Journal of Experimental Psychology:
Human Perception and Performance
1997, Vol. 23, No. 4, 962–979


0096-1523/97/$3.00

# Age and IQ Effects on Stimulus and Response Timing

J. H. Wearden, A. J. Wearden, and P. M. A. Rabbitt
University of Manchester

Normal older participants (aged 60–79 years), with known scores on the Culture Fair Intelligence Test, were tested on 4 timing tasks (i.e., temporal generalization, bisection, differential threshold, and interval production). The data were related to the theoretical framework of scalar timing theory and ideas about information processing and aging. In general, increasing age and decreasing IQ tended to be associated with increasing variability of judgments of duration, although in all groups events could be timed on average accurately. In some cases (e.g., bisection), performance differences between the older participants and students nearly 50 years younger used in other studies were negligible.

Anecdotes and some experimental findings suggest that time experience and behavior may change with increasing age from adulthood. As usual, in the psychology of time an initial problem is to decide on interrelations between diverse findings that may have no common theoretical basis (Wearden, 1994). For example, various authors have reported an apparent speeding up of phenomenological time experience with age (e.g., "Christmas seems to come round quicker every year"), a change claimed to be sufficiently orderly to warrant mathematical description and reification into "laws" (e.g., Doob, 1971; Lejeune & Pouthas, 1991; Lemlich, 1975), but it is unclear what relation this effect may have to judgments that people of different ages make about the duration of events in laboratory experiments (e.g., those that require the production of time intervals with specific durations).

Laboratory studies are few and inconsistent. For example, Feifel (1957) required older and younger participants to produce intervals of 30, 60, 180, and 300 s and found that although productions from both groups underestimated the real-time duration, older participants underestimated it more, a finding not replicated in a similar study by Surwillo (1964), in which no clear age differences were found. McGrath and O'Hanlon (1968) studied the production of longer durations (1–8 minutes) and again found that older participants produced shorter durations, particularly at the shorter time values, and Kline, Holleran, and Orme-Rogers (1980) found the shortest durations produced by older participants when they counted to produce intervals from 30 to 90 s. In one of the few studies to report the variability of performance, as opposed to simply averages, Licht, Morganti, Nehrke, and Heiman (1985) found a higher relative variability of times produced in older (60–70+) participants than younger ones, although the effects on means were not monotonic with increasing age. In a more recent experiment involving learning to produce a duration between 4 and 6 s, Lejeune and Pouthas (1991) found that older participants took longer to learn the task but that they eventually adjusted to the feedback contingency used.

It is clear that previous studies of timing in older people have many deficiencies. For one thing, little attempt has been made to relate data to existing psychological theorizing about time, perhaps because the growth in such theorizing has been recent (for recent work on time psychology, see Block, 1990; Gibbon & Allan, 1984; Macar, Pouthas, & Friedman, 1992), and the range of tasks used has been somewhat limited, but, at the same time, the theoretical perspectives used and general investigative aims have been heterogeneous. Another problem has been that the participants used were, in many cases, institutionalized, thus possibly restricting the generalizability of the results to the normal older population. In addition, little attempt has been made previously to relate time judgments to more general theories of changes in cognitive processes with aging.

An old idea is that timing in humans may be mediated by an internal clocklike mechanism (see Wearden & Penton-Voak, 1995, for a historical review), and this raises questions about possible relations between any change in the speed of the pacemaker of this clock and cognitive processing, particularly the rate of information processing. A large amount of literature documents the slowing of information processing as the earliest, and most marked, symptom of cognitive aging (Birren, Woods, & Williams, 1979; Cerella, 1985; Salthouse, 1985, 1991), although the relation between "clock speed" and a person's information-processing rate may not be straightforward. If slowing of the speed of the pacemaker of an internal clock (e.g., a decrease in the number of "ticks" occurring per unit time) occurs with age, some changes in timing behavior might be expected. However, internal clocks that tick more slowly do not necessarily underestimate durations compared with those that operate at higher speeds, as the participant may learn to recalibrate

J. H. Wearden and A. J. Wearden, Department of Psychology, University of Manchester, Manchester, England; P. M. A. Rabbitt, Age and Cognitive Performance Research Unit, University of Manchester, Manchester, England.

This research was supported by grants from the United Kingdom Medical Research Council. We are grateful to Scott Hofer for helpful discussions about the proper conduct of principal-components analysis.

Correspondence concerning this article should be addressed to J. H. Wearden, Department of Psychology, University of Manchester, Manchester M13 9PL, England. Electronic mail may be sent via Internet to wearden@psy.man.ac.uk.

behavior to take into account the speed change (i.e., some unit time becomes identified with fewer ticks than before). Mathematical explorations of the effects of clock speed on the output of various plausible sorts of internal clocks (e.g., Gibbon, 1977; Killeen & Weiss, 1987), however, indicate that pacemaker slowing generally will increase the variability of time judgments (at least in situations in which pacemaker speed is a major source of behavioral variability; see Gibbon, Church, & Meck, 1984, for a discussion), and some studies (e.g., Lejeune & Pouthas, 1991; Licht et al., 1985) have indeed shown that time judgments become more variable in older individuals.

Any increasing variability of temporal judgments with increasing age may have important implications for understanding age-related changes in a task commonly used in studies of cognitive aging, that of *choice reaction time* (CRT), a task from which evidence for slowing of information-processing rate with increasing age is often adduced. Here, participants must rapidly make one of a number of responses to associated stimuli, and the usual finding is that older people have longer CRTs than younger ones. However, mean reaction times (RTs) are poor summary statistics because RT distributions have large standard errors and large rightward skews. Some authors have (e.g., Rabbitt & Goward, 1994; Rabbitt & Vyas, 1973; Smith & Brewer, 1995) pointed out that in laboratory CRT tasks, from which most data on age-related slowing have been collected, the fastest correct responses made by older and young adults may differ only by 15–50 ms. In the same experiments, differences in the mean CRTs of 200 ms or more are almost entirely attributable to the fact that the variance and skew of the CRT distributions increases markedly with age. Thus, rather than regarding increases in mean CRT with age as the results of slowing of central nervous system processes, it is more plausible to regard them as consequences of increased variability of response times, resulting from progressive loss of temporal control of response speed. Consistent with this is Smith and Brewer's (1995, p. 244) observation that trial-to-trial adjustment of response speed is "coarser" in older people than younger ones, and Rabbitt and Vyas's (1973) observation that "precision of temporal tracking" to determine the optimal RT band at which speed is optimized and errors minimized, deteriorates in old age. On these premises it might be argued that there is a direct link between variability in precision of estimation of brief time intervals, and variability of CRT with age, or intelligence. The study of timing in older individuals thus may have more than just intrinsic interest.

In this article we present data from four experiments investigating timing behavior in normal older participants. An advantage of this research over some previous investigations is that we used a consistent general theoretical framework of non-counting-based timing, and some tasks could be related specifically to recent work with student participants carried out within the framework of *scalar timing theory* (Gibbon, 1977; Gibbon et al., 1984). For present purposes, its minutiae are not directly relevant (see Wearden, 1991a, 1994, and Wearden & Lejeune, 1993, for reviews). We note here only that the theory is an account of observed behavior in the form of a *clock-comparison model.* The basic premise is that participants time event duration and regulate their own behavior in time using the output of an internal clock (although the complete model involves much more than just a clock; see Gibbon et al., 1984). The time representations possessed by participants are supposed to have two critical properties: *mean accuracy* (i.e., the assertion that the average representation of some clock-measured time, *t* seconds, is equal to *t*, or nearly so) and the *scalar property* (i.e., the assertion that the standard deviation of time representations grows linearly with the mean, in effect proposing that a coefficient of variation statistic [standard deviation/mean] remains constant as the duration timed, *t*, changes). The latter property is a form of conformity to Weber's law (Gibbon, 1977) and implies a timing mechanism with constant sensitivity as the duration to be timed varies.

Attempts to apply scalar timing theory to humans have tended to concentrate on situations in which uniquely human timing resources such as explicit chronometric counting are not used (Wearden, 1991a). Rather than attempt to prevent counting by imposing a concurrent interfering task, which would draw attention away from the timing task, which is the focus of interest, most recent experiments have concentrated on humans' timing of short durations (usually less than 1 s), for which chronometric counting may not be spontaneously used or be useful even if attempted (see arguments by Wearden, 1991a, 1991b, 1992, 1994, 1995; Wearden & Ferrara, 1993, 1995; Wearden & McShane, 1988). It is in these situations that control by putative underlying biological timing mechanisms, common to humans and animals, such as certain types of internal clocks (Wearden, 1994), might be expected to manifest itself. We realize that this domain of non-counting-based timing represents only a small proportion of the participant matter of time psychology and that discoveries made within this framework may not transfer to other areas. On the other hand, it may represent a domain in which orderly data can be readily obtained and easily treated with consistent theoretical approaches.

Our research was conducted using active, older community residents for whom psychometric test data on general intellectual ability were available. Two of the experiments (temporal generalization in Experiment 1 and bisection in Experiment 2) derived directly from animal experiments conducted within the framework of scalar timing theory (Wearden, 1991b, 1992), and both used methods recently used with young adults. Experiment 3 (differential threshold determination) was not a direct analog of any animal experiment, and Experiment 4 (interval production), although sharing some aspects in common with a previous study (Wearden & McShane, 1988) that was one of the first to produce evidence for scalar timing in humans, was more concerned with learning about time than with time representations per se.

The four tasks used were chosen because (a) they had not been used previously with older individuals (e.g., even previous interval production experiments concentrated on much longer durations than used here); (b) there were previous data and established theory for three of them (Experiments 1, 2, and 4); and (c) between them they tested a range of timing abilities, including identification of absolute stimulus duration (Experiment 1), comparisons of durations of two (Experiment 3) or three (Experiment 2) stimuli, and motor timing (Experiment 4). In all experiments, one focus of interest was on the variability of the participants' performance, as well as simply the mean because, as noted earlier, any increasing variability of timing behavior with increasing age may have implications for interpreting age-related effects on commonly used RT tasks. In addition to reporting performance on the individual experiments, we also used principal-components analysis (e.g., Hair, Anderson, & Tatham, 1987) to examine possible interrelations between measures of behavior on the tasks as well as any possible relations of behavioral measures with age and intelligence.

Because a large body of evidence (e.g., reviewed by Salthouse, 1991, chap. 2) shows that general intellectual ability as measured by intelligence tests declines with increasing age, age differences on many cognitive tasks are confounded by IQ differences in random samples of older participants. In our studies, on the other hand, we matched our two age groups (60–69 years and 70–79 years) for scores on an IQ test, thus allowing us to examine age and IQ effects in groups in which the two variables were not strongly correlated.

## General Method

### *Participants*

Ninety members of the panel of volunteers at the Aging and Cognitive Performance Research Unit of the University of Manchester served as participants. Twenty-nine were men and 61 were women. The participants were divided into two groups: those aged 60–69 years ($n$ = 41) and those aged 70–79 years ($n$ = 49). The two groups were matched for scores on Scale 2 of the Culture Fair Intelligence Test (Cattell & Cattell, 1960), which had been administered to all participants during the previous 2 years. For purposes of analysis, we assigned participants to three IQ groups: low, mid, and high. A breakdown of participant numbers by age and IQ is given in the *Method* and *Results* sections of the experiments.

