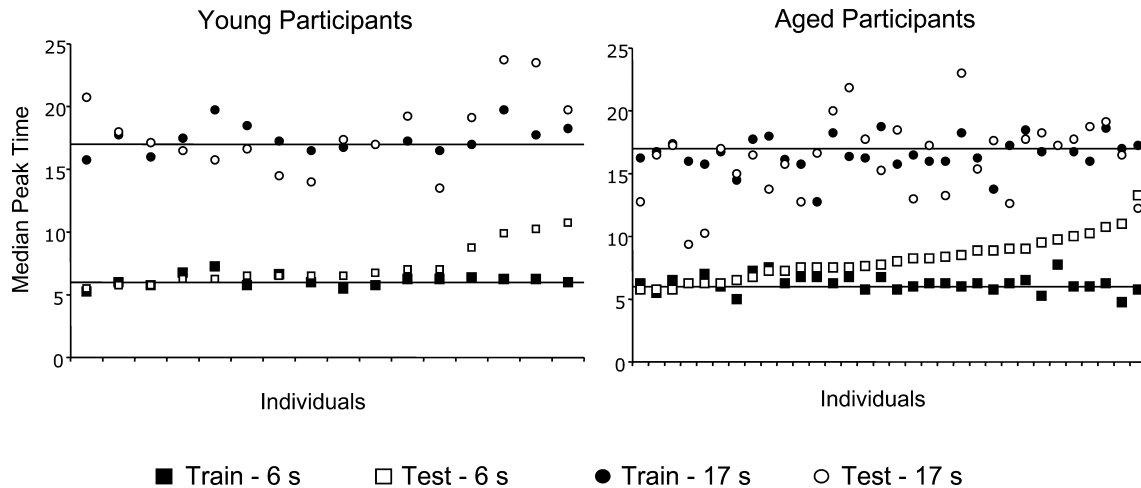


Fig. 2. Young (left panel) and all aged (right panel) participant's median peak times for production of the 6 and 17 s targets in training and test sessions of Experiment 1. The participants are ordered from left to right by increasing peak times for production of the 6 s target during the test session.

participants are ordered left to right by the magnitude of the median peak time of production of the 6 s interval in the test session. As can be seen in both aged and young participants there is little correspondence between either the direction or magnitude of test session production errors for the two target intervals. In addition, those individuals who produce the longest 6 s intervals in the test session are not those who produce either the shortest or the longest 17 s intervals in the test session. Interestingly, it appears that a few of the young participants over-produce the 6 s interval in the test session to the same degree as the worst performing aged participants. The difference between group average magnitudes of over-productions of the 6 s interval in the test session therefore seems to reflect the number of participants in either group making the largest errors, rather than the average size of errors in all participants.

2.2.1. Timing accuracy: Analysis of median peak times

The pattern of errors apparent in Fig. 1 manifests as a marginally significant omnibus interaction among session, group, and duration effects on the median middle. However, the interaction among session, duration, and the contrast between older and younger participants' means was significant ($t(45)=1.82, p<.05$), while the interaction among session, duration, and the contrast between the means of the two groups of older participants was not ($t(45)=1.29, p>.1$). It is important to note that the analysis that included only the young participants and the Aged Group 1 participants produced a significant omnibus interaction of session, group, and duration ($F(1,30)=5.3$, mean square error=2.199, $p<.05$). It was this analysis that motivated the choice of planned contrasts, as well as the expected direction of effects, in the analysis that includes the replication Aged Group 2, reported in full in Table 3.

Post hoc analysis of this interaction indicates that the Aged Group 1 participants' productions of the 6 s interval were 2.07 s longer in the testing session than in the training session ($t(30)=3.73, p<.01$, one-tailed). Similarly, the approximately 1.96 s lengthening of the Aged Group 2 participants' productions of the 6 s interval between the test and train sessions was also reliable ($t(30)=4.01, p<.01$, one-tailed). The Aged Group 1 participants' productions of the 17 s interval did not differ between the training and testing sessions ($t(30)=1.35, p<.1$, one-tailed), despite the apparent shortening of the 17 s productions in the test session evident in Fig. 1 (left panel). The Aged Group 2 participants' productions of the 17 s interval did not differ between the training and testing sessions ($t(30)=.1, p<.1$, one-tailed). Young participants' productions of the short interval in the testing session were also longer than in the training session ($t(30)=2.64, p<.01$), although the effect is a full second smaller than the effect in the aged group. Young participants' productions of the 17 s interval did not differ between sessions ($t(30)=.2$).

The effect of duration was significant, reflecting basic discrimination between the two intervals. The main effect of session, and the interactions between session and duration, and between duration and group are significant, and occur as a result of the triple interaction described above.

2.2.2. Timing variability: Analyses of SIQR and CV of peak times

Absolute timing variability increased with the duration of the target intervals. This can be seen in Fig. 1, where functions representing reproductions of the longer interval are visibly wider than those of the shorter interval. Supporting this observation, the SIQR of peak times, the variable indicating absolute timing variability,

was significantly affected by the duration of the target interval. There was also a trend towards an interaction of duration and group indicating somewhat more variable responding by the young group timing the 17 s signal. This was also reflected in the significant interaction between duration and the contrast between old and young participants ($t(45)=2.27$, $p<.05$), and the absence of a significant interaction between duration and the contrast between the two aged groups ($t(45)=.4$). These effects may reflect the fact that the productions of the 17 s interval are a little shorter for the old participants, especially Aged Group 1, than for the young participants. This is discussed further in the analysis of individual differences in aged participants' productions of the 17 s interval during the test session. Absolute timing variability also increased between the train and test sessions.

