THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2001, *54B* (4), 289–309

Scalar timing without reference memory? Episodic temporal generalization and bisection in humans

J.H. Wearden and Simon Bray

Manchester University, Manchester, UK

Three experiments tested whether the *scalar property* of timing could occur when humans timed short durations under conditions in which it was unlikely that they developed reference memories of temporal "standards". Experiment 1 used an episodic version of a temporal generalization task where judgements were made of the potential equality of two durations presented on each trial. Unknown to the subject, one of these was always 200, 400, 600, or 800 ms, and the other was of variable duration. Temporal generalization gradients showed the scalar property of superimposition at standard values greater than 200 ms. Experiment 2 used a variant of the "roving bisection" method invented by Rodriguez-Girones and Kacelnik (1998) modified so that the scalar property of timing could be observed empirically. Data from bisection with short/long standard pairs of 100/400, 200/800, and 300/1,200 ms showed nearly perfect scalar-type superimposition. Experiment 3 again used episodic temporal generalization, but durations were never repeated and came from three distinct time ranges. Superimposition was found across these ranges except for the shortest visual stimuli timed. The data suggested that scalar timing could occur in humans in conditions where the formation of reference memories of temporal standards was highly improbable.

One of the recent success stories of that part of comparative psychology dealing with comparisons between the behaviour of humans and that of other animals has been the application of *scalar timing theory* (or *scalar expectancy theory*, SET: See Gibbon, 1977, or Gibbon, Church, & Meck, 1984, for discussion) to timing behaviour in humans. SET was originally developed as an explanation of the performance of rats and pigeons on reinforcement schedules that involved various temporal periodicities or that imposed temporal requirements on responding (e.g., see Lejeune & Wearden, 1991). This was later followed by the development of animal timing tasks intended to test specific aspects of SET's predictions, such as temporal generalization (Church & Gibbon, 1982) and "time-left" (Gibbon & Church, 1981). The history and development of SET as it relates to animal timing have been narrated in detail by Gibbon (1991).

Requests for reprints should be sent to J.H. Wearden, Department of Psychology, The University, Manchester, M13 9PL, UK. Email: wearden@psy.man.ac.uk

We are grateful to "TP" groups for running subjects in Experiments 2 and 3 for us and to Dr Don O'Boyle for helpful discussions of statistical analysis.

http://www.tandf.co.uk/journals/pp/02724995.html DOI:10.1080/02724990042000173

The last decade has seen extensive application of SET to timing in humans, especially in situations where humans cannot use their unique resources such as chronometric counting. Allan (1998) reviews the application of SET to human timing in some depth but, in summary, methods and theoretical ideas derived from SET have been used with humans on a range of procedures, some linked directly to animal timing tasks (the so-called *analogue experiments*, see Wearden, 1991a), but others involving judgements that animals cannot perform, such as verbal estimation of duration (Penton-Voak, Edwards, Percival, & Wearden, 1996). A more recent development has been the use of ideas compatible with SET to attack some classical problems in human time perception, such as why "tones are judged longer than lights" (Wearden, Edwards, Fakhri, & Percival, 1998).

The claim that the timing behaviour of humans often conforms to SET rests principally on the manifestation in human behaviour of the *scalar property* of timing, which gives SET its name. One way to demonstrate the scalar property is to construct a *coefficient of variation* statistic (standard deviation/mean) from measures of timed behaviour. This Weber-fraction-like index should remain constant as the absolute duration timed varies, and this was indeed found in the first clear demonstration of scalar timing in humans, that of Wearden and McShane (1988). However, another common test of the scalar property is to try to demonstrate *superimposition* (called *superposition* in the US literature). Superimposition is found when measures of timed behaviour from different absolute time values superimpose when plotted on the same relative scale, hence the term *scalar.* The method of temporal generalization (variants of which are used in Experiments 1 and 3) can be used to illustrate superimposition. In the temporal generalization method developed by Wearden (1992) from the original animal experiment by Church and Gibbon (1982), subjects initially received a stimulus identified as having a "standard" duration (e.g., a 400-ms tone). They then received a range of comparison durations, including values shorter than, longer than, or equal to the standard, and had to decide whether each stimulus presented was or was not the standard duration, with accurate feedback being given after each decision. Orderly behaviour rapidly emerged with this procedure, and when the proportion of identifications of a stimulus as the standard was plotted against duration, the resulting function (the *temporal generalization gradient*) peaked at the standard and fell off progressively as durations became more remote from the standard, in both the longer and the shorter directions. If the generalization gradients from conditions with different absolute standard durations are plotted on the same relative scale (with each comparison stimulus duration being expressed as a fraction of the standard in force) SET requires that the data superimpose, and this has been found when humans timed short durations (e.g., standards of 400 to 700 ms; Wearden, 1992), and longer durations without chronometric counting (e.g., standards of 2, 4, 6, and 8 s, see Wearden, Denovan, Fakhri, & Howarth, 1997). Superimposition on temporal generalization in humans holds over variations of stimulus duration, modality, and stimulus spacing (e.g., linear or logarithmic spacing of durations) although it can sometimes be violated (e.g., Ferrara, Lejeune, & Wearden, 1997; Wearden, 1999). Superimposition has also been found in human timing behaviour on other tasks, such as categorical timing (Wearden, 1995) and bisection (Wearden, 1991b; Wearden & Ferrara, 1995, 1996; Wearden, Rogers, & Thomas, 1997; see also Allan & Gibbon, 1991).

