THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2005, 58A (4), 693–704

Psychology Press
Taylor & Francis Group

Episodic temporal generalization: A developmental study

T. McCormack
Queen's University, Belfast, Northern Ireland, UK

J. H. Wearden
University of Manchester, Manchester, UK

M. C. Smith and G. D. A. Brown
University of Warwick, Warwick, UK

Groups of 5-year-olds, 10-year-olds, and adults completed either an episodic temporal generalization task, in which no stimuli were repeated, or a repeated standard temporal generalization task, in which there was a fixed standard that was repeated on every trial. Significant developmental improvements were found on both tasks. In both tasks, gradients of performance over two different stimulus ranges superimposed well when plotted on the same relative scale. Performance was similar for the adults and 10-year-olds across tasks, but the 5-year-olds performed better on the repeated standard task. These findings suggest that perceptual processes are a source of scalar variability in timing, and that there are developmental changes in levels of such variability.

A key Piagetian claim has been that perception and estimation of duration are dependent on mastery of concepts of time and measurement (see Levin, 1992, for a comprehensive review of this research). The ability to ignore irrelevant stimulus dimensions and to make reasoned temporal judgements is thought to emerge along with other changes in logical thought (see Pouthas, 1993, for discussion). Developmental studies carried out in the Piagetian tradition have provided valuable insights into the nature of children's reasoning about time and into how this can affect their judgements of duration. However, adult research suggests there are some simple timing tasks that may require limited conceptual resources and that may involve processes that are quite different from more complex tasks in which memory and reasoning

---

Correspondence should be addressed to Teresa McCormack, Department of Psychology, Queen's University, David Keir Building, Malone Road, Belfast, Northern Ireland, BT9 5BP. Email: T.McCormack@qub.ac.uk

This research was supported by a grant from the ESRC (R000 23 9002) awarded to the first, second, and fourth authors, and a grant from the BBSRC (88/S15050) awarded to the first and fourth authors.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980443000250

are heavily involved (Block, 1990; Wearden, 1994). Indeed, much recent research on the development of timing has been carried out largely with the framework of Scalar Expectancy Theory (SET; Gibbon, Church, & Meck, 1984), a framework that was originally developed as a model of animal rather than human timing.

SET has been the most widely applied theory of animal timing, and a number of well-specified models of timing that have emerged within this theoretical tradition can account for a broad range of animal data (Allan, 1998). More recently, this model has been successfully extended to account for human data from tasks that are analogous to those used with animals (Allan & Gibbon, 1991; Wearden, 1991, 1992, 1994; Wearden & Ferrara, 1995, 1996). The details of models have varied, as have the descriptions of the decision rules used to make timing judgements (in particular, decision rules may vary between animals and humans; Wearden, 1992; Wearden & Lejeune, 1993). However, all models in this tradition are characterized by the *scalar property*: that the sensitivity of timing judgements is independent of particular time intervals being considered (an instance of Weber's law). Explanations of developmental changes in timing behaviour within this framework typically refer to changes in basic timing mechanisms and/or the memory and decision processes that operate on their output (Droit-Volet, 2002; Droit-Volet, Clement, & Wearden, 2001; Droit-Volet & Wearden, 2001; McCormack, Brown, Maylor, Darby, & Green, 1999; McCormack, Brown, Maylor, Richardson, & Darby, 2002; Wearden, Wearden, & Rabbitt, 1997b). The aim of this paper is to examine sources of developmental improvement in timing more closely. In particular, we examine whether such improvement reflects more accurate perception of, or better long-term memory for, intervals of time.

There are three processing levels in SET clock-comparison models such as that of Gibbon et al. (1984): clock processes, memory processes, and decision processes. At the clock level, perception of the onset of a to-be-timed duration signals the closing of a switch that allows pulses from a pacemaker to an accumulator. Memory representations are formed from the information in the accumulator, are held in working memory, and, if they are of behavioural significance, are transferred to long-term or reference memory. Typically, decision processes compare time representations in working memory with representations that are retrieved from reference memory. The extent of the similarity between the representation of a just-presented interval and the representation of an interval retrieved from memory predicts the response.