Timing variability is “scalar” and adheres to Weber’s law to the extent that the standard deviation of reproduced intervals is proportional to the target interval’s duration. That is, the CV should be equal for productions of different intervals. If inaccurate timing by older participants in the testing session was accompanied by a failure of variability to scale with the duration of the interval, we would anticipate a significant interaction effect of the group, session, and duration on the CV data. However, statistical analysis of the CVs indicated no significant effect of the interaction of session, duration, and group. In addition, neither planned contrast comprising the interaction of session, duration, and group was significant ($t(45)=1.21$ and $t(45)=.22$ for the comparisons between aged and young groups, and between aged groups, respectively). Moreover, there are no significant main effects of session, nor do any of the two-way interactions involving session approach significance. Thus, the conditions that give rise to a significant effect on the duration of production, measured by the median peak times, are not associated with a specific change in relative timing variability. In other words, while the free-recall circumstances of the test session cause changes in timing accuracy, they do not cause a change in the expression of the scalar property of timing variability.

However, the scalar property of timing variability was violated on average as indicated by a significant effect of duration on the CV of peak times. The CV of productions of short interval, averaged over groups and sessions, was 0.109, while the CV of productions of the long interval was 0.098. This effect was not equal for the young and old participants. While the omnibus test of the interaction of group and duration was not significant, the interaction between duration and the contrast between the young participants and old participants was ($t(45)=2.05$, $p<.05$). The means of CVs contributing to this effect are as follows: 0.116 for the aged participants timing the short interval, 0.010 for the aged participants timing the long interval, 0.096 for the young participants

timing the short interval, and 0.0987 for the young participants timing the long interval. The interaction of duration and the contrast between the two aged groups were not significant ($t(45)=.53$). As can be seen, the scalar property is excellently expressed by the young participants but not by the aged participants.

2.2.3. Individual differences in timing accuracy among aged participants

Close inspection of Fig. 1 and Table 2 indicates an apparent shortening of productions of the 17 s target in Aged Group 1. However, no such effect was observed in Aged Group 2, and there was no reliable statistical effect corresponding to this observation. We therefore decided to pool the data from both aged groups to examine individual differences in production. Fig. 3 shows a histogram of the means of median peak times for both intervals in both sessions from all aged participants. The figure indicates that the distribution of the 6 s means, while jagged, is uni-modal. The distribution of 17 s responses is different in that the older participants’ distribution is clearly bimodal. Note that aged participants from both groups fell into both modes (see below). Moreover, the right-most mode of the aged participants’ distribution is centered over the criterion time, while the peak of left mode, indicated by the black arrow, falls short of the criterion. We take this as an indication that there may be two groups of aged participants.

To test this hypothesis, we divided the aged participants into two groups based on test session productions of the 17 s interval. The “minority” were those with median peak times of 13.75 s ($n=9$, six from Aged Group 1, and three from Aged Group 2) or less, and the “majority” were those with peak times of 15.0 s or greater. The next largest median peak time in the majority group was 15.25, emphasizing that the chosen bifurcation point divides the left tail of the distribution of 17 s test session peak times from the more or less smooth distribution of majority group peak times.

Comparing the train session peak times to the test session peak times we found that the majority groups productions of both the 6 and 17 s target intervals were longer than the train session productions (train mean = 6.14, test mean = 8.21, $t(22)=5.72$, $p<.0001$; train mean = 16.63, test mean = 17.68, $t(22)=-2.37$, $p<.05$; for 6 and 17 s targets, respectively). The latter effect is of course a direct consequence of having removed the left tail of the distribution of aged participants 17 s test session production peak times. Nonetheless, it was conceivable that the majority of aged participants could have been found to be accurately timing the 17 s interval during the test session, as per the omnibus statistics and post hoc tests reported earlier. This contrasts with the minority group’s productions of the 17 s interval that were significantly shorter than their train session productions (train mean = 16.53, test mean = 12.22, $t(8)=10.21$,

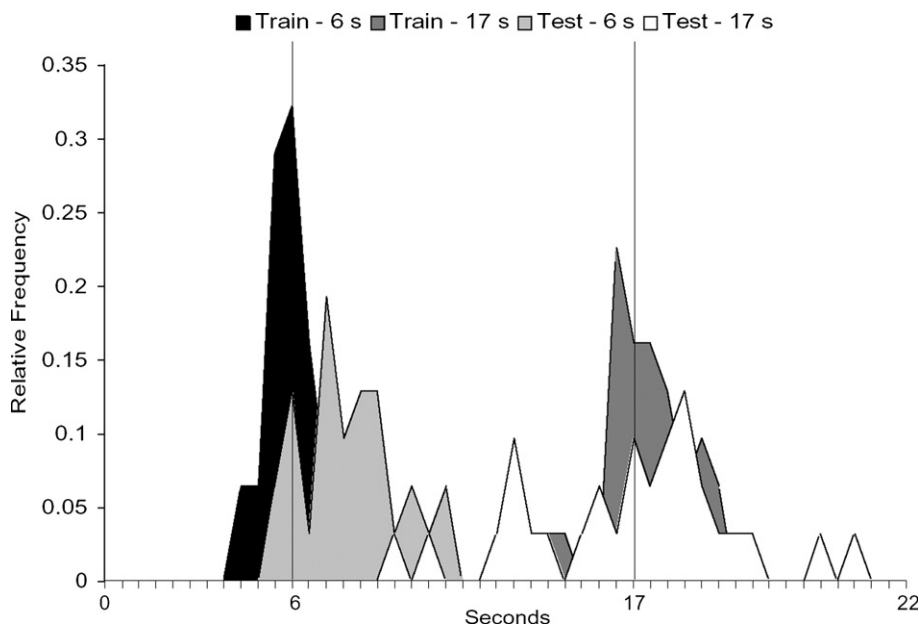


Fig. 3. Relative frequency histograms showing distributions of aged participants' median peak times for production of the 6 and 17 s intervals in the train and test sessions of Experiment 1. These data illustrate the between-subject variability in peak times contributing to the group means presented in Table 2. Unlike Fig. 1 within-subject variability is not represented. The black arrow indicates a second, shorter mode in the distribution of old participants' median productions of the long (17 s) interval during the testing session.