If the scalar property is so often found in human timing, where does it come from? As is well known, SET is a multi-process theory, involving internal clock, memory, and decision processes. A fuller non-technical introduction to SET is provided elsewhere (Wearden, 1994) but, in brief, an internal clock of a pacemaker–accumulator type is proposed to provide the "raw" representations of stimuli whose duration is to be judged. The representation provided by the clock is transferred, more or less without change, to a short-term "working memory", and it can then be transferred to a longer term "reference" memory, if appropriate. The reference memory contains representations of "important times", such as those associated with reinforcement in animal experiments, or identified in some way as standards in experiments with humans. In most tasks, samples from reference memory are compared with representations in working memory, and various types of decision processes are used to generate observed responses.

Most theoretical analysis using SET has assumed that the reference memory is the source of scalar variance, either because scalar variance arises by "transformation" of time representations during transfer from working to reference memory, or because the reference memory generates scalar variance in some other way. Accordingly, quantitative models of performance on timing tasks have usually embodied the assumption that reference memory is the source of scalar variance. For example, Wearden's (1992) model of temporal generalization performance in humans (a variant of an earlier model by Church & Gibbon, 1982) explains behaviour as follows. On each trial, the just-presented stimulus duration, t, is assumed to be timed without variance, and this is compared with a sample s^* drawn from the reference memory of the standard, s. The reference memory of s is assumed to be a Gaussian distribution with mean s and coefficient of variation, c, which is constant as s varies in absolute value between conditions. The model judges that t is "close enough" to s when $|(s^*-t)|/t < b^*$, where b^* is some threshold, which is variable from trial to trial, although there is little loss of predictive accuracy by assuming that b is fixed (Wearden, 1992).

The temporal generalization model provides a good fit to data obtained when students time short durations (<1 s; Wearden, 1992), and when elderly people do (Wearden, Wearden, & Rabbitt, 1997), and it also fits data when stimulus durations more than 10 times as long are used, providing that chronometric counting is prevented (Wearden, Denovan et al., 1997). The problem is that the model outlined is not the only one possible: Variance can be added to t without any loss of predictive precision, as Wearden (1992) showed. More generally, scalar variance can be incorporated in different parts of the SET system, with resulting behavioural predictions being highly similar or even identical (see discussion in Gibbon, 1992).

How, then, can the source of scalar variance in human timing be identified? One approach, suggested by Wearden (1999), and used in the three experiments reported here, is to develop methods that discourage the development of "standards" in reference memory, methods we refer to here as *episodic*. For example, in normal temporal generalization, a single stimulus is presented on each trial, and the subject must compare this with the remembered standard duration. But suppose that, instead, two stimuli are presented during the trial, and the subject must indicate whether they have the same duration or not. If the two stimuli vary markedly and apparently randomly from trial to trial, it is hard to see how developing any kind of "standard" can help performance, as the judgement made on one trial seems of little use in aiding discrimination on the next. Rodriguez-Girones and Kacelnik (1995, 1998) deserve historical priority in developing such methods (and their *roving bisection* method inspired our Experiment 2), but Experiment 1 uses an episodic variant of the conceptually somewhat simpler temporal generalization technique.

EXPERIMENT 1

Experiment 1 reports data collected with a temporal generalization variant that we refer to as *episodic* temporal generalization. On each trial two stimuli (either tones or visual stimuli) are presented, and the subject is asked whether they are equal in duration or not, without feedback being given. There are in fact 28 different stimulus pairs in a block of experimental trials, with "standards" (not, of course, identified in any way to the subject) of 200, 400, 600, and 800 ms each being paired with seven different comparison values. The trials in the block are intermingled, so that, for example, a judgement of the possible equality of 200 and 50 ms could be followed on the next trial by the same judgement made about 1,400 and 800 ms, stimuli that are obviously very different. We assume that subjects will perform this episodic task by using short-term memory representations of the two stimuli present on the trial, without reference memory. Thus, according to the SET architecture, the task involves internal clock, short-term memory, and decision processes, but not reference memory. However, the selection of stimuli in the experiment overall enables the scalar property of timing to be tested by examining potential superimposition. If scalar timing is found in this task then, at least according to our arguments, reference memory is very unlikely to be its source.

Method

Subjects

A total of 30 Manchester University undergraduates were arbitrarily allocated to two equal-sized groups. One group performed in an auditory condition, the other in a visual condition.