One of the most commonly used tasks within the SET framework is the temporal generalization task (Church & Gibbon, 1982; Wearden, 1992; Wearden, Denovan, Fakhri, & Haworth, 1997a). In this task, participants are pre-exposed to a single standard or reference duration. In the test phase of the task, they have to judge whether or not other comparison stimuli are identical in duration to the standard. Plots of the proportion of positive responses given to test stimulus durations yield a generalization gradient that typically peaks at the standard and declines with distance from it. A SET-consistent model would capture performance on such a task as follows. On the basis of the output of the clock processes, a representation of the standard duration is formed in reference memory. On any given test trial, a sample of the standard is retrieved from reference memory and compared to a representation of the just-presented duration. If the two are sufficiently similar, the just-presented duration will be judged as being the standard. SET models describe such processes mathematically (see Wearden, 1992).

Orderly data from such tasks have been obtained from children as young as three (Droit-Volet et al., 2001), and a number of studies have found consistent developmental differences in levels of performance, as indexed by the steepness of the temporal generalization gradient (Droit-Volet, 2002; Droit-Volet et al., 2001; McCormack et al., 1999; Wearden et al., 1997b). Such developmental changes in performance have typically been modelled in one of two ways. Droit-Volet (2002; Droit-Volet et al., 2001) has assumed that there is variability in the sampling of the standard from long-term memory, with the amount of variability being proportional to the magnitude of the duration, yielding the scalar property. Such an assumption is extremely common in SET models of timing (see Wearden, 1999, for discussion). Developmental changes in the steepness of the generalization gradient are then modelled by assuming that this memory variability declines with age (see Wearden et al., 1997b, for a similar model of the decline of timing abilities with old age). This model assumes noise-free perception of the to-be-judged test stimuli. However, McCormack et al. (1999; see also McCormack, Brown, Smith, & Brock, 2004) made the alternative assumption: that the perception of each test stimulus (rather than the memory of the standard) is susceptible to noise that is proportional to the magnitude of the duration, which again yields the scalar property. They modelled developmental changes by assuming that the level of perceptual noise declines with age (see also McCormack et al., 2002).

For the sake of parsimony, both of these alternatives assume a single source of scalar variability, and the developmental data modelled do not allow us to decide which is the correct assumption regarding that source. However, recent research on human adults has specifically examined whether perceptual processes make an important contribution to variability in timing and, if so, whether perceptual variance possesses the scalar property. This research has used tasks in which reference memory of standards cannot be exploited. Wearden and Bray (2001) developed an "episodic" version of the generalization task, in which stimulus durations were never repeated. Instead, each test trial consisted of a pair of to-be-compared stimuli, for which participants made a same/different judgement. One of the stimuli in the pair was labelled the sample, and the other the comparison stimulus, although the sample was never repeated (i.e., there was no standard or reference duration in this task; the sample was drawn from a particular distribution of durations). The ratio of the stimuli to each other was always one of a set of fixed values increasing and decreasing relative to 1, and gradients were plotted of the number of positive responses to comparison pairs as a function of their ratio. This gradient resembled usual temporal generalization gradients in that it peaked at 1 (when the stimuli are in fact identical) and declined in an orderly way with increasing or decreasing ratio. The scalar property was examined by varying the range of the distribution from which the sample stimulus was drawn. Wearden and Bray found very similar generalization gradients over different stimulus ranges, especially with auditorily presented stimuli (i.e., they demonstrated *superimposition*). They concluded that the clock processes themselves may be the source of scalar variability, since, as the sample stimulus was never repeated, long-term memory processes could not have been recruited.