$p < .0001$). The minority group's productions of the 6 s target were not significantly longer in the test session compared to the train session (train mean = 6.44, test mean = 7.99, $t(8) = -1.79$, *n.s.*). Given the mean values it is clear that this is due to a simple lack of statistical power. This is supported by the lack of difference between the subgroups, reported next. From these tests we can say that there are in fact two subgroups of aged participants, one that over-produces both intervals, and a second that over-produces the 6 s interval, and under-produces the 17 s interval.

Comparing between the subgroups we found that during the train session, the subgroups did not differ in productions of either the 6 s target ($t(18) = 1.32$, *n.s.*), or the 17 s target ($t(27) = -.25$, *n.s.*). These tests indicate that the differences in the subgroups' test session performance does not reflect differences in train session performance. The subgroups do not differ in their productions of the 6 s interval ($t(11) = -.27$, *n.s.*). This test supports the assertion made above that the minority subgroup is in fact over-producing the 6 s-target despite the lack of a significant train-test difference. Finally, the test of the difference between the subgroups' test session productions of the 17 s interval is trivially significant ($t(20) = -8.65$, $p < .0001$), given that in this case the subgroups were defined as having non-overlapping distributions of peak times.

Given that these two subgroups were determined to have different time production accuracy during the test session, it is important to know whether the subgroups differ in expression of the scalar property of timing

variability. In fact, the scalar property is observed in both the majority subgroup (mean 6 s CV = .11, mean 17 s CV = .10, $t(22) = .84$, *n.s.*) and the minority subgroup (mean 6 s CV = .13, mean 17 s CV = .11, $t(8) = .84$, *n.s.*). Recalling that the omnibus analysis reported earlier indicated a lack of session effects in the age-related violations of the scalar property evident in the data, it is apparent that the minority subgroup does indeed contribute to the finding that the scalar property is violated in the test session, given that scaling is evident in the variability of the majority subgroup. Nonetheless, this finding must be considered inconclusive due to the null effect in the test of scaling in the minority group, despite the lack of power in the test stemming from the small size of the minority group and the high between-subject variability in the CV data.

To follow-up on this hypothesis, we compared the demographics and neuropsychological test scores available for both aged groups (i.e., those listed in Table 1) to see if the participants falling into the two subgroups differed in other respects. The two subgroups did not differ in age, education or any of the neuropsychological testing scores listed in Table 1. Thus, differences in performance between the two subgroups of aged individuals cannot be directly attributed to differences among these basic descriptors.

2.3. Summary

Overall, training session timing accuracy was good for both young and old participants. Both young and

old participants made errors during the test session. These errors were duration-dependent to the extent that participants over-produced the 6 s target interval while accurately producing the 17 s interval. The aged groups' errors made while producing the 6 s interval were about 2 s in magnitude, about twice the size of the errors made by the young participants. Subsequent analysis of the median peak time data revealed two subgroups of individuals. The majority of aged participants showed a lengthening of productions of the 17 s target interval in the test session, an effect that is consistent with standard timing theories of memory (Gibbon et al., 1984; Meck, 1983), and attention (Meck & Church, 1983; Zakay & Block, 1997). The minority produced very short 17 s intervals, a pattern reminiscent of the PD-related migration effect (Malapani et al., 1998, 2002).

In the omnibus analyses, violations of the scalar property occurred in the aged participants collapsing over sessions, but not in the testing session alone as with the timing inaccuracy effects. Further analysis of subgroups of aged participants indicated two sources of the observed effects: (1) the training session condition and (2) the subgroup of aged participants that under-produced the long target in the testing session.

3. Experiment 2: Line-length production

Perceptual discrimination of line-length was used in a critical study that established patterns of timing deficits among a number of groups of neuropsychological interest (Ivry & Keele, 1989). That study used a perceptual comparison technique intended to complement the perceptual comparisons among time-intervals. Our emphasis is on time production in discrete-trials designs, so we have employed a novel task that requires participants to manually control the length of a line in order to match the length of a previously demonstrated standard. This study is therefore the first report of results from this task. The principle design considerations beyond the use of production instead of comparison were the removal of any temporal cues from the task, and a general structuring of the procedure to match that used in the time production experiments.

3.1. Method

3.1.1. Apparatus

A custom-designed computer application was used to deliver the stimuli and record the responses. Stimuli were presented on a color CRT monitor with a 14 in. (35.56 cm) diagonal. Responses were recorded through the keyboard. Participants were tested in a 9 ft. × 9 ft. windowless room dedicated to running experiments.

3.1.2. Participants

Participants were the same as those in the Aged Group 1 from Experiment 1.

3.1.3. Procedure

The line-length production task was given to participants immediately following administration of the time production task.

3.1.3.1. Training. Line-length production training occurred during the first experimental session. Production of lines 120 pixels and 340 pixels in length were trained in separate blocks of trials. The line-length trained first was counterbalanced across participants. Training for each line-length consisted of a total of 80 trials of various sorts divided into two sets of 10 and one set of 60 trials. The first set of 10 fixed-length trials each presented a red rectangle of a fixed length placed at a random location on the screen. Though the rectangle was placed at a random point on the screen, the long axis of the rectangle was always in parallel with the top and bottom boundaries of the screen.

The second set of trials consisted of 10 production trials with feedback during the ITI. Each production trial presented a blue rectangle of a random length. Participants used two response keys to decrease or increase the length of the rectangle to match the standard that they were shown in the fixed-length trials. Participants were able to decrease the line-length by pressing the "4" key and increase it by hitting the "6" key. The rate of decrement and increment ranged from 2 to 9 pixels per press and was randomized each trial. This meant that participants could not reliably reproduce the length of the rectangle by pressing the keys for either a fixed number of times or a fixed duration. ITI feedback consisted of a pie graph that indicated to the participant the relative error of that trial's production. The participants terminated trials by pressing the "enter" key when they judged that the line on the screen equaled the standard in length.

In the third block, participants were given 15 fixed-length trials, 30 production trials with feedback, and 15 production trials without feedback.

No digit distracters were used during line-length training.