### *Apparatus*

Participants were tested individually in a quiet room. A Wang PC321/20S (IBM-compatible) computer controlled experimental events and recorded the data. Responses were made on the computer keyboard, and displays were presented on a super video graphics array color monitor. The experiments were controlled by programs written in Turbo Pascal, with the timing of the stimulus durations accomplished by a procedure derived from an assembly-language program. All durations were timed to a resolution of at least 1 ms. All stimuli were 500-Hz tones produced by the computer's speaker.

### *General Procedure*

All participants completed all four experiments one after the other, in a single experimental session, lasting about 40 min. The order of presentation of experiments was varied from participant to participant, with the temporal generalization, bisection, and threshold determination experiments presented to roughly equal numbers of participants in all eight possible permutations of order. The production experiment was always presented last because pilot tests suggested that participants were more likely to have difficulty following the instructions for production than for the other procedures.

## Experiment 1: Temporal Generalization

A temporal generalization task involves the identification of a stimulus with a certain absolute duration. In Church and Gibbon's (1982) original experiment, rats received presentations of stimuli (periods of darkness) of different lengths, and leverpresses were rewarded with food only after stimuli of some "standard" duration (e.g., 4 s), whereas responses after stimuli both longer and shorter than the standard were unrewarded. After several sessions of training, the rats produced stable "temporal generalization gradients" in the form of response probability plotted against stimulus duration. These gradients peaked at the standard duration (i.e., the highest probability of response was found at the standard) and were roughly symmetrical in real time, in that equal real-time differences above and below the standard produced about the same level of responding.

Wearden (1992) developed an analog for humans of this procedure, and we followed the method closely in this experiment. Tone durations were the stimuli to be judged, and all durations were short (less than 1 s) to prevent chronometric counting (which, in fact, participants did not spontaneously use). For example, in Wearden's Experiment 1, in the linear spacing condition a 400-ms tone was identified as the standard, and nonstandard tones were 100, 200, 300, 500, 600, and 700 ms. Wearden's students, whose age was not exactly determined but who were at least 45 years younger on average than the participants used in the current study, responded most, like rats, at the real-time standard duration, but their temporal generalization gradients were asymmetrical, with a 500-ms tone being confused more often with the 400-ms standard than a 300-ms tone, a 600-ms tone confused more often than a 200-ms one, and so on.

Wearden (1992) used computer simulations to fit various theoretical models to his data, including one that he called the modified Church and Gibbon (MCG) model, a modification of the model used to fit data from studies of temporal generalization in rats in the original work by Church and Gibbon (1982). Both Church and Gibbon's original model and the MCG model propose that participants carry out the temporal generalization task by comparing a "reference" memory representation of the standard with a short-term representation of the just-presented duration, $t$, and then judging the result of this comparison against a threshold for responding. Both models proposed two sources of variability from trial to trial: variability in the memory representation of the standard, $s$, and variability in the threshold, $b$.

Church and Gibbon's (1982) original model proposed that rats responded when

$$\frac{abs(s^* - t)}{s^*} < b^*,$$

where $t$ is the representation of the just-presented duration (assumed in their model to be accurately timed); $s^*$ is a sample from the memory representation of the standard, which varies from trial to trial but has an accurate mean, $s$; and $b^*$ is a sample from a threshold value that also varies from trial to trial around some mean, $b$, which is on average accurate. Inspection of the previous equation suggests that it produces symmetrical generalization gradients; for example, if $s = 4$ s, then $t$ values of 3 and 5 s both produce the same left-hand side in the equation, and are thus equally likely to meet the response criterion.

To account for the asymmetrical temporal generalization gradients he obtained from humans, Wearden (1992) proposed a changed decision rule in his MCG model. Here, participants were proposed to respond when

$$\frac{abs(s^* - t)}{t} < b^*,$$

where all terms are the same as in Church and Gibbon's (1982) original model. A quick calculation will show that the MCG model predicts asymmetrical generalization functions of the type found (e.g., a 500-ms duration would be more likely to be confused with a 400-ms standard than a 300-ms duration would be).

Along with standard statistical analysis of the data obtained in the current study of temporal generalization in older individuals, the MCG model was fit to the data, permitting an examination of possible age and IQ effects on the model's parameters as well as a comparison with parameter values from students, derived when the MCG model was fit by Wearden (1992).

The temporal generalization experiment presented here closely followed the "linear spacing" group from the first experiment of Wearden (1992). A 400-ms tone was identified as the standard, and participants were required to judge whether various tones, ranging from 100 to 700 ms in length, were or were not the standard.

## Method

*Participants.* Forty-one 60- to 69-year-old volunteers and forty-nine 70- to 79-year-old volunteers participated in this study.

*Procedure.* The standard stimulus duration was 400 ms, and the nonstandard durations were 100, 200, 300, 500, 600, and 700 ms. The experiment was arranged in eight series of trials. For each individual experimental trial, there was a nominal probability of .75 that one of the nonstandard durations would be presented and a nominal probability of .25 that the standard duration would be presented. The six nonstandard durations were presented in eight randomly ordered series. Thus, different nonstandard durations were presented successively (with presentations possibly interspersed with one or more presentations of the standard stimulus) until all six nonstandard stimuli had been presented. Another random order of the six nonstandard stimuli was then composed, and another series was presented. Thus, each nonstandard stimulus was presented eight times. The standard was presented more frequently; the mean number of presentations across all participants was 16.3. The experiment started with five presentations of the standard stimulus accompanied by the display "This is the standard length tone." Each presentation of the standard was followed by a 5-s interpresentation interval. After the standard durations had been presented, participants received the display "Press the spacebar for the next trial." A press on the spacebar was followed by a delay that was chosen randomly from a uniform distribution between 1 and 3 s and then a stimulus presentation. This was followed immediately by the display "Was that the standard length tone? Press Yes (Y) or No (N) key." The response was followed by accurate feedback appropriate to the type of response produced and whether the duration presented was the standard (e.g., "Correct, that was the standard"). The next trial then was started by pressing the spacebar.

## Results

Data from 10 participants were discarded from the analysis, 3 because of failure to understand or comply with instructions and 7 because of a procedural error, which was rectified after running the first 7 participants on the procedure. All participants who understood and complied with instructions were included in the analysis. The sample thus comprised 36 participants aged 60–69 years (mean age = 64.72 years, $SD$ = 2.96; unadjusted test score = 28.52, $SD$ = 6.38) and 44 participants aged 70–79 years (mean age = 73.93 years, $SD$ = 2.89; mean test score = 26.96, $SD$ = 6.78). These 80 participants fell into three IQ groups, 26 with test scores equal to or less than 23 (mean age = 70.08 years, $SD$ = 5.65; mean test score = 19.92, $SD$ = 3.28), 27 with test scores from 24 to 31 (mean age = 70.11 years, $SD$ = 5.45; mean test score = 27.96, $SD$ = 2.14), and 27 with test scores of 32 or more (mean age = 69.19 years, $SD$ = 5.41; mean test score = 34.82, $SD$ = 2.15). For each participant, the proportion of yes responses (i.e., identifications of a presented stimulus as to the standard 400-ms stimulus) produced at each stimulus duration was calculated. An analysis of variance (ANOVA) yielded a significant effect of stimulus duration, $F(6, 372) = 69.01, p < .01$, and significant Age × Stimulus Duration, $F(6, 372) = 3.55$, $p < .01$, Intelligence × Stimulus Duration, $F(12, 372) = 3.31, p < .01$, and Age × Intelligence × Stimulus Duration, $F(12, 372) = 3.32, p < .01$, interactions. The mean proportion of yes responses at each stimulus duration for participants divided into the two age groups and three IQ groups are shown by the unconnected squares in Figures 1 and 2. The lines on these figures represent the fit of the MCG model.

In all groups, the highest proportion of yes responses occurred after the standard. It can be seen that the generalization gradient was flatter in the older participants than the younger ones and that the gradient flattened with decreasing IQ. Furthermore, generalization gradients around the standard were asymmetrical, with more responses occurring at

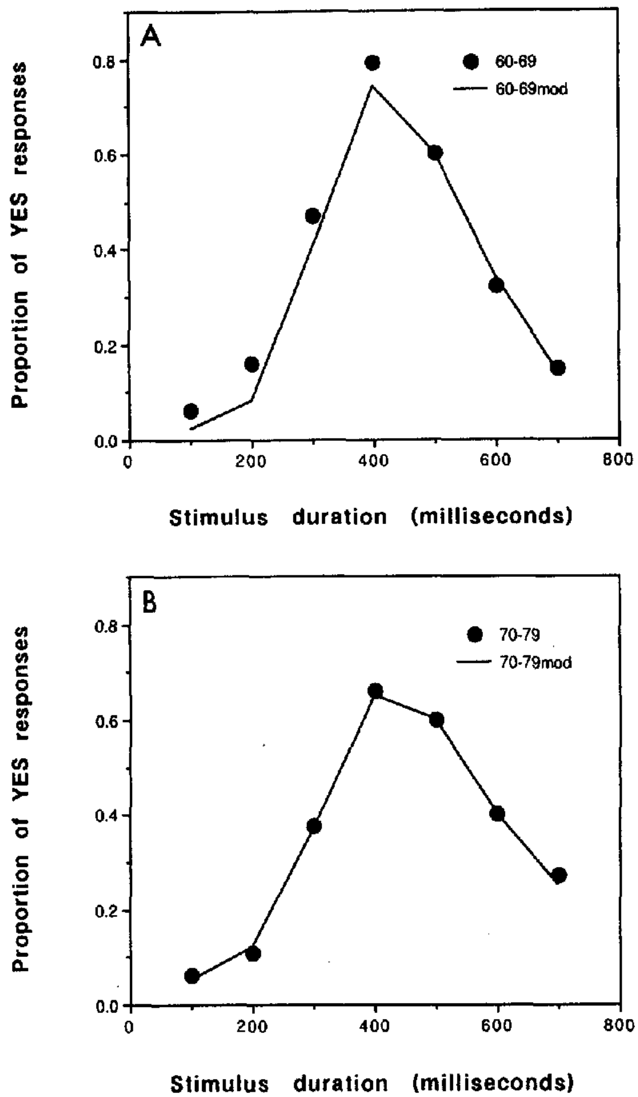


*Figure 1.* Mean proportion of yes responses (i.e., identifications of a presented duration as the 400-ms standard) as a function of stimulus duration for the 60- to 69-year-old (A) and 70- to 79-year-old (B) groups on the temporal generalization task (Experiment 1). The circles represent the obtained data, and the lines (e.g., 60–69mod) connect points derived from the computer model.

the 500-ms stimulus than at the 300-ms one, more at 600 ms than at 200 ms and more at 700 ms than at 100 ms. Wilcoxon matched-pairs signed-ranks tests were performed on the proportion of yes responses for individual participants at the three pairs of stimuli: 300 and 500 ms, 200 and 600 ms, and 100 and 700 ms. All comparisons between 100 and 700 ms and between 200 and 600 ms were significant at least at the .05 level. Comparisons between 300 and 500 ms were significant at the .01 level for the mid-IQ and older age groups, but they failed to reach significance for the younger participants and for the high- and low-IQ groups.