A similar conclusion was reached by Rodriguez-Girones and Kacelnik (2001), using a modified version of a related task (temporal bisection; see also Wearden & Bray, 2001). On each trial, participants heard two stimuli, one after another. They then had to judge whether a third stimulus was more similar in duration to the first or to the second stimulus that they had heard. Rodriguez-Girones and Kacelnik compared two task conditions: In one condition the first two stimuli were always standards that were repeatedly presented on every trial, and in the other condition the first two stimuli were never repeated so that reference memory representations could not be formed and recruited. They reasoned that if the main source of variability in such judgements was reference memory (i.e., *mnemonic* variance), performance should not benefit in the former condition because participants would not use information from the less reliable reference memory. If in contrast the main source of variability was at the level of perceptual processes (i.e., *perceptual* variance), performance should be less accurate in the latter condition. In fact, performance was less accurate when no memory representations could be exploited. Therefore, along with Wearden and Bray (2001), they concluded that perceptual processes are an important source of variance in human timing. It is important to note that Rodriguez-Girones and Kacelnik (2001) use the term "perceptual" in a particular way—that is, to contrast specifically with long-term memorial processes. Thus, as they point out, perceptual processes include the maintenance of representations of just-presented durations in working memory. In the current paper, we also use the terms perceptual and mnemonic in this way.

Allan and Gerhardt (2001) compared performance on the two types of bisection task used by Rodriguez-Girones and Kacelnik (2001), but they also included the traditional bisection task in which participants were pre-exposed to the standards before each block of testing, but did not encounter them again during test trials. In contrast to Rodriguez-Girones and Kacelnik, Allan and Gerhardt found that overall level of performance was very similar in all three bisection tasks, but that the location of the bisection point was affected by type of task. Like Wearden and Bray (2001), they also found good superimposition across different stimulus ranges in all tasks. The fact that level of performance on the traditional bisection task was similar to that on a bisection task in which there were no fixed standards was interpreted as indicating that, contrary to the assumptions of many models of bisection, participants do not actually compare test stimuli to stored long-term memory representations of the standards in the traditional task. Rather, test stimuli are assumed to be compared with a single criterion duration in both types of task (see also Allan, 2002; Wearden & Ferrara, 1995). Allan and Gerhardt modelled their data by assuming that the primary source of scalar variability in timing was perceptual variability, but that provision of fixed standards either before or during test trials can affect the criterion to which test stimuli are compared.

Thus, a number of recent studies with human adults have used timing tasks in which there are no fixed standards in order to examine whether timing has scalar properties under such circumstances and to gain insight into the contribution of perceptual relative to mnemonic processes to variability in timing. The present paper takes advantage of the same basic methodology but for a novel purpose: examination of the possible contributions of perceptual and mnemonic processes to developmental improvements in timing. We devised two tasks that resembled those of Rodriguez-Girones and Kacelnik (2001) and Wearden and Bray (2001; Exp. 3). We chose variants of the temporal generalization task because it is relatively easy to interpret data on this task compared to the bisection task, the theoretical interpretation of which is the subject of intense debate. The same type of test trial was used in both of our temporal generalization tasks: Participants were presented with a pair of stimuli and had to judge whether or not the stimuli were of the same duration. In the *repeated standard* task, one of the two stimuli was always a standard stimulus that participants had been exposed to before the test trials began. The ratio of test comparison stimulus to the standard was always 0.25, 0.5, 0.75, 1, 1.25, 1.5, or 1.75, enabling generalization gradients to be plotted as a function of comparison-to-standard ratio. In the *episodic* task, the ratio between the two stimuli was again always taken from the same set of values (i.e., ranging from 0.25 to 1.75), but no stimulus duration was ever repeated. Instead, one stimulus from the pair was drawn from a particular distribution of durations. To examine the scalar property of timing under these circumstances, two different duration distributions were used in the episodic task ("short" and "long" ranges), and two different standards were used in the repeated standard task. These standards were in fact the mean durations of the short and long ranges in the episodic task.

We can make the following predictions. First, if perceptual variance is scalar in nature, we should expect to replicate previous findings with adults, in that similar generalization gradients should be obtained for the two stimulus ranges in both the episodic task and the repeated standard task (i.e., we should see superimposition even in the absence of reference memory, Allan, 2002; Allan & Gerhardt, 2001; Wearden & Bray, 2001). Superimposition is predicted for the groups of children as well as adults, given that previous findings suggest that timing is scalar even in young children (Droit-Volet, 2002; Droit-Volet et al., 2001). Second, if developmental differences are due to changes in levels of perceptual variance rather than simply changes in mnemonic variance, age differences should be found on the episodic task (in which long-term memory plays no role) as well as on the repeated standard task.