3.1.3.2. Testing. In the second session, participants received one block per line-length of 60 peak trials without feedback. The line-length tested first was counterbalanced between participants independently of the order of training. Digit distracters were not used.

3.1.3.3. Data analysis. The line-length task data were analyzed in a manner analogous to the timing task data. Responses in each production trial were recorded as a single value in pixels indicating the length of the line at the point that the participants terminated the trial. Each

participant contributed one distribution of such values for training and testing of the two standards. Absolute response–frequency distributions with a bin width of 2 pixels were created from these values. Relative response–frequency histograms, average relative frequency histograms, and normalized histograms were created using the same methods described for Experiment 1. The median, SIQR, and CV of lengths derived from these histograms were treated statistically using separate ANOVAs with two within-participants effects (session and length of standard) and one between-participants effect (age group). All probabilities reported with ANOVA effects are computed for 1 and 28 degrees of freedom.

3.2. Results

Fig. 4 shows distributions of responses from Experiment 2. The left panel depicts performance from the training session and the right panel shows data from the testing session. As can be seen in Fig. 2, production of line-lengths was accurate on both days. This contrasted with the distorted timing of older participants during testing in Experiment 1 (see Fig. 1).

Statistical analysis confirmed these observations. (Mean values from this and subsequent analyses are

found in Table 4. The complete results of the ANOVA are found in Table 5.) Only the effect of line-length significantly affected the median produced lengths.

Analysis of the SIQR indicated that productions of the longer line were more variable than productions of the shorter line. In addition, the SIQR was larger for the older participants than for the younger. The interaction of length and group was also significant reflecting the substantially larger increase in variability between production of short and long lines for the older participants. The following effects were not significant: session, the interaction of session and group, the interaction of session and length, and the interaction of session, length, and group.

The statistical analysis of the CVs indicates that the CVs of the older group are larger than those of the younger group. The effect of session was also significant, reflecting a reduction in relative production variability for both groups and both line-lengths in the testing session. This effect is similar to the one observed in the time production task and may stem from participants’ effort to vary their responding by trying new strategies on a trial-to-trial basis in response to the feedback. There was a marginal effect of the interaction between length and group, reflecting a tendency for the young group to be slightly more variable in production of the shorter line

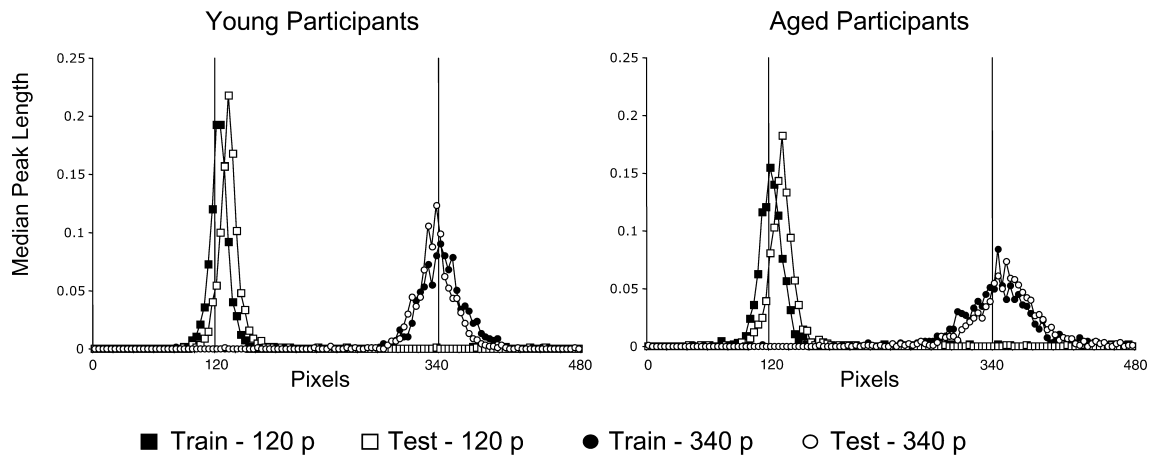


Fig. 4. Average relative frequency histogram showing distributions of individual trials’ line-length responses for 120 and 340 pixel targets during training and test sessions for the young (left panel) and aged (right panel) participants in Experiment 2. Each function is the within-group average of histograms derived for each individual participant.

Table 4
Experiment 2—Mean and standard errors of median, SIQR, and CV of produced length

Session	Group	Standard (p)	Median	SIQR	CV
Training	Young	120	120.47 (0.93)	5.67 (0.43)	0.047 (0.004)
		340	339.93 (2.20)	14.57 (1.22)	0.043 (0.004)
	Aged	120	119.80 (1.66)	7.83 (0.74)	0.065 (0.006)
		340	342.80 (3.90)	22.20 (2.43)	0.065 (0.007)
Testing	Young	120	130.67 (4.29)	6.28 (0.96)	0.047 (0.006)
		340	332.5 (6.67)	11.78 (1.33)	0.036 (0.004)
	Aged	120	129.20 (11.16)	7.36 (0.98)	0.057 (0.006)
		340	351.27 (18.36)	21.78 (3.54)	0.061 (0.007)

Table 5
Experiment 2—ANOVA table by effects (rows) and dependent variables (columns)

Effect	df	Median		SIQR		CV	
		MSE	F	MSE	F	MSE	F
<i>Source—Between subjects</i>							
Group	1	712.97	0.54	818.37	8.79*	0.010	7.84*
Error (group)	28	1309.22		93.053		0.001	
<i>Source—Within Subjects</i>							
Session	1	798.25	0.63	17.73	0.68	0.0007	3.89**
Session × group	1	427.52	0.34	3.13	0.12	0.00004	0.24
Error (session)	28	35646.92		25.92		0.0002	
Line-length	1	1407358.50	1814.89*	3497.85	90.00*	0.0003	1.69
Line-length × group	1	1059.10	1.37	388.35	9.99*	0.0006	3.89**
Error (line-length)	28	21712.58		38.87		0.0002	
Session × line-length	1	646.35	0.86	20.94	0.85	0.00002	0.10
Session × line-length × group	1	522.92	0.69	22.425	0.91	0.0002	1.48
Error (session × line-length)	28	752.82		24.67		0.0002	

Note. ANOVA results reflect independent univariate analysis of peak lengths. MSE is the mean squared error for the indicated effect. *F* is the ratio of the MSE for the indicated effect and the MSE for the error listed next below it.