In the *Discussion* section, we use the MCG model to provide measures of variability of the memory of the standard duration in different groups, but this comparison mixes both between- and within-participants variability. To provide some index of response dispersion from individuals, as well as comparisons that do not depend on the specific assumptions of the MCG model, we calculated the proportion of total yes responses that occurred to the target duration and the two durations immediately adjacent (i.e., 300, 400, and 500 ms). Obviously, this measure would approach 1.0 if all yes responses were clustered closely around the standard value and would be lower if individual temporal generalization gradients were flatter. Data from 4 participants were lost because of a computer error, but with the remaining 76 we obtained a significant difference between the two age conditions (with the 60- to 69-year-olds showing more peaked distributions, $t(74) = 2.22$, $p < .05$, and significant differences among the three IQ groups ($ps < .05$). The proportion of yes responses to the three central stimuli were, on average, ordered in terms of IQ group: The highest IQ was most peaked, the lowest IQ group was the least peaked, and the mid-IQ group fell in between.

We conducted various correlation and regression analyses with the entire participant sample using the hit score (i.e., the proportion of yes responses made to the standard 400-ms stimulus), chronological age, and IQ in the form of the test score. Hits and age were significantly negatively correlated ($r = -.297$, $p = .007$), whereas IQ score was significantly positively related to hits ($r = .258$, $p = .02$), and age and IQ were negatively but nonsignificantly correlated ($r = -.174$, $p = .12$). A partial correlation analysis showed that age and hits remained significantly negatively correlated ($r = -.265$, $p = .018$) when IQ was partialed out but that the positive relation between hits and IQ just failed to reach significance ($r = .219$, $p = .052$) when age was partialed out. The measure of dispersion of responses around 400 ms, discussed earlier, was significantly negatively correlated with age ($r = -.26$, $p < .05$) and positively related to the test score ($r = .53$, $p < .05$). These correlations confirmed the result from the ANOVA that dispersion tended to be greater in older and lower IQ participants. Both the negative age and positive IQ correlation remained significant (values $= -.25$ and $.52$, respectively) when the other variable (age or IQ) was partialed out.

### *Discussion*

The results of this experiment parallel those of Wearden (1992) that undergraduates produced the most responses at the standard duration and that their temporal generalization gradient was asymmetrical with more responses occurring at durations longer than the standard than those shorter by the same amount. In contrast to Wearden's (1992) results, however, the asymmetry between the older individual participants' responses at 300 and 500 ms was not significant in all groups, suggesting some interparticipant variability in the data. These results also differ from Wearden's (1992) in that the current temporal generalization gradients were flatter than those Wearden obtained from the undergraduates (i.e., older participants responded less at the standard dura-

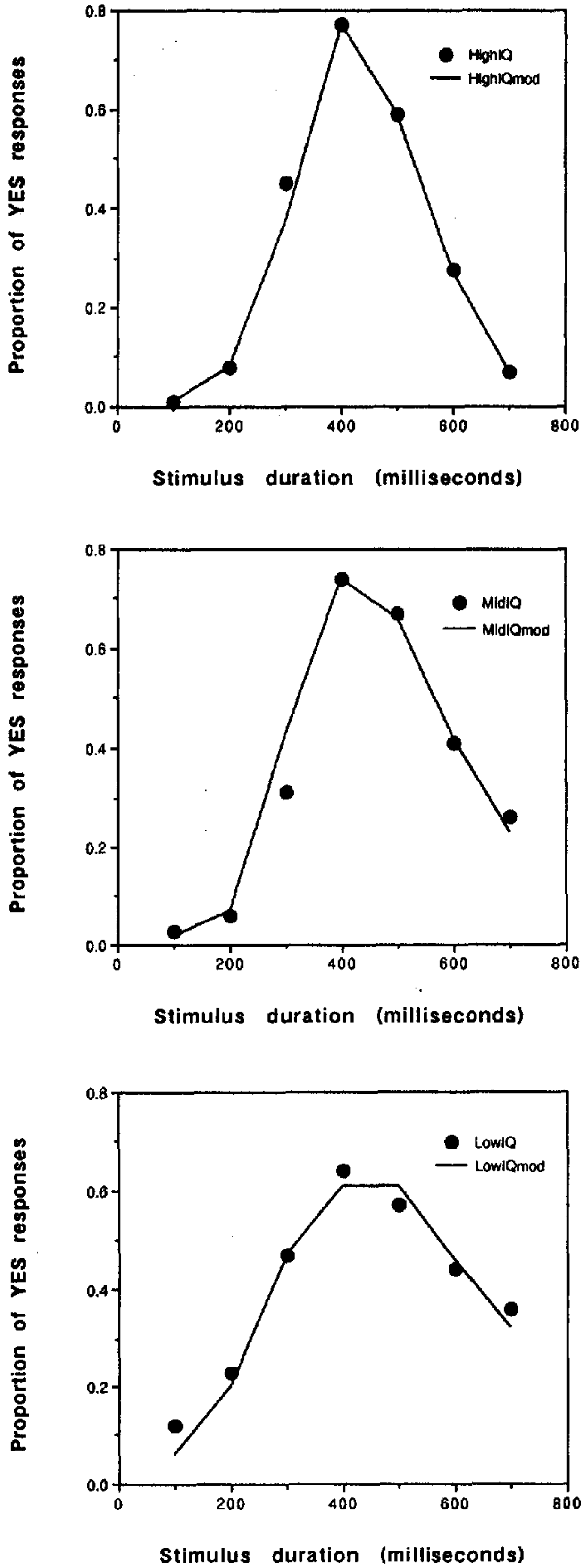


*Figure 2.* Mean proportion of yes responses from the temporal generalization task (Experiment 1) arranged as a function of participant IQ. mod = model.

tion and more often at nonstandard durations than did the students).

The MCG model described earlier was embodied in a Turbo Pascal computer program that generated 120 trials at each of the stimulus durations used. The parameters *c* (the coefficient of variation of the memory of the standard duration), *b* (the mean threshold value), and *x* (the standard deviation of the threshold) were varied over a wide range to determine the values that best fit each data set, using a criterion of smallest total absolute deviation. Table 1 shows parameter values for the best fits of the MCG model to the data from the five groups as well as the mean absolute deviation of the theoretically derived points from the data points. It also reproduces parameter values for student participants taken from Wearden (1992) in two conditions in which 50% of all presentations were the standard and 10% of all stimulus presentations were the standard.

It can be seen that with variation in either age or IQ, the best fit was obtained by varying *c*, the coefficient of variation of the memory representation of the standard, as participant characteristics changed, whereas the values for *b* and *x* remained fairly constant for all age and IQ test score groups. It seems that both age and IQ affected the degree of variability of the memory representation of the standard, with older and lower IQ participants having a higher coefficient of variation; in psychological terms, their memory representations of the standard duration (400 ms) were more variable. These relations between memory variance and age and IQ measures, derived from the MCG model, were thus consistent with the empirical measure of variability (i.e., proportion of total responses occurring to the 300-, 400-, and 500-ms stimuli) discussed in the previous section.

The mean absolute deviations between the output of the model and the empirical data were generally small, certainly no larger than those found by Wearden (1992), confirming the impression gained from inspection of Figures 1 and 2 that the theoretical model fit the data well. The goodness-of-fit values support the view that the basic assumptions of the model (e.g., the average accuracy of time and the type of decision process used) were applicable to older participants. Thus, overall, we found that decision processes and the mean accuracy of memory representation of the standard do not vary with either age or IQ but that the variability of the memory representation of the standard changes with both of these variables.

## Experiment 2: Temporal Bisection

Church and Deluty (1977) introduced a bisection technique that has been used in several animal experiments (e.g., Maricq, Roberts, & Church, 1981; Meck, 1983) and that has formed the basis of analog experiments on bisection carried out with humans by Wearden (1991b) and Allan and Gibbon (1991). Rats initially received discrimination training in which different responses (e.g., leverpresses on the right- and left-hand lever of a Skinner box) were reinforced after stimuli of different durations (e.g., a left response after a short stimulus 2 s long, a right response after a long stimulus 8 s long). This discrimination is not difficult for rats and was

Table 1
*Parameter Values From Fits of the Modified Church and Gibbon Model to Data From Experiment 1*

| Group | Parameter | | | |
|---|---|---|---|---|
| | *c* | *b* | *x* | MAD |
| 60–69 years | .26 | .28 | .13 | .04 |
| 70–79 years | .37 | .30 | .15 | .01 |
| Low IQ | .50 | .33 | .15 | .03 |
| Mid-IQ | .30 | .31 | .15 | .03 |
| High IQ | .24 | .26 | .12 | .02 |
| Students (50%) | .16 | .25 | .14 | .02 |
| Students (10%) | .20 | .25 | .10 | .05 |

*Note.* *c* is coefficient of variation of memory of the standard 400-ms duration, *b* is the threshold mean, *x* is the threshold standard deviation, and MAD is the mean absolute deviation in the proportion of yes responses between the obtained data and those obtained from the model. The two student groups were participants in Wearden's (1992, Experiment 2) research.

mastered to a high degree of accuracy after a few sessions of training. When discrimination of the standard short and long durations had been established, rats received a range of durations (e.g., ranging from 2 to 8 s in linear or logarithmic steps), and their response to each was noted in a test session conducted without reinforcement. If the response initially reinforced after the long standard stimulus (8 s in our example) is defined as the long response, the obtained data invariably show an increasing, and usually monotonic, tendency to emit this long response as stimulus duration increases. This psychophysical function can be analyzed in various ways, but one measure that has attracted considerable interest is the location of the "bisection point," which is the stimulus duration that gives rise to 50% long responses. Experiments with animals almost invariably show that the bisection point is located at the geometric mean of the short and long standards (where the geometric mean here is the square root of their product and 4 s in our example) rather than at the "linear" midpoint of the range spanned by the short and long standards, the arithmetic mean (5 s in our example).

Wearden (1991b) developed the basic technique of Church and DeLuty (1977) to produce a bisection method usable with humans. In an illustrative condition, participants initially received presentations of two stimulus durations (200 and 800 ms), with the shorter identified as a short standard and the longer as a long standard. Participants then received a range of durations (e.g., 200–800 ms in 100-ms steps) in which each stimulus had to be classified in terms of its perceived similarity to the short or long standards. This method quickly produced extremely orderly data, with almost all participants showing a monotonic change in the proportion of long classifications produced (going from near zero at 200 ms to near 100% at 800 ms) as stimulus duration increased. The bisection points obtained by Wearden (1991b) from two conditions with a 200- and 800-ms short and long standard pair were 0.43 and 0.44 s, values that were between the geometric mean of the standard durations (0.4 s) and its arithmetic mean (0.5 s). In bisection conditions with 100- and 900-ms pairs, however, the bisection point obtained was clearly much closer to the arithmetic mean (0.5 s) than to the geometric mean (0.3 s). Wearden (1991b) developed a theoretical model of bisection that fit the data reasonably well (discussed later). In a bisection experiment with human participants using short and long pairs with much smaller ratios than those used by Wearden (1991b), with the maximum being 2 to 1, Allan and Gibbon (1991) obtained evidence for a geometric mean bisection from humans. For a more complete discussion of the different experiments, along with a review and analysis of various theoretical models, see Wearden and Ferrara (1995, 1996).

In this experiment we used a bisection technique copied directly from Wearden (1991b), in which the standard short duration was 200 ms and the standard long duration was 800 ms. The principal difference between this experiment and that of Wearden (1991b), as well as some conditions from Wearden and Ferrara (1995), was that participants classified five series of durations ranging from 200 to 800 ms, not 10, as in the earlier experiments.

## Method

*Participants.* The same volunteers who served in the temporal generalization experiment participated in this experiment.

*Apparatus.* The same apparatus that was used in Experiment 1 was used in this study.