The third issue concerns the predicted levels of performance on the episodic task relative to the repeated standard task. The findings of Rodriguez-Girones and Kacelnik (2001) lead to the prediction of better performance on the repeated standard task, which would indicate that relatively precise long-term memory representations can be used to moderate relatively noisy perceptual processes on this task. However, in contrast, Allan and Gerhardt's (2001) findings lead to the expectation that performance levels should be similar on both tasks, indicating that long-term memories are not recruited in the repeated standard task. A further possibility is that the age groups may differ in the extent to which they actually recruit the long-term memory representation of the standard in the repeated standard task. As Rodriguez-Girones and Kacelnik point out, the use of reference memory is in some sense optional in such a task, given that the standard is repeated on every trial. If groups differ in their use of reference memory in the repeated standard task, we might expect to see variation across age groups in any differences in levels of performance between that task and the episodic task.

Method

Participants

A total of 26 five-year-olds (*M* = 5;2, range = 4;11–5;8; 10 males and 16 females), 26 ten-year-olds (*M* = 10;3, range = 9;5–10;7; 15 males and 11 females), and 25 adults (13 males and 12 females) completed the repeated standard task. A total of 26 five-year-olds (*M* = 5;1, range = 4;8–5;6; 15 males and 11 females), 26 ten-year-olds (*M* = 10;3, range = 9;7–10;6; 13 males and 13 females), and 26 adults (10 males and 16 females) completed the episodic task. Children were recruited from primary schools in Birmingham, UK, and were tested individually in a quiet room in their schools. Adult participants were primarily undergraduates from the University of Warwick, who took part voluntarily and were paid 4 pounds sterling for their participation.

Materials

An Apple Macintosh laptop was used to present the auditory and visual stimuli, with stimulus presentation controlled by the Psyscope computer program. The auditory stimuli were 500-Hz pure tones that varied in duration; these were presented through external speakers attached to the computer. The visual stimuli consisted of cartoon images of a pair of green birds that were identical except for their size.

Procedure

In the repeated standard task, participants completed two generalization tasks, the first with a "short" standard of 495 ms and the second with a "long" standard of 825 ms. The comparison durations in both tasks had ratios of 0.25, 0.5, 0.75, 1, 1.25, 1.5, or 1.75 to the standard. Participants completed six blocks of seven trials in each task, with each block of trials containing one example of each comparison ratio. Each trial began with the appearance of the larger of the two cartoon birds on the left hand side of the computer screen. This bird remained on the screen for 3 s in total. After this bird had been on the screen for 500 ms, the standard stimulus was presented. As soon as the larger bird disappeared from the screen, a second smaller bird appeared on the right hand side of the screen. This bird remained on the screen until the participant keyed in a response. At a given interval after the appearance of the second bird, the comparison stimulus was presented. This interval varied randomly between 400 and 600 ms.

At the start of the experiment, the structure of each trial was demonstrated to participants in an initial block consisting of six trials. In particular, it was explained to the participants that the larger bird was "mummy" bird and that the smaller bird was "baby" bird and that mummy bird was trying to teach baby bird to sing. Participants were encouraged to help baby bird by informing him when he had correctly made the same sound as mummy bird and when he had incorrectly made a different sound to mummy bird. It was then explained that participants could inform baby bird by pressing a "same" or "different" button on the computer. After the initial demonstration block, participants received two training blocks of six trials in which they received feedback about whether baby bird's sounds were the same as or different from mummy bird's sounds. Prior to the blocks of trials featuring the short standard, participants were informed that mummy bird would be making a short sound for baby bird to copy. In contrast, prior to the blocks of trials featuring the long standard, participants were informed that mummy bird would be making a long sound for baby bird to copy. Half of the participants were presented with blocks of trials featuring the short standard first, and half of the participants were presented with blocks of trials featuring the long standard first. Apart from the demonstration and practice blocks, no feedback was given in this or in the episodic task.