Apparatus

An Opus 16X (IBM-compatible) computer controlled all experimental events. The computer speaker produced the tones used in the auditory condition, and the stimuli for the visual condition were presented on a standard colour monitor. The computer keyboard registered responses, and the experimental programs were written in the MEL language (Micro-Experimental Laboratory: Psychology Software Tools, Inc.), which assured millisecond accuracy for timing of stimuli and responses.

Procedure

All subjects received a single experimental session lasting about 25 min. The procedure for the auditory and visual conditions was identical except for the stimuli whose durations had to be compared, and some attendant changes in displayed instructions. The procedure for the auditory condition was as follows. The stimuli to be judged were 500-Hz tones produced by the computer speaker. On each trial, two stimuli were presented, separated by a random gap running from 400 to 600 ms. To produce the stimuli, the subject pressed the spacebar of the computer keyboard in response to an appropriate prompt, and this response was followed after a random duration, selected from a uniform distribution running from 2,000 to 4,000 ms, by the stimulus presentations. When the two stimuli had been presented, the subject was asked whether they had the SAME duration and had to respond YES (Y key) or NO (N key). No feedback was given as to performance accuracy. The experimental trials were arranged into five blocks of 28 stimulus pairs, and within each block the stimulus pairs were randomly selected until all had been presented. A different random order of trials was used for each subject and each block. The 28 trials of the block were made up by comparing four "standard" durations with seven "comparison" durations. The standards were 200, 400, 600, and 800 ms, and comparison values for the 400, 600, and 800-ms conditions were

400: 100, 200, 300, 400, 500, 600, and 700 ms; 600: 150, 300, 450, 600, 750, 900, and 1,050 ms; 800: 200, 400, 600, 800, 1,000, 1,200, and 1,400 ms. The intended comparison durations for the 200-ms standard were 50, 100, 150, 200, 250, 300, and 350 ms but, as a result of a programming error, a value of 10 ms was substituted for 100 ms. This error only affected one of the 28 trials in each block. On each trial of the block the subject received a standard and a comparison interval, with the standard first on a random 50% of trials and the comparison first on the rest. There was no constraint on trial order within a block, so the subject could receive very different absolute durations on consecutive trials.

The procedure for the visual condition was identical except that the stimuli used were 14 × 14-cm light-blue squares displayed in the centre of the computer screen. The subject was instructed to judge whether or not the duration of presentatin of the stimuli was the same, again without feedback. The programming error (substitution of 10 ms for 100 ms on trials with a 200-ms standard) was also present in the visual condition.

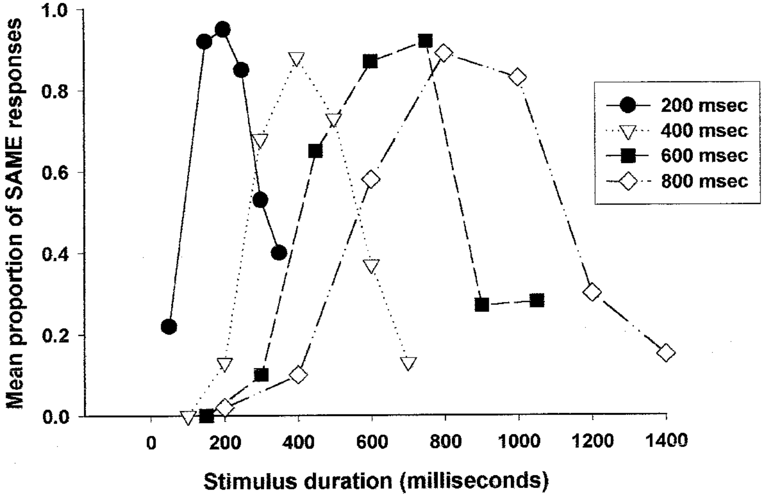
Results

Data were taken from the last four of the five blocks of the experiment. The upper panel of Figure 1 shows the mean proportion of judgements that the stimuli presented on the trial had the SAME duration, plotted against the duration of the comparison stimulus. Data are shown separately for conditions with the 200, 400, 600, and 800-ms standards, and separately for conditions with the 200, 400, 600, and 800-ms standards, and separately for auditory stimuli (upper panel) and visual stimuli (lower panel). The 10-ms comparison duration used in error in the 200-ms condition was hardly ever confused with the standard, and data from this trial type are omitted from the figures and the following analyses.

Inspection of the data suggests that the response functions varied in an orderly way with changes in comparison stimulus durations: Decreasing the duration difference between the standard and comparison stimuli systematically produced more SAME judgements. In the auditory case, the response functions for the 200, 400, and 800-ms standards peaked when the standard/comparison difference was zero; the response function from the 600-ms condition peaked at the comparison value (750 ms) just greater than the standard. In the visual condition, the response functions peaked at the zero standard/comparison difference when 200, 600, and 800-ms standards were used, and at the value just greater than the standard in the 400-ms condition. The effect of comparison stimulus duration on proportion of SAME responses is obvious in Figure 1, but it was also confirmed statistically in all conditions.