*Procedure.* There were two standard stimulus durations: The short standard was 200 ms and the long standard was 800 ms. The nonstandard durations were 300, 400, 500, 600, and 700 ms. On each individual experimental trial, the participant was presented with one of the seven stimuli and asked to make a decision about it. There were five series of experimental trials, with each series consisting of one presentation of each of the seven stimuli. The order of presentation of stimuli within the series was determined by the computer picking at random one of the stimuli, presenting it, eliminating it from the list of stimuli, and then picking the next stimulus at random from the remaining list and so on until all the stimuli had been used. Participants received a display instructing them to start the experiment by pressing the spacebar. There then followed five alternating presentations, separated by a delay of 3 s, of the short and long standard durations accompanied by displays stating "This is the short standard duration" and "This is the long standard duration," respectively. A new display then instructed participants to press the spacebar for the next trial. After a delay chosen at random from a uniform distribution between 1 and 3 s, participants received the first experimental trial of the series, that is, the presentation of one of the seven stimuli immediately followed by the display "Was that more like the standard SHORT or LONG? Press SHORT (S) or LONG (L) key." After the participant had responded, the display "Press the spacebar for the next trial" was repeated, the next stimulus was presented in the same manner, and so on until the end of the series. Because there was no "correct" answer on this task, no feedback was given. Each series except the last was followed by two presentations of the short and long standards arranged as at the beginning of the experiment.

## Results

Because all participants understood and complied with instructions, the data from all 90 were included in the analysis. The total sample thus comprised 41 participants aged 60–69 years (mean age = 64.81 years, $SD$ = 2.86; mean test score = 28.66, $SD$ = 6.22) and 49 participants aged 70–79 years (mean age = 73.94 years, $SD$ = 2.79; mean test score = 26.63, $SD$ = 6.99). These 90 participants fell into three IQ groups: 31 with test scores less than or equal to 23 (mean age = 70.49 years, $SD$ = 5.41; mean test score = 19.87, $SD$ = 3.04), 28 with test scores from 24 to 31 (mean age = 70.00 years, $SD$ = 5.38; mean test score = 27.89, $SD$ = 2.13), and 31 with test scores of 32+ (mean age = 68.87 years, $SD$ = 5.35; mean test score = 34.61, $SD$ = 2.11). For each participant, the number of long responses at each stimulus duration was calculated and plotted against stimulus duration. For all but 8 participants, the number of responses at each stimulus duration increased monotonically with increases in stimulus duration. An ANOVA showed a main effect of stimulus duration, $F(6, 504) = 804.8$, $p < .01$), but no effects of age, intelligence, or significant interactions.

Figure 3 shows the proportion of long responses plotted against stimulus duration at each of the stimulus durations for the two age groups (A) and the three IQ groups (B). It is obvious that the functions produced by each of the groups were nearly perfectly superimposed on each other, consistent with the absence of any statistical effect of age or IQ. Next, the bisection point was determined by eye from plots for individual participants (following Raslear, 1983; see also Wearden & Ferrara, 1995, for a discussion of different methods of determining the bisection point). This was accomplished easily for all but 3 participants, for whom the bisection point was not clear. The bisection points (in milliseconds) for the two age and IQ groups were as follows: 60–69 ($n$ = 41), $M$ = 435.49, $SD$ = 51.83; 70–79 ($n$ = 46), $M$ = 444.02, $SD$ = 55.14; low IQ ($n$ = 29), $M$ = 434.69, $SD$ = 52.86; mid-IQ ($n$ = 27), $M$ = 445.0, $SD$ = 55.46; and high IQ ($n$ = 31), $M$ = 440.65, $SD$ = 53.63. For the pooled data from 87 participants for whom the bisection point could be determined, the mean bisection point was 440.00 ms ($SD$ = 53.47 ms). An ANOVA carried out on the bisection points showed no effect of either age or intelligence, and no type of correlation analysis (e.g., simple, partial, or multiple regression) revealed any significant correlation between the bisection point value and either age or IQ.

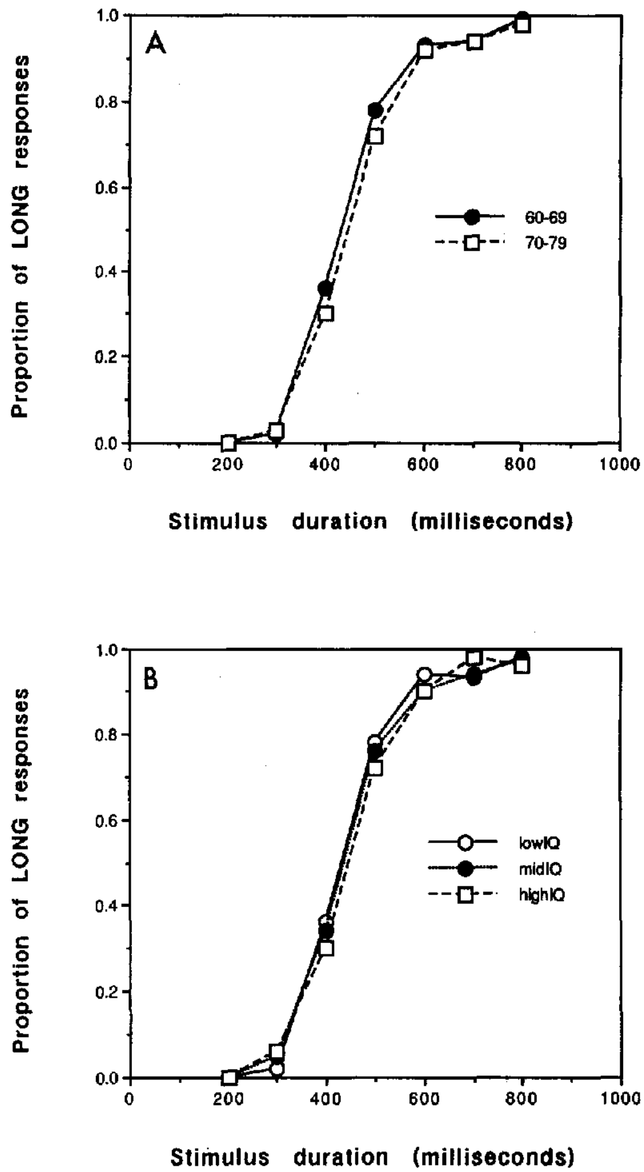


*Figure 3.* Mean proportion of long responses as a function of stimulus duration on the temporal bisection task (Experiment 2). A: Data arranged according to age. B: Data arranged according to IQ.

## *Discussion*

The data produced by the elderly participants on this bisection task were strikingly orderly, with no systematic variation in mean bisection points between groups. The mean bisection point of 440 ms derived from our data fell somewhere between the arithmetic and geometric means of the two standard stimuli, as was the case for data from students in the earlier studies.

Wearden (1991b), in two experiments, obtained data from 30 student participants on a bisection task that was virtually identical to that described earlier. To compare results with those of the current study, the proportions of long responses at each stimulus duration were averaged over the student group, and data from all the older participants in our study also were averaged together (a procedure justified by the finding that there were no effects of age or IQ on performance). Figure 4A shows the performance of these two groups on the bisection task.

Inspection of the data points in Figure 4 shows that the performance of the older participants in the current study and students from Wearden (1991b) was highly similar. However, there was a slight tendency for the elderly participants to produce fewer long responses to the longest stimulus durations presented. Given the similarity of the data, one would expect that theoretical models applied to data from the different groups would produce similar conclusions, and this was verified by applying the model of bisection developed by Wearden (1991b). This model fol-

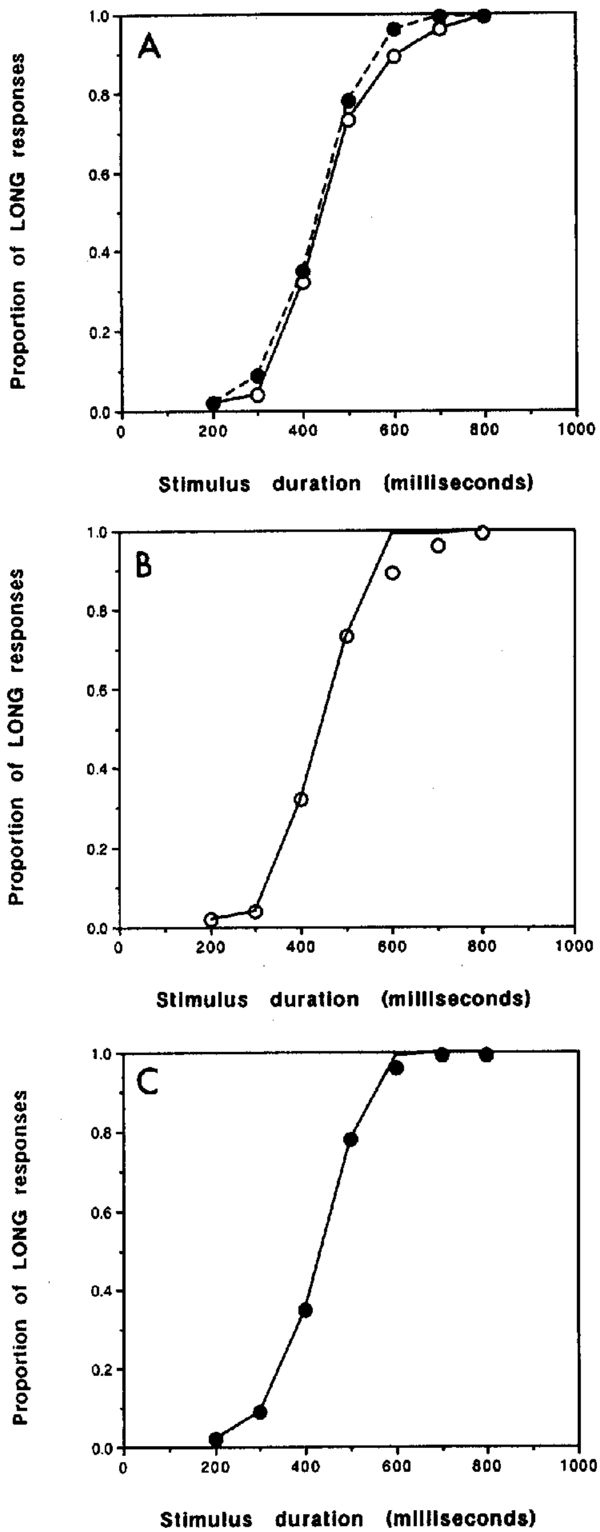


*Figure 4.* Comparisons between data obtained from older participants in Experiment 2 and students used by Wearden (1991b). A: Mean proportion of long responses for both types of participants (older: open circles; students: filled circles) as a function of stimulus duration. B: Data points from older participants (open circles) and the line corresponding to the predictions of the theoretical model. C: Data points from students (filled circles) and predictions of the theoretical model.

lows the basic principles of scalar timing theory and assumes that participants have representations of the short and long standard durations that are, on average, accurate but that vary from trial to trial around this accurate mean, being picked randomly from a distribution with a constant coefficient of variation, *c*. Thus, on each trial, the operative representations of the short and long standards are $s^*$ and $l^*$. The just-presented stimulus duration, *t*, is assumed to be timed accurately, and judgments of similarity of *t* to the short and long standards are governed by the absolute differences between *t* and $s^*$ and between *t* and $l^*$. If these absolute differences are much different (i.e., *t* is "obviously" closer to $s^*$ than it is to $l^*$), the response is governed by the smaller difference; if *t* is "ambiguous" (i.e., the difference between the absolute differences is less than some threshold, *x*), the model responds long. For a more complete account of this treatment of bisection, see Wearden (1991b, pp. 70–76).