In the episodic task, the sample stimulus was randomly drawn from either a "short" set or a "long" set. Participants received 42 trials in which the sample stimulus was drawn from the short set, and 42 trials in which it was drawn from the long set, with half the participants receiving the short trials first, and half the participants receiving the long trials first. The shortest sample stimulus in the short set was 369 ms, with the other stimuli in the set increasing in 6-ms increments up to 615 ms. The shortest sample stimulus in the long set was 615 ms, with the other stimuli in the set increasing in 10-ms increments up to 1,025 ms. In the episodic task, the comparison durations had ratios of 0.25, 0.5, 0.75, 1, 1.25, 1.5, or 1.75 to the sample, with six examples of each ratio for each range. In all other respects the trial and block structure for the episodic task was identical to that employed on the repeated standard task. The instructions for the episodic task were also identical to that for the repeated standard task. Before the short trials, participants were told that mummy bird would be making some short sounds for baby bird to copy, and before the long trials they were told that she would be making some long sounds for baby bird to copy.

Results

Figure 1 shows the proportion of positive responses in the episodic task as a function of comparison-to-sample ratio for each age-group; Figure 2 shows the proportion of positive responses in the repeated standard task as a function of comparison-to-standard ratio for each age group.

In the episodic task, there appear to be age-related improvements in performance, particularly between the youngest age group and the two older groups, with the youngest age group showing less steep gradients. Superimposition is generally good across the three

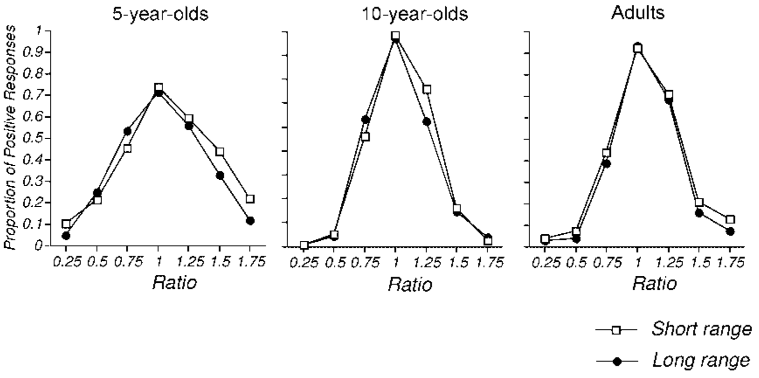

Figure 1. Temporal generalization gradients from the episodic task. The mean proportion of positive responses (indicating that the two stimuli on the trial have the same duration) is plotted against comparison/sample ratio. Left panel: data from 5-year-olds. Centre panel: data from 10-year-olds. Right panel: data from adults. Within each panel data are shown separately from the short and long duration ranges used.

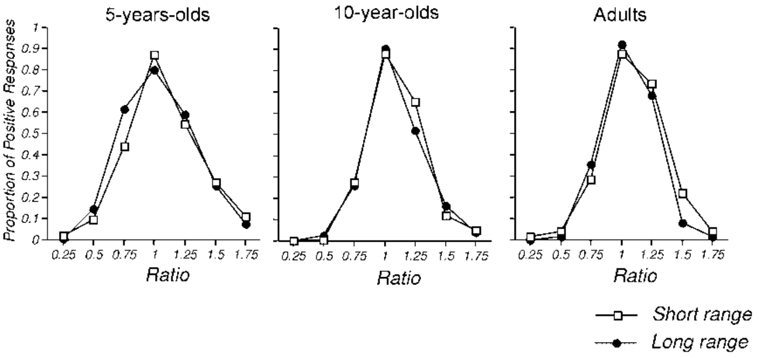

Figure 2. Temporal generalization gradients from the repeated standard task. The mean proportion of positive responses is plotted against comparison/standard ratio. Left, centre, and right panels show data from the 5-year-olds, 10-year-olds, and adults, respectively. Within each panel data obtained from the short and long duration ranges are shown separately.