* Significant effect at the $p < .05$ level.

** Marginal effect at the $p < .1$ level.

with respect to the median length of reproductions. There were no effects of the interaction of session and group, length, the interaction of session and length, and the interaction of session, length, and group. Overall, lack of significant effects involving line-length indicates that for both groups of participants, line-length reproduction variability is described by Weber's law.

To summarize, production of line-length became more variable with age and in the presence of feedback while preserving Weber's law, and accuracy was excellent in all conditions. This is in striking contrast to the dramatic distortion in accuracy found in the timing task when older participants performed in the absence of feedback.

4. Experiment 3: Line-length production

An important difference between the line-length production procedure reported in Experiment 2 and the time production tasks reported in Experiment 1 was the absence of the digit distracter used in the timing task. We originally conceived of the digits as a means of preventing chronometric counting (Rakitin et al., 1998). The line-length task reported in Experiment 2 was created without any dynamic component other than the participants' own intermittent key-presses. The digits were therefore excluded because timing was occurring and because the participant could simply alternate responding and reading the digits. However, the digits can also be conceived as an attention-demanding task in its own right that may contribute to the overall difficulty of the timing task. It is therefore possible that the observed difference in relative variability in time and line-length production does not reflect the intrinsic variability in

timing and line-length production but rather the relative difficulty of a dual-task condition (time production with digits) compared to a single task condition (line-length production without digits). To address this concern we developed a second novel line-length production task that has a dynamic component, and included the digit distracter. We were particularly interested in whether the relative variability of line-length production in this new task would be more similar to that observed in the previous line-length task or that observed in the time production experiments.

4.1. Method

4.1.1. Apparatus

Data were collected using a Macintosh G3 Powerbook equipped with a custom data collection program. Data from aged participants were collected in the same senior centers as Aged Group 2 in Experiment 1. Data from young participants were collected in the NYSPI facility described in Experiment 1.

4.1.2. Participants

Sixteen young participants (six males and 10 females) were recruited from Columbia University. The mean age of these participants was 24.3 years, and they had an average of 16 years of education. None had a history of neurological or psychiatric disorders, or were currently taking psychoactive medication.

Sixteen aged individuals (six males and 10 females) participated in this study. Six of the 16 aged participants in this experiment were also included in Aged Group 2 of Experiment 1. The remainder were recruited from the same senior centers from which Aged Group 2 of Experiment 1 were recruited. The demographics and neuropsych-

chological profile are found in the third column of Table 1. Overall this group was very similar to the Aged Group 2 participants from Experiment 1 with the exception of the estimated IQ and WAIS-III colored matrices measures, which were closer to that of Aged Group 1.

4.1.3. Procedure

This experiment was conducted approximately one year after the data were acquired from the Aged Group 2 participants of Experiment 1.

4.1.3.1. Training. Line-length production training occurred during the first experimental session. Productions of lines 120 pixels and 340 pixels in length were trained in separate blocks of trials. Half of the participants were trained on the 120 pixels first and half were trained with the 340 pixels first. The line-length to be trained first was counterbalanced among participants. Training for each interval consisted of a total of 80 trials divided into two sets of 10 and one set of 60 trials. The first two sets of trials, or pre-training blocks, served to acquaint participants with the length to be tested in the third set of trials, the training block.

The first pre-training block consisted of 10 “fixed-length trials.” On each trial a blue line was placed at a random location on the monitor. The line would elongate at varying speeds until it reached its target length. After the target length had been reached, the blue rectangle turned magenta. Participants were instructed to try to remember the length of the line delineated by the size of the rectangle on the screen up to the subsequent color change. Participants were informed that they should refrain from counting or tapping so as to avoid extraneous timekeeping. Random digits were intermittently placed above the center of the rectangle, and participants were instructed to read the digits aloud. The algorithm used to determine the inter-digit intervals was the same as that described in Experiment 1. The digits were presented in a random font size (between 8 and 48 point) so that the size of the digit would not serve as a reliable reference for the lines’ length.

The second pre-training block consisted of 10 “peak trials” in which participants were asked to reproduce the line-length that they had just observed in the preceding fixed-length trials. In these peak trials, the blue rectangle did not turn red, but instead continued to elongate and remained blue. Participants were instructed to respond by pressing the spacebar just before they felt that the length of the target line-length had been reached and should create a window of response around the target line-length. That is, participants were to “bracket” the target length by pressing, holding, and the releasing the space bar. Digit distracters were employed, as described above. During the ITI, participants were shown a pie graph that indicated to them the relative error of their trial’s production. Additional messages informed the

participants if they had responded too short or too long with respect to the target length.

In the third block of the training session, participants received 15 fixed-length trials, 30 peak trials with feedback during the ITI, and 15 peak trials without ITI feedback. Digit distracters were used in all trials for both groups.

4.1.3.2. Testing. In the second session, participants received only one block of 60 peak trials without feedback for each of the two target lengths. Distracter digits were used in all trials. The order of testing of the two intervals was counterbalanced so that half of the participants trained first on the 120 pixel line were tested first on the 120 pixel line.

4.1.3.3. Analysis. The dependent variables in this Experiment were derived in a manner similar to those from the time production experiments. The length of the line when the subject began responding and terminated responding were averaged to determine the “peak length” line-length production on each trial. The median, SIQR, and CV of these peak lengths were then computed within condition.

The statistical model used in the ANOVA was identical to that described in Experiment 2.