The model was embodied in a Turbo Pascal program and produced 200 classifications of each of the stimulus durations presented. The two parameter values (*c*, the coefficient of variation of the memory of the short and long standards, and *x*, the ambiguity threshold) were varied over a wide range, and the best-fitting values in terms of mean absolute deviation were found.

Figure 4B shows the data points and line derived from the model for the older participant group, and Figure 4C shows the data points and theoretical values for the student participants. It is clear that, at least by usual psychological standards, the model fit the data well (i.e., the mean absolute deviation of the proportion of long responses for the older group was .03 and .02 for the younger participants), although it did tend to overestimate the proportion of long responses produced by both groups at the longest stimulus durations. Parameter values for the two fits were highly similar, with $c = .26$ for both groups and $x = .1$ (older) and .12 (younger). Thus, in the case of bisection, the model suggests that representations of the standard durations had the same variability in the young and older participants. The application of the Wearden (1991b) model illustrates that the basic principles of scalar timing (e.g., mean accuracy and scalar property of variance) can fit data from both younger and older participants.

The lack of effect of age or IQ on bisection performance, and the similarity of bisection performance of our older participants and students (who were in their late teens or early 20s), is in striking contrast to the data from the temporal generalization experiment, presented earlier. The reasons for this are unclear, but one possibility is that the bisection experiment, by repeatedly presenting "refresher" examples of the short and long standard durations, enables the older participants to develop less variable representations of the standard durations than they are able to do in temporal generalization. At the least, the results of the bisection experiment suggest that the performance of older people is not always more variable than that of younger participants on timing tasks.

Our Experiments 1 and 2 may be the first time that the same people have been tested on the temporal generalization and bisection tasks, and the "dissociation" between the effects of age and IQ on the two tasks poses problems for existing theory. As noted earlier, theoretical models compatible with scalar timing theory fit the data well in both temporal generalization and bisection, but variability in the proposed underlying memory of "critical" durations is subject to an age and IQ effect in temporal generalization, but not apparently in bisection. This may result from procedural factors, as discussed earlier, but it also may have more profound theoretical implications. Suppose that temporal memories do become more variable with increasing age and decreasing IQ, as the temporal generalization data suggest. The problem then becomes that of finding an appropriate bisection model that can incorporate these age- and IQ-related differences yet still produce performance that does not differ from that of students, whom temporal generalization experiments suggest have less variable temporal memories. Simulations of bisection performance using Wearden's (1991b) model with underlying coefficients of variation derived from temporal generalization produce much flatter bisection functions for the older people than for students, in contrast to our data. Other models of bisection (such as those given by Allan & Gibbon, 1991, and Wearden & Ferrara, 1995) also will encounter difficulties because, in all of them, the underlying variability of the memory representation of the standard short and long durations (Allan & Gibbon, 1991) or the mean of all the stimulus durations presented in the bisection task (Wearden & Ferrara, 1995) plays a role in determining the shape of the bisection function.

## Experiment 3: Threshold Determination

The third procedure used was an assessment of differential thresholds for tone duration. On all trials, one of two tone stimuli presented had a duration of 600 ms, and the other had a different, variable value. Participants simply had to judge whether the two stimuli had the same duration. Experimental trials were arranged in four conditions. For two conditions, the constant 600-ms stimulus came first, and the variable stimulus (longer or shorter than 600 ms) came second; for two others, the variable stimulus came first and the 600-ms stimulus second. We did this to counterbalance the possibility of time–order errors, which are effects on the judgment of two consecutive stimuli depending on their order of presentation (see Hellstrom, 1985, for a review of time–order errors in general and Wearden & Ferrara, 1993, for a review of time–order errors relating to judgments of duration).

### *Method*

*Participants and apparatus.* The participants and apparatus were the same as in the preceding experiments.

*Procedure.* The order of presentation of the four experimental conditions was randomized between participants. On each experimental trial, participants were presented with two tones and had to decide whether they had the same duration. On any trial, a response that the two tones were not of the same duration caused the next trial to be drawn from the same condition, whereas a response that the two tones were of the same duration caused the next trial to be drawn from the next condition, until all four conditions had been presented. In Condition 1, the second tone of the pair was always 600 ms, whereas the first tone started at 320 ms on the first trial in the condition and then incremented by 20 ms from one trial to the next if the participant judged that the tone durations were different. For Condition 2, the first tone of the pair was always 600 ms long, whereas the second tone started at 320 ms on the first trial in the condition and incremented by 20 ms from one trial to the next after "different" responses. For Condition 3, the second tone was always 600 ms, whereas the first tone started at 880 ms on the first trial and decreased in 20-ms steps from one trial to the next if the participants responded "different." Finally, for Condition 4, the first tone was always 600 ms long, whereas the second tone started at 880 ms on the first trial in the condition and decreased in 20-ms steps from one trial to the next after "different" responses.

Participants started the experiment by pressing the spacebar. They then received the first pair of stimuli separated by a delay chosen at random from a uniform distribution ranging from 300 to 500 ms. The tones were followed immediately by the display "Did the two tones have the same length? Press YES (Y) or NO (N) key." After the participant had responded, there was a delay of 1 s before the next pair of stimuli were presented. The experiment continued as described earlier until the participant had responded yes (i.e., judged that the stimulus durations were the same) four times.

### *Results*

Because all participants understood and complied with instructions, all were included in the analysis that follows. Each participant produced four measures, one for each condition, of the difference (in milliseconds) between the two tones on the trial on which the participant said both tones were of the same duration (hereafter called the *threshold*). An ANOVA revealed a main effect of condition, $F(3, 252) = 18.02$, $p < .01$, and of intelligence across all conditions, $F(2, 84) = 3.49$, $p < .05$, with the low-IQ participants showing a higher threshold than the mid- and high-IQ participants. There was no main effect of age, but there was an Age × Condition interaction, $F(3, 252) = 3.67$, $p < .05$. Across all participants, the mean thresholds in Conditions 1–4 were, respectively, 233.99, 220.16, 261.48, and 246.91 ms. Thus, participants found the pairs containing the longer and middle tones harder to tell apart than the pairs containing the shorter and middle tones. Figure 5 shows the mean thresholds at the four conditions for the two age groups (A) and three IQ groups (B).

It can be seen that although age had no overall effect across conditions, in Condition 2 (which had the lowest mean threshold overall across all participants), older participants produced higher thresholds than younger ones. Similarly, the difference between the low-IQ and other participants, which was significant across all conditions taken together, was most marked in Condition 2.

We next correlated the four threshold measures together and with participant age and IQ. All four threshold measures correlated positively together, and most intercorrelations were significant at the .05 level. IQ score was negatively correlated with all four threshold measures, but the correlation was never statistically significant, and correlations

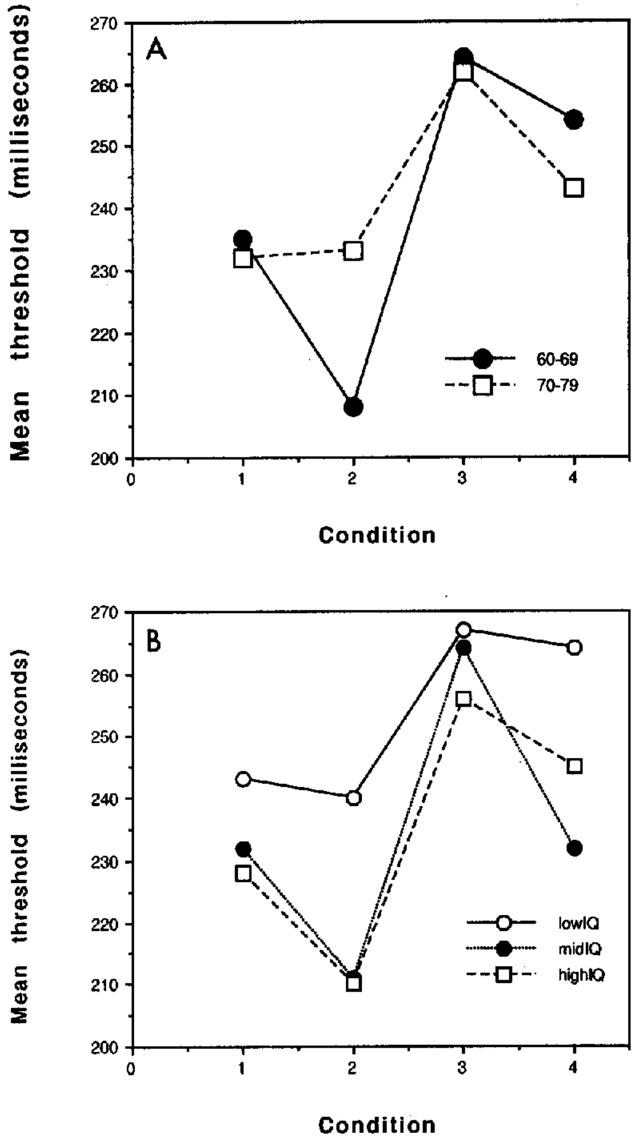


*Figure 5.* Mean differential thresholds obtained from Experiment 3 in the four determination conditions. Data are arranged in terms of age group (A) and IQ group (B).

between age and threshold measures were small, of variable sign, and never significant. No partial correlation or multiple regression analyses using age and IQ to predict threshold values produced significant results.

### *Discussion*

Although the experimental procedure used was simple, the pattern of results obtained was, at least at first sight, complex. Both age and IQ were found to have effects on some threshold measurement conditions but not on others. Furthermore, although all conditions involved comparisons of a 600-ms tone with another shorter or longer time, they produced consistently different threshold values. The key to the complex effects observed might have been floor effects. In general, the between-groups effects of age and IQ were most marked (or even only present) in the second threshold determination condition, which produced the smallest mean threshold. In the current experiment it seems reasonable to regard the threshold values obtained as a measure of task difficulty, so Condition 2 also was the easiest (i.e., the condition in which the perceived threshold of difference was smallest).

Verbal reports from the participants suggested that they found the threshold determination task difficult, a suggestion reinforced by the impression of the experimenters themselves when developing the task. Perhaps between-groups threshold differences resulting from age or IQ differences would have been more clearly manifested in an easier task or when many more trials were given on the task used. Considerations of available time and resources precluded this in the current study, but our results suggest that a more elaborate repetition of the task in another study might well provide clearer effects of age and intelligence.

## Experiment 4: Interval Production

The final task we used was one of *interval production.* This requires participants to produce responses with some specified temporal characteristic (e.g., by holding down a response key for a specified time period or pressing a key twice with some specified temporal gap between presses). Feedback on performance accuracy may or may not be given after responses. The temporal characteristics of the population of times produced are generally assumed to be some reasonably direct reflection of the characteristics of underlying timing processes. For example, Wearden and McShane (1988) required young adults to repeatedly produce intervals ranging from 500 to 1,300 ms by spacing two buttonpresses, with accurate feedback given after each response. They found that the population of times produced conformed almost exactly to the theoretical requirements of scalar timing theory in that the (a) mean time produced matched almost perfectly the required time and (b) coefficient of variation of times produced (standard deviation/mean) remained constant (at about 0.13) as the interval to be produced varied, a manifestation of the scalar property of time.