age groups. A three-way analysis of variance (ANOVA) of these data, with a between-subjects factor of age group and within-subjects factors of ratio (0.25–1.75) and range (short vs. long), confirmed these observations.[1] The main effect of age group was significant, $F(2, 75) = 5.69$, $p < .01$, indicating that the groups differed in overall number of positive responses. The interaction between age group and ratio was also significant, $F(8.54, 450) = 11.46$, $p < .001$, indicating age differences in the steepness of the generalization gradients and thus in overall levels of performance. The effect of range just reached significance, $F(1, 75) = 4.10$, $p = .05$, indicating a difference in the overall number of positive responses between the two ranges. However, the interaction between range and ratio was not significant, $F(4.32, 450) = 1.78$, $p > .05$, indicating that the shapes of the generalization gradients were similar for the short and long ranges (i.e., that there was superimposition). Furthermore, the three-way interaction between age, range, and ratio was also not significant, $F(8.63, 450) = 1.36$, $p > .05$. Thus, the analyses confirm that there was age-related improvement on the episodic task and that superimposition was found.

For the repeated standard task, it can be seen that performance also improves with age, with the youngest group having less steep generalization gradients than the two older groups. Superimposition again appears generally good for each of the age groups. A repeated measures ANOVA was carried out on the proportion of positive responses, with a between-subjects factor of age group and within-subjects factors of ratio (0.25–1.75) and range (short vs. long). The main effect of age group was significant, $F(2, 74) = 4.38$, $p < .02$, as was the interaction between age group and ratio, $F(8.07, 444) = 5.94$, $p < .001$, indicating changes in overall levels of performance. The main effect of range was not significant ($F < 1$), indicating similar levels of positive responses in both ranges, but the interaction between range and ratio was significant, $F(3.93, 444) = 2.92$, $p < .05$, as was the three-way interaction between range, ratio, and age, $F(7.87, 444) = 2.73$, $p < .01$. Further analysis revealed that the interaction between ratio and range was not significant for the adult group, $F(2.99, 144) = 2.28$, $p > .05$, and only marginally significant for the 10-year-old group, $F(3.44, 150) = 2.63$, $p = .05$, but significant for the youngest group, $F(3.40, 150) = 3.42$, $p < .05$. Post hoc tests showed that the difference between ranges was significant only at the 0.75 and the 1 ratio for the 5-year-olds and at the 1.25 ratio for the 10-year-olds (a significance level of $p < .05$ was set for these analyses). Thus, superimposition was excellent for the adults and reasonably good for the two groups of children.

Cross-task comparisons. In order to compare performance across task, the data were collapsed over the short and the long ranges. Figure 3 shows the performance of each age group across the two tasks, with the data collapsed across range. It can be seen that the groups of adults and 10-year-olds performed at remarkably similar levels across the tasks. However, the 5-year-old group who completed the repeated standard task appear to perform more accurately than those who completed the episodic task. A two-way ANOVA was conducted on proportion of positive responses, with a between-subjects factor of age group and task and

---

[1] In this and subsequent ANOVAs, there was some evidence of departure from the sphericity assumption (ε values ranged from .49 to .72). Thus, all reported degrees of freedom and probability values have been adjusted using the Greenhouse–Geisser correction.

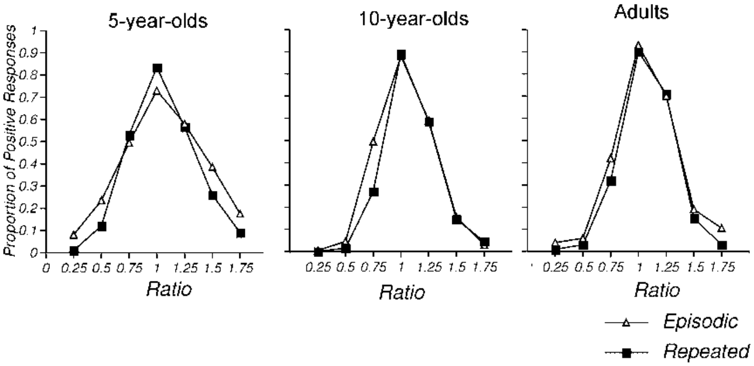

Figure 3. Temporal generalization gradients from both the episodic task and the repeated standard task, collapsed across short and long duration ranges. The mean proportion of positive responses is plotted against comparison/standard ratio. Left, centre, and right panels show data from the 5-year-olds, 10-year-olds, and adults, respectively.