4.2. Results

Fig. 5 shows distributions of responses from Experiment 3. The left panel depicts performance from the training session and the right panel shows data from the testing session. As can be seen in Fig. 5, production of line-lengths was accurate on both days. As with the results of Experiment 2, these results contrasted with the distorted timing of older participants during testing in Experiment 1 (see Fig. 1). The main difference between the data presented in Figs. 4 and 5 is that the response distributions are wider in Fig. 5.

The mean values and standard errors of the dependent variables are reported in Table 6. Table 7 reports the results of the ANOVA. The main pattern of statistical results is similar to those reported in Experiment 2. The length of the target line was the only significant determinant of the median lengths of the produced lines. Overall, production accuracy was excellent and equivalent to that reported in Experiment 2.

The absolute variability of production measured by the SIQR of peak length productions was significantly greater in the aged group compared to the young group. Absolute variability was greater for the production of the longer line, and this difference was significantly larger in the aged group than in the young group. Relative variability measured by the CV of median production was greater for the aged group than for the young group, but was not significantly affected by either the

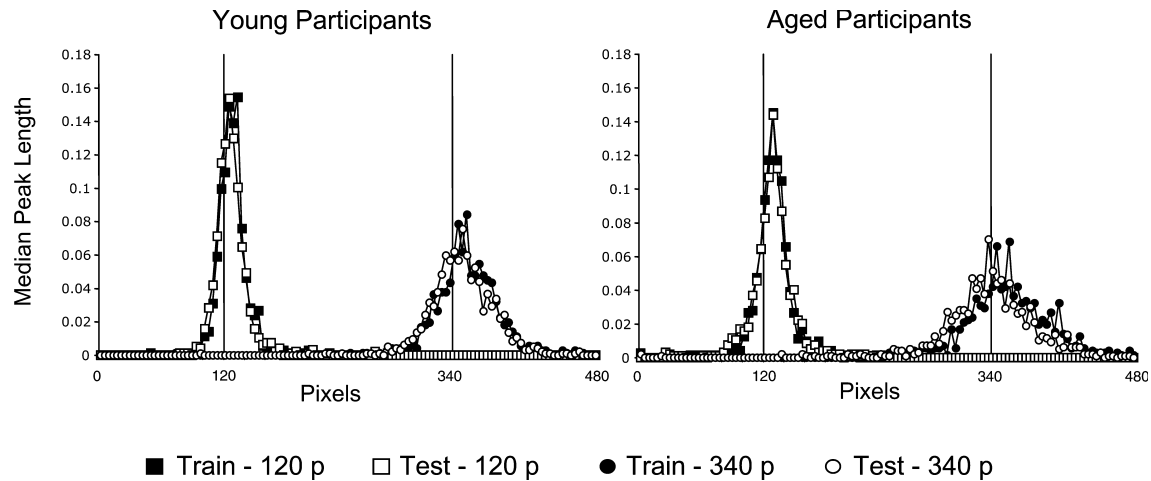


Fig. 5. Average relative frequency histogram showing distributions of individual trials' peak lengths for 120 and 340 pixel targets during training and test sessions for the young (left panel) and aged (right panel) participants in Experiment 3. Each function is the within-group average of histograms derived for each individual participant.

Table 6
Experiment 3—Mean and standard errors of median, SIQR, and CV of peak length

Session	Group	Standard (p)	Median	SIQR	CV
Training	Young	120	127.50 (4.65)	7.53 (0.66)	0.060 (0.005)
		340	341.50 (6.47)	18.01 (1.44)	0.052 (0.004)
	Aged	120	126.53 (3.30)	9.98 (1.51)	0.077 (0.009)
		340	352.19 (2.29)	26.09 (3.24)	0.074 (0.009)
Testing	Young	120	123.00 (2.85)	8.30 (0.84)	0.067 (0.007)
		340	344.84 (7.29)	18.53 (1.23)	0.054 (0.004)
	Aged	120	125.88 (4.59)	11.35 (1.97)	0.089 (0.013)
		340	335.69 (10.82)	27.29 (4.03)	0.081 (0.011)

Table 7
Experiment 3—ANOVA table by effects (rows) and dependent variables (columns)

Effect	df	Median		SIQR		CV	
		MSE	F	MSE	F	MSE	F
<i>Source—Between subjects</i>							
Group	1	23.63	0.03	999.18	5.85*	0.016	5.17*
Error (group)	30	688.93		170.70		0.003	
<i>Source—Within Subjects</i>							
Session	1	670.70	1.44	29.91	0.72	0.002	3.42
Session × group	1	512.00	1.10	3.32	0.08	0.0002	0.36**
Error (session)	30	466.78		41.47		0.0005	
Line-length	1	1518370.95	2775.05*	5563.48	105.19*	0.002	3.37
Line-length × group	1	0.281	0.00	256.30	4.85*	0.0002	0.30
Error (line-length)	30	547.15		52.89		0.0006	
Session × line-length	1	128.00	0.24	0.34	0.01	0.0002	0.49
Session × line-length × group	1	1122.20	2.12	0.001	0.00	1E-7	0.00
Error (session × line-length)	30	528.94		35.97		0.0003	

Note. ANOVA results reflect independent univariate analysis of peak length. MSE is the mean squared error for the indicated effect. F is the ratio of the MSE for the indicated effect and the MSE for the error listed next below it.

* Significant effect at the $p < .05$ level.

** Marginal effect at the $p < .1$ level.

length of the target or the interaction of target line-length and age group. The absence of these two latter effects indicates that the changes in absolute variability are in proportion to the length of the lines produced by

the two groups in accordance with Weber's law. Inspection of Table 6 indicates that scaling is in fact excellent by both groups with a maximum difference of 0.008 demonstrated by the young participants during the

training session and the aged participants during the testing session.