Available time and resources did not permit a replication with our older participants of Wearden and McShane's (1988) study, so we concentrated instead on a briefer production procedure involving the learning, maintenance, and retention of production of a single time interval: 1 s. Participants received 45 production trials: an initial practice period of 5 trials, 10 further trials without feedback, 20 trials with accurate feedback, and 10 additional trials in which feedback was withdrawn. This simple design enabled us to examine stable performance with and without feedback, and any possible changes in production that occur when feedback is initially introduced, and later when it is withdrawn.

## Method

*Participants and apparatus.* The participants and apparatus were the same as those used in the preceding experiments.

*Procedure.* Participants were required to produce a 1-s interval by pressing on the spacebar on the computer keyboard once, waiting 1 s, and then pressing again; they received 45 trials overall. No feedback was given for the first 14 trials; after the next 20 trials, participants received accurate feedback, and feedback was discontinued after the last 11 trials. In other words, participants received feedback after responding on Trials 15–34 inclusive but not on any other trials. They could use this feedback to modify their responses on Trials 16–35. There were thus three sequential experimental conditions: prefeedback, with feedback, and postfeedback. To allay participants' anxieties about performing the task, the first 5 trials were described as practice trials and were accompanied by slightly different displays on the screen, but otherwise they were no different from the other trials. Participants received instructions from the experimenter describing the task and then received the display "The first few trials are for practice. Press the spacebar twice, leaving one second in between the presses. Start when you are ready." After participants made their two responses, they received the display "Please wait for the tone." This was accompanied by a delay that was chosen at random from a uniform distribution between 2 and 3 s, followed by the presentation of a 500-Hz tone with a duration chosen at random from a uniform distribution between 100 and 200 ms. The delay and tone were introduced into the procedure to space trials because pilot tests had shown that some participants tended to press the spacebar repeatedly. After the tone, for Trials 2–5, participants again received the display "Press the spacebar twice, leaving one second between the presses. Start when you are ready." This procedure was repeated until the start of the 6th trial, when participants received the display "Practice is over. The experiment starts now. Press when you are ready." Participants then responded and received the delay and the tone as before. On Trials 7–15, participants received the simple instruction "Press when you are ready" at the start of the trial. The experiment continued as described until participants had responded on the 15th trial, at which point they immediately received the display "Time between presses . . . [followed by the time in seconds to three decimal places between the two presses they had just made]. Please wait for the tone." The feedback and response spacing message remained on the screen for a duration chosen at random from 2 to 3 s and was followed by the tone as before. The 16th–34th trials were identical with the fifteenth. On Trials 35–44, participants received the simple response spacing message after responding, and on the final trial a message to thank them for participating in the experiment. The computer recorded the time between the two presses for all trials.

## Results

The data from 5 participants were discarded from the analysis because they failed to understand or comply with instructions. This left a sample of 40 participants aged 60–69 years (mean age = 64.86 years, $SD = 2.86$; mean test score = 28.55, $SD = 6.23$) and 45 participants aged 70–79 years (mean age = 73.77 years, $SD = 2.72$; mean test score = 27.29, $SD = 6.28$). These 85 participants fell into three IQ groups: 27 with test scores of 23 or less (mean age = 69.74 years, $SD = 5.30$; mean test score = 20.33, $SD = 2.17$), 28 with test scores from 24 to 31 (mean age = 70.00 years, $SD = 5.38$; mean test score = 27.89, $SD = 2.13$), and 30 with test scores of 32+ (mean age = 69.03 years, $SD = 5.21$; mean test score = 34.67, $SD = 2.12$). For each participant, the 45 times produced were sorted into nine 5-trial blocks. Blocks 1, 2, and 3 contained data from the 15 prefeedback trials, including the 5 practice trials that formed Block 1. Blocks 4–7 contained data from the 20 trials on which participants could use feedback to modify their responses, and Blocks 8 and 9 data from the 10 postfeedback trials. For 26 participants, the data from one or more trials had to be discarded from the analysis for a variety of reasons, including failure to wait for the tone, holding down the spacebar, and failure to read feedback. Because the experimenter was present during the experiment, it was easy to note and identify unusable trials. In these cases, either the block had been treated as holding fewer data or the next consecutive piece of data had been added to the block. For Blocks 1, 3, 4, 7, 8, and 9, the mean time produced, mean absolute deviation from 1.0 s, standard deviation, and coefficient of variation (standard deviation/mean) of the five responses were calculated, and data from these five-trial blocks were grouped into six conditions to be reported: two from trials before feedback (BEF1 = first 5 trials; BEF2 = last 5 trials), two from trials with feedback (FEED1 = first 5 trials; FEED2 = last 5 trials), and two from trials after feedback was discontinued (AFT1 = first 5 trials; AFT2 = last 5 trials). The data for the means, mean absolute deviation, and coefficient of variation were then entered into three separate ANOVAs.

Figure 6 shows the mean times produced in the different conditions averaged across the two age groups (A) and three IQ groups (B). It can be seen that participants in all age and IQ groups produced overestimates of 1 s in the two prefeedback conditions, with younger participants producing larger overestimates than older ones and lower IQ participants producing larger overestimates than the other two IQ groups. The onset of feedback had a marked effect on responding in all groups, and, by the end of the feedback trials, all groups were producing intervals with means close to 1 s. The accuracy of the mean was maintained well in all participant groups after the offset of feedback. An ANOVA indicated that there was a main effect of condition, $F(5, 395) = 34.2$, $p < .01$, over all participants. Paired comparisons produced by the Newman–Keuls test showed that production in bins BEF1 and BEF2 (i.e., the condition without feedback) differed significantly ($p < .01$) from that in other conditions (i.e., with and after feedback). There was a main effect of intelligence, $F(2, 79) = 7.60$, $p < .01$, over all conditions and an Intelligence × Condition interaction, $F(10, 395) = 6.25$, $p < .01$, with intelligence having a significant effect on BEF1, $F(2, 79) = 5.62$, BEF2, $F(2, 79) = 8.22$, and FEED1, $F(2, 79) = 5.38$, $ps < .01$, but not on later conditions. Therefore, it would seem that, compared with high- and mid-IQ participants, participants with low IQ scores were less able to produce an accurate estimate of a second before they were given any indication of its true length. They then adjusted their productions more in the light of feedback than did the other two groups, but they were still less accurate over the first five feedback trials than mid-IQ participants, who in turn were less accurate than

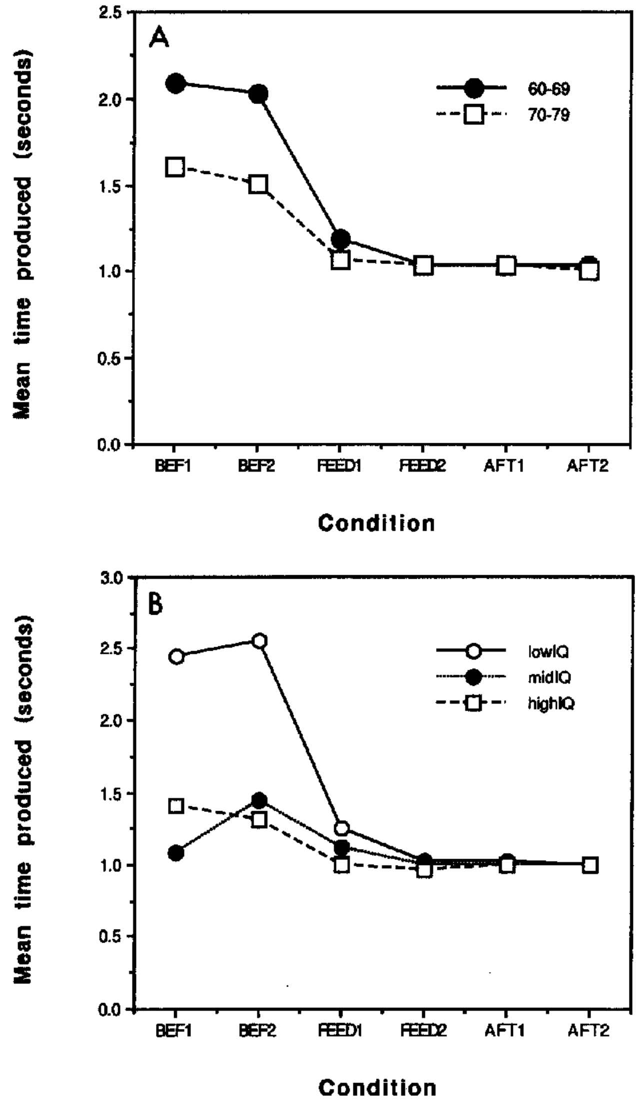


*Figure 6.* Mean times produced in the interval production task (Experiment 4) in the six response conditions. Data are arranged in terms of age group (A) and IQ group (B). BEF1 and BEF2 = before feedback; FEED1 and FEED2 = with feedback; AFT1 and AFT2 = after feedback.

high-IQ participants. By the end of the feedback trials, however, differences between the three groups had virtually disappeared. There was a main effect of age, $F(1, 79) = 4.14$, $p < .05$, over all conditions, and an Age × Condition interaction, $F(5, 395) = 3.33$, $p < .01$, with age having an effect that just failed to reach significance ($ps = .059$ and $.051$) on responding before feedback (BEF1 and BEF2). Finally, there was an Age × Intelligence × Condition interaction, $F(10, 395) = 2.72$, $p < .01$, reflecting that it was the young, low-IQ participants whose initial estimates of 1 s were the least accurate, with the older, low-IQ participants showing less of a difference from the other two older IQ groups.

The average mean absolute deviations (i.e., the unsigned difference between the time produced on a trial and 1 s, averaged across participants in the two age and three IQ groups), mirrored those for data on the mean time produced, with younger and lower IQ participants producing larger deviations from 1.0 on the prefeedback trials than did mid- and high-IQ participants. Hence, these data are not shown separately to save space. In contrast to the data on the mean time produced, however, over conditions with and after feedback, the mean absolute deviations of intervals produced by the low-IQ group continued to be higher than those for the other two IQ groups. This shows that although the means of the intervals produced in these conditions were accurate, there was a greater spread of durations around the mean.

Figure 7 shows the averaged coefficients of variation (standard deviation divided by the mean of times produced by individuals) in each of the six conditions for the two age groups (A) and three IQ groups (B). An ANOVA showed main effects of condition, $F(5, 395) = 18.39$, $p < .01$, and intelligence, $F(2, 79) = 12.25$, $p < .01$, but no effect of age or interactions between intelligence or age and condition. This can be seen in Figure 7, in which the coefficient of variation clearly differs from one condition to another and the plots for the two age groups are nearly superimposed (top), whereas the plots for the three IQ groups follow nearly parallel paths (bottom). In all groups, the coefficient of variation was higher in BEF1 than BEF2 (indicating some reduction in variability even before feedback was given), higher in FEED1 than FEED2, and higher in AFT1 than AFT2, that is, at the start of each experimental condition (i.e., before feedback, with feedback, and after feedback), participants' productions were more variable than at the end of the condition. The increase in variability in FEED1 probably reflected participants attempting to adjust their productions in light of feedback and initially overcompensating in one direction or the other, thus oscillating around the 1-s value. The offset of feedback also affected variability without having an effect on the mean accuracy of the intervals produced.

We conducted regression analyses on the various performance measures from the different bins of the production task and related these to age and IQ. Because there were many intercorrelations, many of which were significant, we do not present all of them in detail; in the summary that follows, correlation coefficients are described as significant if they reached the .05 level and as nonsignificant otherwise, but the significant correlations were usually associated with much smaller probability values than .05. The mean times produced from the different conditions (BEF1 to AFT2) were all significantly correlated; thus, performance on production before feedback was correlated with performance when feedback was given. All intercorrelations between the mean absolute deviations were significant, except for those from FEED1 and FEED2. On the other hand, coefficients of variation from the different conditions, although almost always intercorrelated positively, reached significance only in about half the cases.