a within-subjects factor of ratio. The interaction between ratio and task was significant, $F(4.26, 894) = 3.66$, $p < .001$, indicating task differences in overall levels of performance. However, the three-way interaction between ratio, age, and task was also significant, $F(8.53, 894) = 3.85$, $p < .001$. Further analysis showed that the interaction between task and ratio was not significant for the adult group, $F(4.04, 294) = 1.12$, $p > .05$. This interaction was significant for the 10-year-old group, $F(3.12, 300) = 5.06$, $p < .01$; however, t tests showed that the task difference was only significant for the 0.5 ratio, $t(50) = 2.07$, $p = .05$, and the 0.75 ratio, $t(50) = 4.13$, $p < .05$. Participants made more erroneous positive responses in episodic task on these ratios, although it can be seen from Figure 3 that performance was extremely similar for all other ratios. The interaction between task and ratio was significant for the 5-year-old group, $F(4.16, 300) = 4.38$, $p < .01$, with t tests showing that the task difference was significant for all ratios, $t(50) > 2$, $p < .05$, with the exception of the 0.75 and 1.25 ratios.

Discussion

Consistent with the findings of previous similar studies (Allan, 2002; Allan & Gerhardt, 2001; Wearden & Bray, 2001) we found evidence that perceptual processes exhibit the scalar property that is central to SET models of timing. Superimposition of performance across two stimulus ranges was good for all groups in the episodic task, in which no long-term memory representations of standards could be used. Superimposition was also reasonably good in the repeated standard task. Although the groups of children (in particular, the 5-year-old group) showed a small amount of deviation from superimposition in this task (see Figure 2), this deviation was not systematic. Thus, we attribute this to the higher probability of obtaining noisy data from young children, rather than to a qualitative difference in the properties of temporal processes in this group. Droit-Volet and Wearden (2001) found even more marked deviation from superimposition in their study of temporal bisection in children, but since the deviation was not systematic they did not take it as evidence against the claim that timing is scalar in children. Indeed, Droit-Volet et al. (2001) found good superimposition across two very different stimulus ranges in a study of temporal generalization using groups of children as young as 3 years old, as did Droit-Volet (2002). Thus, the general picture of findings would suggest that timing can reasonably be assumed to have the scalar property in children as well as adults.

Developmental improvements in performance were found on both the episodic task and the repeated standard task. The finding that performance improves developmentally on a timing task even when there is no long-term memory component to the task implies that future developmental models of timing must consider perceptual as well as mnemonic changes. Some previous developmental models of timing within the SET framework have assumed either perceptual variance with error-free long-term memory (McCormack et al., 1999), or mnemonic variance with error-free perception (Droit-Volet et al., 2001). Our findings suggest that error-free, developmentally invariant perception cannot be assumed. However, it remains to be established exactly which aspects of the set of processes that we have categorized here as "perceptual" change developmentally. Recently, Droit-Volet (2003) has found that alerting young children to the impending onset of to-be-timed stimuli improved performance, and argued that the attentional processes that trigger the timing of stimuli are poorer in young children. Although attentional processes may indeed contribute to developmental change, we note that what we have categorized for present purposes as perceptual processes include not only the clock processes that are used to time durations, but the processes involved in the maintenance of to-be-compared durations in short-term memory. These short-term memory processes may also change developmentally.

We now turn to the cross-task comparisons. The findings from the groups of adults and 10-year-olds resemble those of Allan and Gerhardt (2001), who found similar levels of performance in tasks with or without fixed standards. In this respect our results differ from those of Rodriguez-Girones and Kacelnik (2001), who found better performance on a repeated standard task. Allan and Gerhardt point out, however, that the cross-task comparisons in the Rodriguez-Girones and Kacelnik study might be misleading because the ranges of the stimulus durations in the two tasks were not equated. In the current study, we attempted to equate for stimulus range across the tasks.