The line-length production variability found in this experiment was somewhat greater than that found in Experiment 2, both in absolute and relative terms. Comparing the values in Tables 4 and 6, one can see that the absolute variability in the current experiment's training session was about 2–4 pixels greater than those found in Experiment 2. The difference between the testing sessions' values ranged from 2 to 7 pixels with both the minimum and maximum differences found in the young groups' data. The differences in relative variability ranged from a minimum of 0.009 found between the young groups' training session production of the long line, to 0.032 found between the aged groups' testing session productions of the short line. Overall, it is clear that production of a line the length of which is changing dynamically in the presence of distracting digits is more variable than line-length production that occurs under the simpler conditions found in Experiment 2. However, these changes neither increased variability to the degree typically found in time production, nor induced age-related production inaccuracies.

To summarize, these results concur with those of Experiment 2, and emphasize that while age-related production inaccuracy is specific to timing, age-related increases in variability are not specific to timing.

5. Discussion

The current study extends previous work on older adults' interval timing by examining the effects of multiple durations and an extended retention period between training and test. We found an intriguing set of results: the majority of older adults showed a simple rightward shift of their peak functions from training to test, as might be predicted by age effects on either the attention or memory encoding–decoding speed (K^*) aspects of information-processing models such as SET (Meck, 2002). However, a subset of older adults showed a very different pattern, with a leftward shift of the function for the longer duration. The “migration” of the long duration towards the short duration shown by this subset of older adults is reminiscent of previous findings in PD patients, where the effect is modulated by dopamine administration (Malapani et al., 1998, 2002). These data raise the possibility that findings of the migration effect in apparently healthy older adults may be a very early, sensitive measure of age-associated cognitive decline.

Accurate time production by aged participants has been observed previously (Block et al., 1998; Wearden et al., 1997). In the present case accurate production occurred in the test session, which included both feedback and demonstrations of the target, two factors previously suggested to constitute favorable conditions for

timing in aged individuals (Wearden et al., 1997). Over-production of time intervals, as well as a tendency for age-related over-productions to reflect an exaggeration of effects observed in young participants, have been also reported previously (Perbal et al., 2002). New to this study is the association of this age-related effect, and the conditions of the test session, which omitted feedback and followed a 24 h retention interval. This suggests that this effect in both aged in young participants is due to age-related changes in temporal memory, not simply psycho-motor slowing (Perbal et al., 2002). Moreover, the tendency for aged individuals to over-produce intervals is consistent with SET predictions regarding the effects of memory on timing to the extent that productions errors were in the same direction for both target intervals. Supporting this argument is the evidence for the scalar property of timing variability in the majority of aged participants, a feature associated with K^* effects in SET (Gibbon et al., 1984; Meck, 1983, 2002). The findings of a smaller effect in young participants, and a complete absence of production errors in either line-length production task argues against these effects being due to a general decline in episodic memory performance commonly found in elderly individuals (Raz, 2000). If these production errors are due to memory effect, they are isolated to temporal memory.

However, in considering a memory explanation of the majority's timing errors one must consider a few additional facts. First, the production errors found in those participants that over-produced both target intervals were larger for the shorter interval. Second, the aged individuals considered as a single group violated the scalar property. While the asymmetry in error magnitudes is not a specific feature of attention hypotheses in timing theory, neither SET nor AGM specifically exclude the possibility. However, violations of the scalar property are a specific feature of the SET implementation of attention (Meck & Church, 1983). Nonetheless, neither the young participants nor the majority subgroup of aged participants significantly violated the scalar property. Taken together the data weigh in favor of the memory explanation.

The pattern of timing errors shown by the minority of older adults that over-produce the 6 s interval and under-produce the 17 s interval are similar, but not identical to those previously shown by PD patients timing without dopaminergic medication (Malapani et al., 1998, 2002). Healthy older adults errors occur only in the testing session, whereas PD patients off dopamine replacement therapy will show the migration effect in either the training or testing sessions. In addition, it is unclear whether the testing session errors apparent in the healthy older adults exhibit the scalar property of variability absent in the PD migration effect. These differences suggest at a minimum that the age-related deficit is less severe than the PD-related deficit, and so

can be overcome by the availability of feedback. On the other hand, the differences may imply an additional mechanism at play in PD that obviates the usefulness of feedback and adds a non-timing related variability component.

Violations of the scalar property by this subset of older adults in the test session, if ultimately determined to be reliable, might suggest that their errors were the result of differences in motor or decision processes, rather than timing per se. However, the results of the line-length tasks weigh against this possibility. The line-length tasks were very similar to the timing task in terms of their demands on motor and decision processes, but did not reveal the same migration effects in accuracy, or violations of the scalar property in variability, in any of the participants. However, the line-length production tasks did show age related increases in variability that adhered to Weber's law. It is therefore possible that there are four separate factors at work in the aged participants' timing data: (1) a general increase in production variability, (2) a timing specific increase in variability that violates the scalar property, (3) a tendency to over-produce time intervals, and (4) a PD-like migration effect.

To summarize, our application of the "encode–decode" procedure previously used with PD patients (Malapani et al., 1998, 2002) to a group of apparently healthy older adults revealed a striking separation between two different groups of older adult participants. One group showed an overly long production of both the short and long durations, in keeping with previous findings of older adults' timing performance and information-processing models such as SET (Gibbon et al., 1984; Meck, 1983; Meck & Benson, 2002). Another group showed an overly long production of the short duration and an overly short production of the long duration, similar to PD patients.

The different direction of effects for the short and long durations in the minority of aged participants is very difficult to explain using attention and memory encoding–decoding speed components of information-processing models such as SET. Two early attempts at theoretical accounts of this account are based on published artificial neural-network models of timing (Miall, 1996). The first permits duration-dependent errors by altering the accumulator so that its current value is an increasing but decelerating function of real time (Malapani & Rakitin, 2003). The second introduces duration-dependent effects in a physiologically viable model by permitting linking of coincidence-detection neurons in the striatum (Matell & Meck, 2000, 2004). Both models could also be used to create hybrid memory effect that would account for the asymmetry in error magnitudes in over-productions of both target intervals.