For the 200-ms auditory and visual conditions, only six different comparison durations were entered into the analysis of variance (ANOVA; i.e., the intended 100-ms value was omitted), and a significant effect of duration was found, $F(5, 70) = 32.01$ for auditory and 28.82 for visual, both $p < .001$. For the other conditions, all seven comparison durations were entered into the ANOVA, and all standard values in both the auditory and visual modalities showed significant effects of stimulus duration on proportion of SAME responses, smallest $F(6, 84) = 22.39$, $p < .001$.

Figure 2 shows the same data as in Figure 1, but this time rescaled so that each comparison stimulus value is expressed as a fraction of the standard value in force. Data from the auditory conditions are shown in the upper panel, those from the visual conditions in the lower one. Inspection of the data in both panels suggests that property of superimposition was manifested, except for the 200-ms visual and, particularly, auditory conditions, which tended to



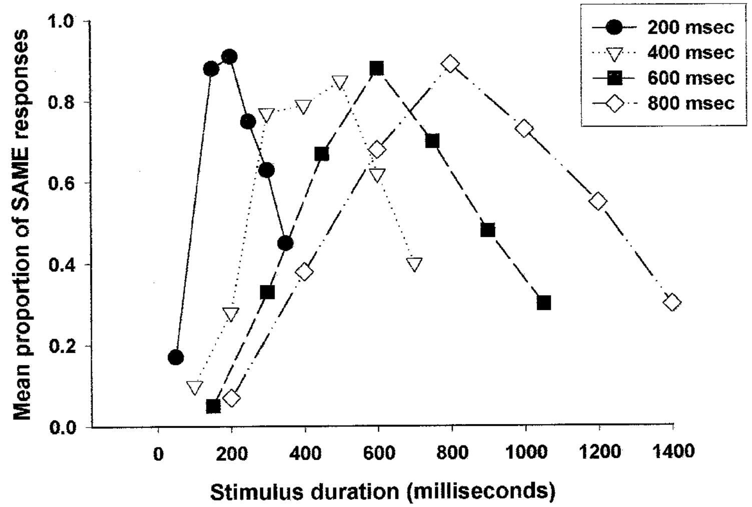


Figure 1. Temporal generalization gradients from Experiment 1. Upper panel shows data from the group tested with auditory stimuli, the lower panel data from the group tested with visual stimuli. The mean proportion of SAME responses (i.e., judgements that the two stimuli presented on the trial had the same duration) is plotted against comparison stimulus duration in ms. Data are shown separately for the conditions with the 200, 400, 600, and 800-ms standards.

generate higher proportions of SAME responses than the other standards for the same comparison/standard values.

As the comparison values from the different conditions were always the same ratios of standard for the condition, one possible test for superimposition is by ANOVA of the proportion of SAME responses using comparison/standard ratio as the independent variable. A

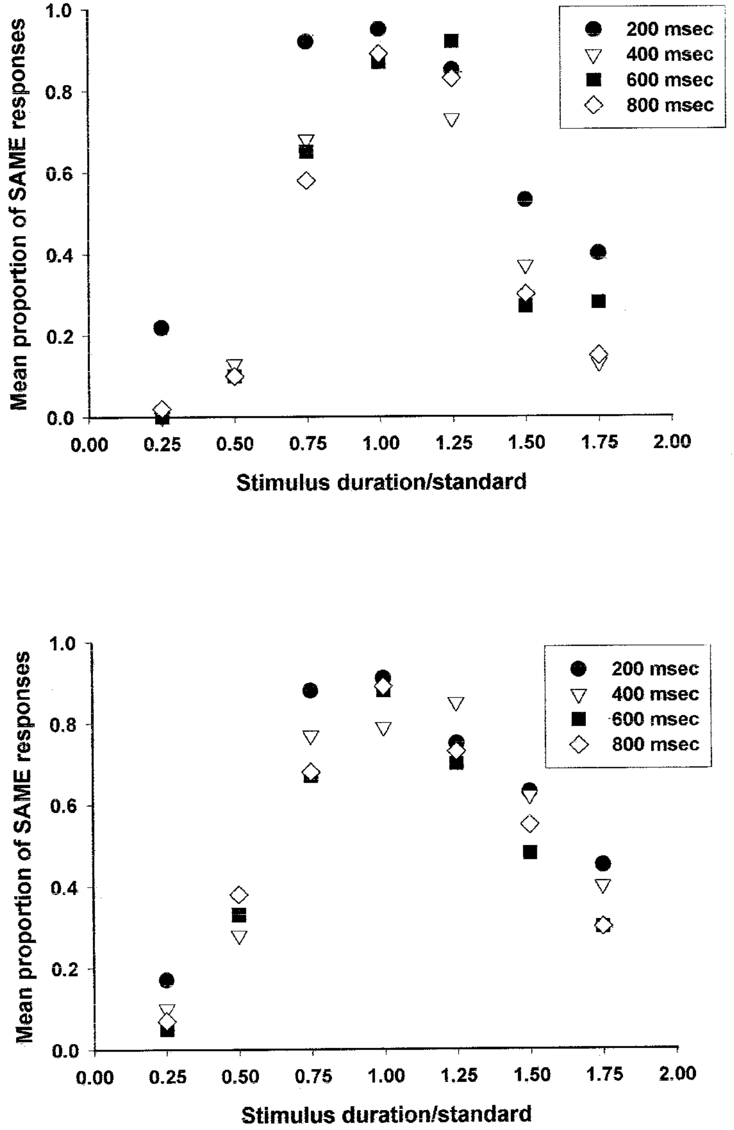


Figure 2. Temporal generalization gradients from Experiment 1, shown with mean proportion of SAME responses plotted against comparison stimulus value divided by the standard value in force. Other details as in Figure 1.