Chronological age correlated negatively, but never significantly, with mean time produced from all six conditions. IQ

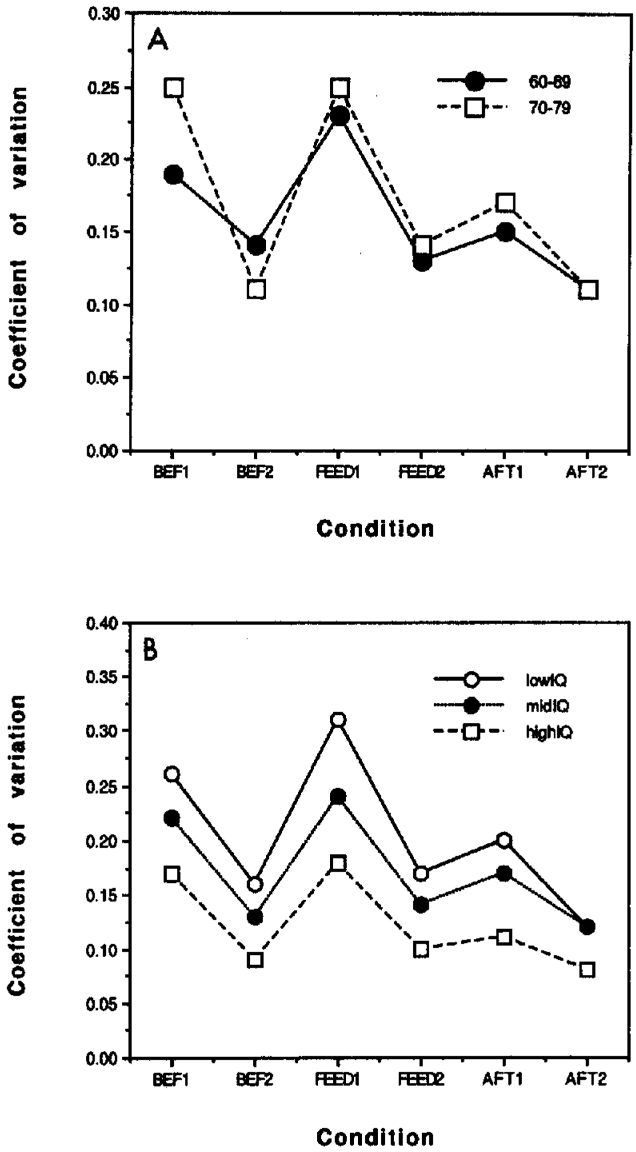


*Figure 7.* Coefficient of variation of times produced (standard deviation/mean) from Experiment 4 in the different response conditions (BEF1 to AFT2). Data are arranged in terms of age group (A) and IQ group (B). BEF1 and BEF2 = before feedback; FEED1 and FEED2 = with feedback; AFT1 and AFT2 = after feedback.

also correlated negatively with the mean time produced, and correlations were significant from all bins except FEED2 (the last trials with feedback, $p = .51$) and AFT2 (the last bin, after feedback was withdrawn). The mean absolute deviation scores similarly were always negatively, but never significantly, correlated with age, whereas IQ scores were negatively, and always significantly, correlated with the mean absolute deviation. Overall, therefore, considering the mean time produced and the mean absolute deviation as measures of "production accuracy," this variable was negatively related to both age and IQ but was usually only significantly correlated with IQ. Coefficients of variation of times produced from the various bins had correlations of variable sign with both age and IQ, and only two of these correlations (both with IQ) were significant. Thus, according to the correlation analysis, variability around the mean (as assessed by the coefficient of variation) was less well predicted by age or by IQ than was deviation from the target time of 1 s, which was well predicted by IQ scores.

### *Discussion*

Perhaps the most interesting finding of this experiment was that, when given feedback, older people were able to produce 1-s estimates that were extremely accurate, both in terms of mean accuracy and small coefficient of variation. In data from FEED2, the last five trials on which participants could use feedback, the mean coefficients of variation for each participant group were as follows: 60–69, .14; 70–79, .14; low IQ, .18; mid-IQ, .14; and high IQ, .10. Wearden and McShane (1988) found coefficients of variation in the range of .10–.17 from their student participants, with an average value of about .13; thus, both the 60- to 69-year-old and 70- to 79-year-old groups in the current study produced values similar to those produced by the students, with the highest IQ participants even producing values at the lower end of the range found by Wearden and McShane (1988). Because the coefficient of variation from production experiments is conventionally interpreted as a measure of timing sensitivity, these results show that older participants can approach, or even match, the timing sensitivity of much younger students (who also might have had higher IQs, although they were not measured) when feedback is given consistently. This result thus parallels that obtained in the bisection study (see Experiment 2) that the performance of older participants is not always more variable (or at least not much more variable) than that of much younger people.

The regression analysis suggested that the main individual difference effect on interval production was the effect of IQ on mean accuracy (mostly in nonfeedback conditions), in which lower IQ participants tended to produce longer intervals (which thus resulted in larger absolute deviations because most participants overestimated 1 s without feedback). Although chronological age was generally negatively related to the mean time produced, this correlation was rarely significant.

Previous work on interval production in older people (e.g., Feifel, 1957; Kline et al., 1980; Licht et al., 1985; McGrath & O'Hanlon, 1968; Surwillo, 1964) has used much longer intervals (between 16 s and 8 min) and has not used feedback. In general, previous studies have shown that older participants tend to produce shorter estimates than do younger ones, sometimes because older participants underestimate real time more than younger participants and sometimes because older participants overestimate real time less than younger participants. The current results likewise show that, before feedback, older participants produced shorter estimates of 1 s than did younger participants, although virtually all participants overestimated 1 s, with the younger, low-IQ participants overestimating the most dramatically.

## Principal-Components Analysis

Our four experiments produced a data set in which participants produced a range of measures of timing behaviors, as well as having known age and intelligence test scores. As mentioned earlier, data from a few participants were discarded from each of the experiments because of errors or failures to understand experimental instructions, and which participants were discarded differed from task to task. Overall, 70 of our participants completed all four tasks successfully, and we began by deriving measures of performance on the four tasks, and intercorrelating these over the 70 participants, with age and intelligence test score also included. We used the following measures: scores from the Culture Fair Intelligence Test; age (chronological age in years); hits (the proportion of correct identifications of the standard duration on the temporal generalization task); peak (the location of the peak of yes responses on the temporal generalization task); a measure of asymmetry around the peak on the temporal generalization task (the number of responses made to durations longer than the standard minus the number made to durations shorter than the standard); the proportion of total yes responses made to the 300-, 400-, and 500-ms stimuli on the temporal generalization task; the bisection point from the bisection task; the four threshold values from the four conditions of the stimulus discrimination experiment; the mean time produced on the six conditions of the production task; mean absolute deviations from the six conditions reported in the production task; and the coefficients of variation (standard deviation/mean) of times produced from the six conditions of the production task. Because the correlation matrix revealed that the 4 threshold values, the 6 mean time values, the 6 mean absolute deviation values, and the 6 coefficient of variation values showed large positive correlations within each type of measure, we used within-participants average values for each. This generated a total of 11 variables, which, with 70 participants, produced an adequate (but not large) participant-to-variable ratio, according to Hair et al. (1987).

We examined the scree plot (Gorsuch, 1983, pp. 165–169) from a principal-components analysis (using SPSS for Windows) to determine how many factors should be extracted, and this suggested that a three-factor solution was appropriate. The varimax-rotated orthogonal factors obtained are shown in Table 2. Together, they accounted for 60% of the data variance.

Inspection of the factor loadings suggested that Factor 1 was related mainly to the mean time produced in the production task (and, because producing durations longer than 1 s was the most common deviation in the before-feedback conditions, to the mean absolute deviation). The other largest loading on this factor was a negative loading with general ability (the Culture Fair Intelligence Test score), thus making it another version of the result shown in the bottom of Figure 7 that less able participants produced longer times on the production task. Factor 2 appeared to relate specifically to the location of the peak of yes responses on temporal generalization, with higher peak values leading to more asymmetry; a smaller proportion of yes responses to the 300-, 400-, and 500-ms stimuli; and fewer hits. This factor loaded slightly positively with age (so older participants were more likely to err in the direction of peaks above 400 ms) and negatively with general ability (suggesting that the more able participants had lower and therefore probably more accurate peak locations).

Table 2
*Factors and Factor Loadings Obtained From the Principal-Components Analysis*

| Variable | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Age | −.30 | .21 | −.58 |
| ASYM | .12 | .86 | −.06 |
| BISPT | −.29 | −.03 | −.01 |
| CF | −.54 | −.16 | .49 |
| Hits | −.11 | −.35 | .69 |
| MAD | .95 | .09 | −.01 |
| Mean | .95 | .03 | .01 |
| CV | .14 | .20 | .45 |
| THR | .19 | −.29 | −.45 |
| MID3 | −.25 | −.47 | .73 |
| Peak | .06 | .94 | −.11 |

*Note.* Age = chronological age in years; ASYM = measure of asymmetry around the peak on the temporal generalization task (i.e., the number of responses made to durations longer than the standard minus the number made to durations shorter than the standard); BISPT = bisection point from the bisection task; CF = Culture Fair Intelligence Test; Hits = proportion of correct identifications of the standard duration on the temporal generalization task; MAD = mean absolute deviations from the six conditions of the production task; Mean = mean time produced on the six conditions of the production task; CV = coefficient of variation (standard deviation/mean) of the times produced from the six conditions in the production task; THR = threshold values from the four conditions in the stimulus discrimination task; MID3 = proportion of total yes responses made to the 300-, 400-, and 500-ms stimuli on the temporal generalization task; Peak = location of the peak of yes responses on the temporal generalization task.

Factor 3 was perhaps the most psychologically interesting of the three, although it accounted for less of the variance in data than the other two. The main loadings here were on the proportion of yes responses to the 300-, 400-, and 500-ms stimuli and hits, and hits, with high positive loadings on these two measures of "accuracy" on temporal generalization and a negative loading on threshold. Taken together, these three loadings might suggest that the factor measured "stimulus timing sensitivity" (i.e., it was positively related to accurate temporal generalization performance and low threshold). It loaded positively on IQ and negatively on age, suggesting that sensitivity was higher in younger, and more able, participants, a common theme of several of our experiments. This simple interpretation is complicated by a substantial positive loading on the coefficient of variation from the production task, suggesting that this factor promoted more variable responding in production. This is difficult to interpret in light of the data shown in Figure 7, which show more variable responding in lower IQ participants. Taken together, the results from the principal-components analysis generally support the interpretation derived from the ANOVA that "poor" timing performance, in terms of mean inaccuracy or, more commonly, higher variability, tends to be associated with increasing age and decreasing IQ.