Besides their theoretical interest, these results may have important practical implications. Does the migra-

tion effect shown by this subset of apparently healthy older adults reflect some difference in their neurocognitive status versus that of older adults who do not show the effect? This question is especially interesting given the need for early indicators of Parkinson's disease, and perhaps Alzheimer's disease, and other age-related dementias to identify targets for rapid treatment. Our results suggest that the encode–decode paradigm may be an important tool for such early detection. Currently, we are following up on the results reported here by using behavioral, pharmacologic, and neuroimaging methods to examine the performance of young adults, normal older adults, and PD patients in related procedures.

References

- Agid, Y., Cervera, P., Hirsch, E., Javoy-Agid, F., Lehericy, S., Raisman, R., et al. (1989). Biochemistry of Parkinson's disease 28 years later: A critical review. *Movement Disorders*, 4, S126–S144.
- Agid, Y., Ruberg, M., Dubois, B., & Pillon, B. (1987). Anatomoclinical and biochemical concepts of subcortical dementia. In S. M. Stahl & S. D. Iversen (Eds.), *Cognitive neurochemistry* (pp. 248–271). Oxford: Oxford University Press.
- Arnsten, A. F., Cai, J. X., Steere, J. C., & Goldman-Rakic, P. S. (1995). Dopamine D2 receptor mechanisms contribute to age-related cognitive decline: The effects of quinpirole on memory and motor performance in monkeys. *Journal of Neuroscience*, 15, 3429–3439.
- Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgements: A meta-analytic review. *Psychology and Aging*, 13, 584–596.
- Catania, A. C. (1970). Reinforcement schedules and psychophysical judgments: A study of some temporal properties of behavior. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 1–42). New York: Appleton-Century-Crofts.
- Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics*, 61, 549–560.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77.
- Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, 7, 170–184.
- Fahn, S. (1995). Parkinson's disease. In L. P. Rowland (Ed.), *Merritt's textbook of neurology* (9th ed.). Baltimore: Williams & Wilkins.
- Goldman-Rakic, P. S., & Brown, R. M. (1981). Regional changes of monoamines in cerebral cortex and subcortical structures of aging rhesus monkeys. *Neuroscience*, 6, 177–187.
- Ivry, R. B., & Hazeltine, E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 3–18.
- Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1, 134–150.
- Krampe, R. T., Engbert, R., & Kliegl, R. (2002). The effects of expertise and age on rhythm production: Adaptations to timing and sequencing constraints. *Brain and Cognition*, 48, 179–194.
- Lustig, C., & Meck, W. (2001). Paying attention to time as one gets older. *Psychological Science*, 12, 478–484.

- Malapani, C., Deweer, B., & Gibbon, J. (2002). Separating storage from retrieval dysfunction of temporal memory in Parkinson's disease. *Journal of Cognitive Neuroscience*, *14*, 31–322.
- Malapani, C., & Rakitin, B. C. (2003). 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* (pp. 486–514). Boca Raton, FL: CRC Press.
- Malapani, C., Rakitin, B. C., Fairhurst, S., & Gibbon, J. (2002). Neurobiology of timing. *Cognitive Processing*, *3*, 3–20.
- Malapani, C., Rakitin, B. C., Levy, R., Meck, W. H., Deweer, B., Dubois, B., et al. (1998). Coupled temporal memories in Parkinson's disease: A dopamine related dysfunction. *Journal of Cognitive Neuroscience*, *10*, 316–331.
- Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behaviour. *BioEssays*, *22*, 94.
- Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: Coincidence-detection of oscillatory processes. *Cognitive Brain Research*, *21*, 139–170.
- McCormack, T., Brown, G., Maylor, E., Darby, R., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology*, *35*, 1143–1155.
- Meck, W. H. (1983). Selective adjustment of the speed of the internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 171–201.
- Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research*, *3*, 227–242.
- Meck, W. H. (2002). Choline uptake in the frontal cortex is proportional to the absolute error of a temporal memory translation constant in mature and aged rats. *Learning & Motivation*, *33*, 88–104.
- Meck, W. H., & Benson, A. M. (2002). Dissecting the brain's internal clock: How frontal-striatal circuitry keeps time and shifts attention? *Brain and Cognition*, *48*, 195–211.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 320–334.
- Miall, R. C. (1996). Models of neural timing. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks and movement* (pp. 69–94). North-Holland: Elsevier Science, Amsterdam.
- Perbal, S., Droit-Volet, S., Isingrini, M., & Pouthas, V. (2002). Relationships between age-related changes in time estimation and age-related changes in processing speed, attention, and memory. *Aging Neuropsychology and Cognition*, *9*(3), 201–216.
- Rakitin, B. C., & Stern, Y. (2002). Parkinson disease. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (Vol. 3, pp. 458–462). London: Nature Publishing Group.
- Rakitin, B. C., Gibbon, J., Penney, T. B., Malapani, C., Hinton, S. C., & Meck, W. H. (1998). Peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 19–33.
- Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik, & T. A. Salthouse, (Eds.), *The handbook of aging and cognition*. (2nd ed., pp. 1–90). Mahwah, NJ: Erlbaum.
- Rao, S., Mayer, A., & Harrington, D. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, *4*, 317–323.
- Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 242–268.
- Schecter, D., Singer, T. M., Kuperman, J., & Endicott, J. (1998). Selection of normal control participants for psychiatric research: Update on a model for centralized recruitment. *Psychiatry Research*, *79*, 175–185.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs*, *77*, 1–13.
- Vanneste, S., & Pouthas, V. (1999). Timing in aging: The role of attention. *Experimental Aging Research*, *25*, 49–67.
- Vanneste, S., Pouthas, V., & Wearden, J. H. (2001). Temporal control of rhythmic performance: A comparison between young and old adults. *Experimental Aging Research*, *27*, 83–102.
- Wearden, J., Wearden, A., & Rabbitt, P. (1997). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 962–979.
- Wenk, G. L., Pierce, D. J., Struble, R. G., Price, D. L., & Cork, L. C. (1989). Age-related changes in multiple neurotransmitter systems in the monkey brain. *Neurobiology of Aging*, *10*, 11–19.
- Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science*, *6*, 12–16.