complication is that the 200-ms condition only involves six usable stimulus values so, to compensate for this, the second shortest stimulus value in each of the other conditions (400, 600, and 800-ms standards) was omitted.

Consider first the auditory stimuli. ANOVA revealed significant effects of standard duration, $F(3, 42) = 15.97, p < .001$, and comparison/standard ratio, $F(5, 70) = 119.05, p < .001$, but no standard duration by comparison/standard ratio interaction, $F(15, 210) = 1.28$. However, a further analysis suggested that the significant effect of standard duration came solely from the presence of data collected with the 200-ms standard. If this was omitted, and data from conditions with the 400, 600, and 800-ms standards were compared, there was no significant effect of standard value, $F(2, 28) = 0.25$, nor any significant interaction, $F(10, 140) = 1.09$, although there was, obviously, an effect of comparison/standard ratio, $F(5, 70) = 97.62$, $p < .001$.

Another way of identifying the source of deviation from superimposition in the data is to compare data from the 200-ms condition with those obtained with longer standards. One method is to construct a mean of the data from all the different comparison/standard ratios for a particular standard value, and compare the means using a paired-samples t test. When this is done, judgements made with a 200-ms standard produced more SAME responses than did those made with other standards, which, as the earlier ANOVA shows, did not differ amongst themselves: 200 ms vs. 400 ms, $t(14) = 6.27$; 200 ms vs. 600 ms, $t(14) = 4.56$; 200 ms vs. 800 ms, $t(14) = 6.45$; all $p < .001$.

Consider next data obtained using visual stimuli. Including six stimuli from each condition, there was a significant effect of standard duration, $F(3, 42) = 4.70$, $p < .01$, and comparison/standard ratio, $F(5, 70) = 84.72$, $p < .001$, but no significant standard duration by comparison/standard ratio interaction. However, omitting data collected with the 200-ms standard, but still using six stimuli, removed the significant effect of standard duration $F(2, 28) = 3.037$, $p = .06$, and there was in addition no standard value by comparison/standard ratio interaction, $F(10, 140) = 1.12$, but the effect of comparison/standard ratio remained, $F(5, 70) = 59.03$, $p < .001$. The same analysis as that used for the auditory stimuli showed that judgements made with a 200-ms standard involved more SAME responses than did those made with the 600-ms standard, $t(14) = 2.66$, $p = .02$, and with the 800-ms standard, $t(14) - 2.45$, $p < .03$, but did not differ from those made with the 400-ms standard, $t(14) = 1.49$.

Discussion

The results of Experiment 1 may be simply summarized. When people performed on an episodic temporal generalization task, with either auditory or visual stimuli of different durations, they produced orderly temporal generalization gradients, which peaked at or close to the comparison value that was equal to the standard. Superimposition of data from all four standard values (200, 400, 600, and 800 ms) failed for both auditory and visual stimuli (e.g., Figure 1), but the source of the failure in both cases was the presence of data from the 200-ms condition. When this was omitted, the temporal generalization gradients superimposed well.

It seems, therefore, that our episodic temporal generalization procedure produced data that obeyed SET predictions well, at least with standard values above 200 ms. The "standards" were not identified in any special way and varied unpredictably from trial to trial, so reliance on any kind of reference memory seems improbable. Thus, our data suggest that scalar timing (as evidenced by superimposition) can occur even in the likely absence of reference memory. This does not, of course, mean that some kind of reference memory cannot be an important source of scalar variance when it is formed, as in non-episodic temporal generalization, only that scalar timing can occur in its probable absence.

To confirm and extend the results of Experiment 1, Experiment 2 addressed the issue of whether similar scalar properties of timing could occur without reference memory by using an episodic version of another popular procedure, bisection.

EXPERIMENT 2

Experiments with humans using temporal bisection predate the development of SET (e.g., Bovet, 1968), but most recent work has employed analogues of an animal experiment developed by Church and DeLuty (1977). In that study, rats were trained in a two-lever operant chamber and were reinforced for responding on one lever after a short stimulus (e.g., 2 s long) had been presented, and on the other lever after a longer stimulus (e.g., 8 s long). With the time values cited, this discrimination is not difficult for rats, and after a few sessions of training the discrimination between the 2- and 8-s stimuli was highly accurate. Next, a range of stimulus values, including 2 and 8 s but with other values between these extremes, was presented, and the response to each stimulus was noted. During this test phase, no reinforcement was given. If we define a response on the lever appropriate to the 8-s stimulus as a *long* response, then rats showed an monotonically increasing proportion of *long* responses as stimulus duration increased, from near zero at 2 s to nearly 100% at 8 s.