Different levels of performance on the tasks can be interpreted as reflecting the variability of perceptual processes relative to mnemonic process. Similar levels of cross-task performance are interpreted as indicating that perceptual processes are not associated with higher levels of noise than mnemonic processes. If mnemonic processes were less noisy than perceptual ones, then participants would be able to use such processes to moderate perceptual variance, leading to better performance on the repeated standard task. In fact, only the 5-year-old group showed any marked advantage from the presence of a fixed standard in the repeated standard task. The 10-year-old group showed a limited advantage from the presence of a fixed standard, and the adult groups performed very similarly on both tasks. This would suggest that one important developmental change may be the magnitude of perceptual variance relative to mnemonic variance. It seems possible that perceptual processes are particularly noisy in young children, perhaps due to poor attentional skills, as Droit-Volet (2003) has suggested.

The findings of the current study suggest that an important aim of future developmental research should be to partition developmental changes in variance into changes in either perceptual or mnemonic variance and then to further examine which specific aspects of perceptual and memory processes contribute to developmental change. The current experiment was not designed to enable such partitioning; rather, its findings simply rule out the possibility that developmental changes are entirely due to changes in mnemonic variance. However, it is a strength of the SET framework that it allows for the construction of formal models that potentially separate different perceptual and mnemonic sources of variance. Although it may not be straightforward to do so, future studies of developmental changes in timing could attempt to exploit this strength, in order to yield a more detailed account of the specific processes that underpin developmental improvements in timing abilities.

REFERENCES

Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*, 101–117.

Allan, L. G. (2002). Are the referents remembered in temporal bisection? *Learning and Motivation*, *33*, 10–31.

Allan, L. G., & Gerhardt, K. (2001). Temporal bisection with trial referents. *Perception & Psychophysics*, *63*, 524–540.

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

Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Droit-Volet, S. (2002). Scalar timing in temporal generalization in children with short and long stimulus durations. *Quarterly Journal of Experimental Psychology*, *55A*, 1193–1209.

Droit-Volet, S. (2003). Alerting attention and time perception in children. *Journal of Experimental Child Psychology*, *85*, 372–384.

Droit-Volet, S., Clement, A., & Wearden, J. (2001). Temporal generalization in 3- to 8-year-old children. *Journal of Experimental Child Psychology*, *80*, 271–288.

Droit-Volet, S., & Wearden, J. (2001). Temporal bisection in children. *Journal of Experimental Child Psychology*, *80*, 142–159.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. G. Allan (Eds.), *Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.

Levin, I. (1992). The development of the concept of time in children: An integrative model. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action, and cognition: Towards bridging the gap* (pp. 13–32). Dordrecht, The Netherlands: Kluwer.

McCormack, T., Brown, G. D. A., Maylor, E. A., Darby, R. J., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology*, *35*, 1143–1155.

McCormack, T., Brown, G. D. A., Maylor, E. A., Richardson, L. B. N., & Darby, R. J. (2002). Effects of aging on absolute identification of duration. *Psychology and Aging*, *17*, 363–378.

McCormack, T., Brown, G. D. A., Smith, M. C., & Brock, J. (2004). A timing specific memory distortion effect in young children. *Journal of Experimental Child Psychology*, *87*, 33–56.

Pouthas, V. (1993). Ontogenesis of temporal learning in the child: Experimental evidence and perspectives. *Psychologia Belgica*, *33*, 171–183.

Rodriguez-Girones, M., & Kacelnik, A. (2001). Relative importance of perceptual and mnemonic variance in human temporal bisection. *Quarterly Journal of Experimental Psychology*, *54A*, 527–546.

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

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. (1999). Beyond the fields we know: Exploring and developing scalar timing theory. *Behavioural Processes*, *45*, 3–21.

Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory? Episodic temporal generalization and bisection in humans. *Quarterly Journal of Experimental Psychology*, *54B*, 289–309.

Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997a). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 502–511.

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., Wearden, A. J., & Rabbitt, P. M. A. (1997b). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 962–979.

Original manuscript received 8 July 2003
Accepted revision received 22 March 2004