## General Discussion

Two frameworks can be used to summarize the effects we obtained in our experiments. In one framework, we used two ways of dividing up the older participants, age and IQ, and organized the effects obtained around these two variables. In the other framework, we compared our older participants (or subgroups of older people) with students used in previous similar experiments. Taking the former type of comparison first, in some cases (e.g., temporal generalization and threshold) stimulus timing variability tended to increase with increasing age and with decreasing IQ. This was not found in the bisection study, as noted earlier. In the interval production study, ANOVA results suggested that lower IQ was associated with more variable response distributions, whereas the correlational analysis suggested that the main effect was that of IQ on production accuracy in the no-feedback conditions (rather than variability), with lower IQ participants producing longer times. However, all three experiments using stimulus timing suggested that durations were, on average, accurately timed by all participants because theoretical models of temporal generalization and bisection that assumed mean accuracy fit the data well. The interval production experiment also suggested that older participants could use feedback to produce intervals accurately and with coefficients of variation that overlapped with, or were only slightly larger than, those produced by students.

Our older participants showed the first property of scalar timing: mean accuracy. There also was evidence from the bisection experiment (the only one in which a model involving variability in representations of two different times, that is, the short and long standards, was used) that their time representations embodied the scalar property of constant coefficient of variation. A model making this assumption fit the data well. In general, therefore, comparisons within the older group suggested that older, and lower IQ, participants may have more variable representations of stimulus durations than younger and higher IQ participants but that, on average, all participant groups conform to scalar timing requirements.

A comparison of the behavior of the older participants in our research and data from the students reported in various studies by Wearden (e.g., 1991a, 1991b, 1992; Wearden & McShane, 1988) also suggests that timing in our older participants, although sometimes more variable than in the student comparison group, did not have any qualitatively different properties. We note that the undergraduates are not, in any proper sense, a "control" for the performance of our older population; for one thing, the students' IQ scores were not known and the testing conditions also differed. Nevertheless, this comparison may be useful as some general indication of the extent of changes in timing behavior over the 40 or 50 years between youth and old age. Our results suggest that, overall, the changes are slight, at least on the tasks we used. During temporal generalization, older participants apparently had more variable representations of the standard duration than students. However, because the MCG model fit data from both sorts of participants well, it would seem that the threshold values and decision processes used by older and much younger people on temporal generalization are much the same. On the bisection task, differences between older participants and a student comparison group were even slighter, with small differences found in the obtained data and parameter values derived from a theoretical model. The other experiment in which some data from students were available (from Wearden & McShane, 1988) was the interval production task. Here, older participants were able to produce the target time with near-perfect average accuracy, and the coefficients of variations of times produced, which are usually considered to be a measure of timing sensitivity, were close to or within the range of values obtained from students.

Overall, our results suggest that older participants may be able to perform well in many conditions involving stimulus or response timing, particularly if conditions are favorable to their forming accurate temporal representations. Such "favorable" conditions might include the provision of consistent, prompt, and accurate feedback, as in our production experiment, or frequent repetition of distinctive standard stimuli, as in our bisection study. Less favorable tasks for older people might be those that involve only a few distinct presentations of standard durations (as in our temporal generalization task), or when the task requires difficult discriminations, as our threshold determination appeared to do. Even in these cases, longer exposure to the experimental conditions may help to reduce performance differences between older and younger participants (as in Lejeune & Pouthas, 1991).

Our findings suggesting broadly similar performance in the older and younger participants might provoke two contradictory reactions. One possible conclusion is that timing performance degenerates only slightly with age (and also varies, but only slightly, with general ability as assessed by IQ test scores). The high accuracy of older people's performance contrasted with their expressed misgivings that their performance would be poor. An alternative conclusion is that age-related timing effects really do occur (as anecdotes suggest) and that either (a) our procedures were not sensitive enough to detect them or (b) the changes were manifest in forms of timing other than the ones we used. Although this second position cannot be disproved because it could be raised regardless of the tasks we used, it may in practice be a reasonable one. We make no claim that our procedures explored anything other than a small part of the large domain of time psychology. Modern theory of duration judgment (see Hicks, 1992; Hicks, Miller, & Kinsbourne, 1976) distinguishes between prospective and retrospective timing. The latter occurs when unexpected questions are asked about the duration of some event after its termination and participants' performance is said to depend on the amount of cognitive activity, or the number of "contextual changes" (Block, 1982, 1992) that they experienced during the event. By contrast, all our tasks were prospective ones in which participants were alerted in advance that duration was important. Furthermore, they were prospective tasks in which chronometric counting was not used. Many phenomena in the psychology of time are not concerned with duration per se but with other aspects of time experience (Michon, 1990). Therefore, although our results suggest that the existing data and theory, obtained from studies with students, apply reasonably well to timing in older people in the particular situations we investigated, they can by no means preclude the discovery of more dramatic age-related differences in other sorts of work. Our research, however, provides a reasonable starting point because we used tasks that produced orderly data in most conditions and ones to which recently developed quantitative theories could be applied successfully.

Our data offer some, albeit indirect, support for suggestions mentioned in the introduction that CRT performance in older people might be affected by their having more variable time judgments than younger people. In three of our four experiments (the exception was Experiment 2), the data suggested that increasing age (and, more clearly, decreasing general ability as measured by IQ) tended to increase timing variability on tasks involving judgments of single durations (Experiment 1), two durations (Experiment 3), and interval production (Experiment 4). Thus, if older people are performing on CRT tasks by progressively speeding and slowing RTs to optimize both speed and accuracy, their increased variability of time judgments may make their performance more variable than that of younger people; in other words, age-related slowing of CRTs may be attributable to increases in CRT variability. Models for CRTs are divided on the question of whether variability of CRTs is actively controlled by tracking against some internal standard of what is an optimally fast and accurate response time (Rabbitt & Goward, 1994; Rabbitt & Vyas, 1973; Smith & Brewer, 1995) or whether variability reflects the variability of some passive process of information acquisition (Laming, 1966; Luce, 1986). In either case, the implication is that it may be more fruitful to think of perceptual-motor efficiency as being limited by the variability, rather than merely by the duration, of functional processes. In this framework of description, our current findings of variability of timing behavior can be linked directly to changes in mean CRTs, with associated changes in the variance and skew of the CRT distribution, which characterize individual differences associated with old age and also with differences in intelligence (Rabbitt & Goward, 1994; Vernon, 1983).

## References


Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation, 22,* 39–58.

Birren, J. E., Woods, A. M., & Williams, M. V. (1979). Speed of behavior as an indicator of age changes and of the integrity of the nervous system. In F. Hoffmeister & C. Muller (Eds.), *Brain function and old age* (pp. 10–44). New York: Springer-Verlag.

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 530–544.

Block, R. A. (1990). (Ed.). *Cognitive models of psychological time.* Hillsdale, NJ: Erlbaum.

Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition* (pp. 141–152). Dordrecht, The Netherlands: Kluwer.

Cattell, R. B., & Cattell, A. K. S. (1960). *Handbook for the individual or group Culture Fair Intelligence Test.* Champaign, IL: Institute for Personality and Ability Testing.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98,* 67–83.

Church, R. M., & DeLuty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 216–228.

Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes, 8,* 165–186.

Doob, L. W. (1971). *Patterning of time.* New Haven, CT: Yale University Press.

Feifel, H. (1957). Judgment of time in younger and older persons. *Journal of Gerontology, 12,* 71–74.

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

Gibbon, J., & Allan, L. G. (1984). (Eds.). *Timing and time perception.* New York: New York Academy of Sciences.

Gibbon, J., Church, R. M., & Meck, W. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences, 423,* 52–77.

Gorsuch, R. L. (1983). *Factor analysis.* Hillsdale, NJ: Erlbaum.

Hair, J. F., Jr., Anderson, R. E., & Tatham, R. L. (1987). *Multivariate data analysis.* New York: Macmillan.

Hellstrom, A. (1985). The time-order error and its relatives: Mirrors of cognitive comparing. *Psychological Bulletin, 97,* 35–61.

Hicks, R. E. (1992). Prospective and retrospective judgments of time: A neurobehavioral analysis. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition* (pp. 97–108). Dordrecht, The Netherlands: Kluwer.

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

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review, 94,* 445–468.

Kline, D. W., Holleran, L., & Orme-Rogers, C. (1980). The estimation of short time intervals as a function of age and metronome pacing. *International Journal of Aging and Human Development, 11,* 127–134.

Laming, D. R. (1966). *An information theory of choice reaction times.* New York: Academic Press.

Lejeune, H., & Pouthas, V. (1991). Production de duree chez la personne agee: Comparaison avec le jeune adulte. *L'Annee Psychologique, 91,* 489–504.

Lemlich, R. (1975). Subjective acceleration of time with aging. *Perceptual and Motor Skills, 41,* 235–238.

Licht, D., Morganti, J. B., Nehrke, M. F., & Heiman, G. (1985). Mediators of brief time intervals in elderly domiciled males. *International Journal of Aging and Human Development, 21,* 211–225.

Luce, D. R. (1986). *Response times: Their role in inferring elementary mental organisation.* Oxford, England: Oxford University Press.

Macar, F., Pouthas, V., & Friedman, W. F. (Eds.). (1992). *Time, action and cognition.* Dordrecht, The Netherlands: Kluwer.

Maricq, A. V., Roberts, S., & Church, R. M. (1981). Metamphetamine and time estimation. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 18–30.

McGrath, J. J., & O'Hanlon, J. F. (1968). Relationships among age, intelligence and rate of subjective time. *Perceptual and Motor Skills, 26,* 1083–1088.

Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 171–201.

Michon, J. A. (1990). Implicit and explicit representations of time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 37–58). Hillsdale, NJ: Erlbaum.

Rabbitt, P., & Goward, L. (1994). Age, information processing speed, and intelligence. *Quarterly Journal of Experimental Psychology, 47A,* 741–760.

Rabbitt, P. M., & Vyas, S. M. (1973). An elementary preliminary taxonomy for some errors in choice reaction time tasks. *Acta Psychologica, 33,* 56–76.

Raslear, T. G. (1983). A test of the Pfanzagl bisection model in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 49–62.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam: North-Holland.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Smith, G. A., & Brewer, N. (1995). Slowness and age: Speed–accuracy mechanisms. *Psychology and Aging, 10,* 238–247.

Surwillo, W. W. (1964). Age and the perception of short intervals of time. *Journal of Gerontology, 19,* 322–324.

Vernon, P. A. (1983). Speed of information processing and intelligence. *Intelligence, 7,* 53–70.

Wearden, J. H. (1991a). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation, 22,* 59–83.

Wearden, J. H. (1991b). Human performance on an analogue of an interval bisection task. *Quarterly Journal of Experimental Psychology, 43B,* 59–81.

Wearden, J. H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 134–144.

Wearden, J. H. (1994). Prescriptions for models of biopsychological time. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology* (pp. 215–236). London: Academic Press.

Wearden, J. H. (1995). Categorical scaling of stimulus duration by humans. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 318–330.

Wearden, J. H., & Ferrara, A. (1993). Subjective shortening in humans' memory for stimulus duration. *Quarterly Journal of Experimental Psychology, 46B,* 163–186.

Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology, 48B,* 289–310.

Wearden, J. H., & Ferrara, A. (1996). Stimulus range effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology, 49B,* 24–44.

Wearden, J. H., & Lejeune, H. (1993). Across the Great Divide: Animal psychology and time in humans. *Time and Society, 2,* 87–106.

Wearden, J. H., & McShane, B. (1988). Interval production as an analogue of the peak procedure: Evidence for similarity of human and animal timing processes. *Quarterly Journal of Experimental Psychology, 40B,* 363–375.

Wearden, J. H., & Penton-Voak, I. S. (1995). Feeling the heat: Body temperature and the rate of subjective time, revisited. *Quarterly Journal of Experimental Psychology, 48B,* 129–141.

Received November 18, 1993
Revision received January 30, 1996
Accepted April 8, 1996 ■