Wearden (1991b) and, independently, Allan and Gibbon (1991) developed bisection methods for use with humans that were based on Church and Deluty's technique. Subjects were initially presented with examples of stimuli described as *short* or *long* standards (*S* and *L*), then were asked whether other stimuli (including *S* and *L* and values in between them) were more similar in duration (Wearden, 1991b) or were identical (Allan & Gibbon, 1991) to *S* or *L*. In these experiments, and in the others cited later, the methods used rapidly yielded very orderly data from human subjects, often with experimental sessions of 30 min or less.

A number of other articles have looked at issues in temporal bisection in humans which are not directly pertinent to the present work, but these will be referenced here for interested readers. For example, Wearden and Ferrara (1995) examined effects of stimulus spacing (between *S* and *L*), and Wearden and Ferrara (1996) examined effects of stimulus range (e.g., possible effects of the ratio of *L* and *S*). Wearden, Rogers et al. (1997) investigated the effects of using longer stimulus durations with humans (some of the same durations as used in animal studies), when counting was prevented by a concurrent task, and Penney, Allan, Meck, and Gibbon (1998) looked at stimulus modality (auditory or visual) effects on bisection. Wearden, Wearden et al. (1997) also used bisection in a study of timing in the elderly, Nichelli, Alway, and Grafman (1996) observed effects of cerebellar degeneration on bisection, and McCormack, Brown, Maylor, Darby, and Green (1999) studied bisection in both children and old people.

The focus of interest of the present article is the issue of where variance comes from in bisection procedures. As in the case of temporal generalization, most theoretical models have located this variance in the long-term reference memories of *S* and *L*. So, for example, the model proposed by Wearden (1991b) assumed that subjects developed representations of *S* and *L* in reference memory, and that on each trial the just-presented duration, *t*, was compared with samples drawn from the memory representations of *S* and *L*, by analogy with the temporal generalization model discussed earlier. The duration *t* was assumed to be timed without variance, and the sole source of variance in bisection was located in reference memories of *S* and *L*. Other articles (Allan & Gibbon, 1991; Wearden & Ferrara, 1995) have developed slightly different models of bisection performance, but all share the view that the source of variance in bisection performance from one trial to the next is the reference memory of some kind of standard duration.

Rodriguez-Girones and Kacelnik (1995, 1998) developed a bisection method (their *roving bisection*), which, like our Experiment 1, imposed conditions that made it unlikely that subjects were using any kind of reference memory as the basis for their bisection decisions. Each trial involved presentation of three stimuli. The first two were standards, and one of these was always longer (between 1.5 and 8 times longer) than the other. The third stimulus was, in most cases, some value between the short and long standards previously presented (although sometimes it lay outside their range), and the subject's task was to judge whether the third stimulus was more similar to the longer or shorter of the previously presented durations.

On this task, the durations presented from one trial to another varied randomly, so it is highly improbable that subjects would find that their performance was aided by attempting to develop any kind of reference memory. Unfortunately, the very fact of the random variability of stimulus durations from one trial to another made it impossible to test the scalar property of timing by simple analysis of obtained behaviour in Rodriguez-Girones and Kacelnik's (1995, 1998) procedure, although modelling conducted in their 1995 paper implied that the scalar property was present. In Experiment 2, we used a slight variant of their method, modified to allow the scalar property to be easily detected, if it is present. On each trial, three stimuli were presented, and subjects were required to judge, in effect, whether the third one was more similar to the shorter or longer of the previous two. On 40% of trials, the standard durations presented were randomly chosen (one always being four times the duration of the other one), but on the remaining 60% of trials, three different standard *S* and *L* stimuli were systematically compared with various comparison durations. As trials were randomly intermixed, it is unlikely that subjects would develop any kind of reference memory of standard durations, but the trial organization permitted a test of the scalar property of timing, as time values were systematically varied within the set of durations presented. Our Experiment 2 used *S*/*L* pairs of 100/400, 200/800, and 300/1,200 ms, interspersed with random *S*/*L* pairs. The basic question of interest was whether the scalar property of timing would be manifest in comparison of these three *S*/*L* pairs. If it was, then, as in Experiment 1, it would be highly unlikely that the source of scalar variance was reference memory, as developing any kind of reference memory would probably be unhelpful during our episodic bisection task.

Method

Subjects

A total of 20 Manchester University undergraduates served as subjects.

Apparatus

The apparatus was the same as that used in Experiment 1

Procedure

All subjects received a single experimental session lasting about 20 min. Only auditory stimuli (500-Hz tones produced by the computer speaker) were used. The experimental session comprised three blocks, each of 35 trials, with each trial in the block being different. After the response to a "Press spacebar for next trial prompt", the three stimuli on the trial were presented after a random delay ranging from 2,000 to 3,000 ms. The stimuli were presented with a random gap ranging from 500 to 1,000 ms between the first and second, and between second and third. The first two stimuli on the trial were standards, one of which was always four times the duration of the other one, and the third a "comparison". The subject's task was to judge whether the duration of the comparison was more similar to the first or the second stimulus presented, and no feedback was given. For seven trials in the block, one of the standards was 100 ms long and the other 400 ms long, and comparison stimulus durations were 100, 150, 200, 250, 300, 350, and 400 ms. For another seven trials the standards were 200 and 800 ms, and the comparison stimuli 200, 300, 400, 500, 600, 700, and 800 ms. For another seven trials, the standards were 300 and 1,200 ms, with the comparisons being 300, 450, 600, 750, 900, 1,050, and 1,200 ms. We refer to these trials as the 100/400, 200/800, and 300/1,200 trials, respectively. Data from these trials were stored and used to test superimposition. The other 14 trials of the block were composed as follows. A "short standard" value was chosen randomly from a uniform distribution running between 75 and 325 ms (thus overlapping with the short standards from the other conditions). The "long standard" was four times as long as the "short standard", and the comparison stimulus was randomly chosen from a uniform distribution running from the short to the long standards. Data from these "random" trials were discarded, as they could not be used to test superimposition. One of the first two stimuli on the trial was always four times the duration of the other one, and the order of the short and long stimuli was randomized from trial to trial. The 35 trials in the block (the three target conditions and the random trials) were randomly ordered for each block and for each subject. When the third (comparison) stimulus had been presented the subject had to indicate (by pressing the "1" or "2" key) whether they judged that the comparison was more similar in duration to the first or the second stimulus. This response was followed by the "Press spacebar to next trial" prompt. For data analysis purposes, these responses were transformed into judgements of perceived similarity of the third stimulus duration to the S and L stimuli on the trial.

Results

The upper panel of Figure 3 shows psychophysical functions in the form of the proportion of judgements that a comparison stimulus was more similar to L than to S (*long* responses) for the different S/L pairs, 100/400, 200/800, and 300/1,200 ms. The proportion of *long* responses increased from near zero when the comparison stimulus was S to 90% and above when the comparison was L, and the increase was nearly monotonic with increasing comparison stimulus duration in all conditions. The effect of comparison duration on proportion of *long* responses is obvious in Figure 3, but we also confirmed statistically that there was a significant effect for all three S/L pairs: 100/400, $F(6, 114) = 71.5$; 200/800, $F(6, 114) = 96.4$; 300/1,200, $F(6, 114) = 93.5$; all $p < .001$.

A measure of bisection performance that has attracted considerable interest both in studies with animals and in those with humans is the *bisection point*, the stimulus value giving rise to 50% *long* responses. Wearden and Ferrara (1995) showed that different methods of determining this bisection point usually lead to nearly identical results, and here the method we used derived the bisection point from each individual subject by interpolation by eye and averaged the values obtained. Resulting bisection point values are shown in Table 1 for the three different S/L pairs. Also shown are the arithmetic mean, $(S + L)/2$, and geometric mean (square root of $S \times L$) for the three pairs.

Comparison of the calculated bisection point values with the psychophysical curves shown in the upper part of Figure 3 suggests that the calculation represented the bisection point well. Inspection of bisection point values also shows that they changed systematically with respect to the arithmetic and geometric means of S and L with increases in the absolute values of S and L. For the 100/400-ms pair, the bisection point was very close to the arithmetic mean of S and

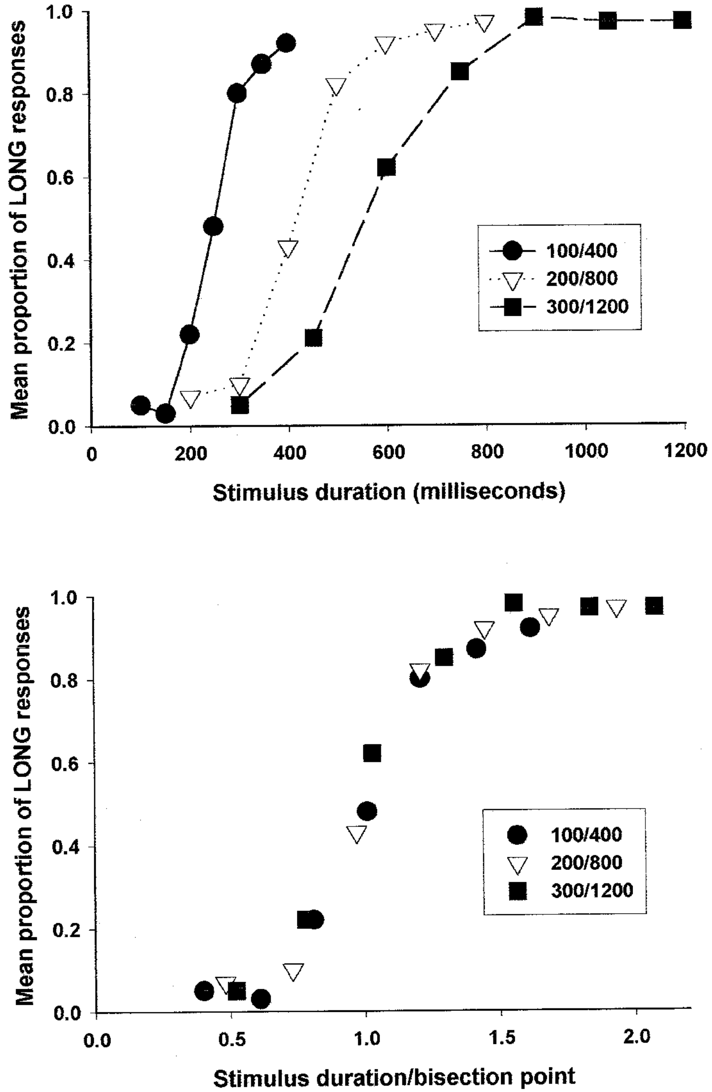


Figure 3. Upper panel: Psychophysical functions from Experiment 2. Mean proportion of LONG responses (i.e., judgements that the presented stimulus duration was more similar to the *long* run than to the *short* standard) plotted against stimulus duration. Data are shown separately for the 100/400, 200/800, and 300/1,200 ms standard conditions. Lower panel: Psychophysical functions from Experiment 2 plotted against stimulus duration, which has been divided by the bisection point calculated for each condition (see Table 1).

L (and therefore well above the geometric mean); for the 200/800-ms pair, the value was close to the geometric mean; for the 300/1,200-ms pair, the bisection point was below the S/L geometric mean. We compared the obtained bisection points statistically with both the arithmetic and geometric means of the appropriate conditions. For the 100/400-ms pair, the bisection point did not differ significantly from the arithmetic mean of S and L, $t(19) = 0.25$, but it was

TABLE 1
Short and long "standard" values and means from the three conditions of Experiment 2

S/L values	*BP*	*AM*	*GM*
100/400	251	250	200
200/800	434	500	400
300/1200	570	750	600

Note: S = short; L = long; BP = bisection point; AM = arithmetic mean; GM = geometric mean. All values in ms.

significantly lower than the geometric mean, $t(19) = 8.44, p < .001$. For the 200/800-ms pair, the bisection point was significantly lower than the *S/L* arithmetic mean, $t(19) = -4.2, p < .001$, and just significantly higher than the geometric mean, $t(19) = 2.20, p = .04$. For the 300/1,200-ms pair, the bisection point was significantly lower than the *S/L* arithmetic mean, $t(19) = -7.14, p < .001$, but did not differ significantly from the geometric mean, $t(19) = -1.19$. We discuss these effects in more detail later, but for present purposes the focus of interest is on a test of superimposition using the episodic bisection data.

According to Allan and Gibbon (1991) the appropriate method of testing superimposition in bisection is to plot the proportion of *long* responses against stimulus duration, where stimulus duration is expressed as a fraction of the bisection point obtained for the condition of interest. The bisection points shown in Table 1 were used for this purpose, and the resulting plot is shown in the lower panel of Figure 3. Obviously, superimposition was close to perfect when comparison values were "normalized" by the bisection points.

Discussion

Data from our episodic bisection study confirmed the general trends obtained in Experiment 1, using temporal generalization. The episodic procedure produced orderly psychophysical functions (upper panel of Figure 3), which superimposed almost perfectly when the bisection point rescaled comparison stimuli (lower panel of Figure 3). Thus, once again, it appears that the scalar property of timing can be manifest in a procedure where it is unlikely that any kind of reference memory of standards was formed. Indeed, such reference memory formation is particularly unlikely in Experiment 2, as 40% of trials were random ones where the *S/L* values were repeated only by chance. Presumably, this would make the detection of the 100/400, 200/800, and 300/1,200-ms standard pairs even less likely than would otherwise be the case.

The quality of the superimposition noted in the lower panel of Figure 3 appears, at least by eye, to be as good as that obtained from standard bisection where *S* and *L* standards were explicitly identified and, presumably, stored in reference memory. For example, compare data in the lower panel of Figure 3 with those in Allan and Gibbon (1991), Figures 5 and 9, (pp. 48 and 53), Wearden and Ferrara (1996, Figure 2, p. 31), and Wearden, Rogers et al. (1997, Figure 3, p. 88).

In spite of the fact that the psychophysical functions obtained from the episodic bisection procedure were orderly, and good superimposition was found, the data from the episodic bisection procedure were not exactly like those obtained from normal bisection. In particular, the location of the bisection point, and its apparent dependence on the absolute values *S* and *L* (with the bisection point being relatively smaller as *S* and *L* become absolutely larger) require some discussion.

Studies with animals, and some experiments with humans (e.g., Allan & Gibbon, 1991), find the bisection point at or near the geometric mean of *S* and *L*, whereas data from our laboratory with *S*/*L* ratios of 1:4, like the ones used in the present study, almost always find it to be near (more precisely, just below) the arithmetic mean (e.g., Wearden, 1991a; Wearden & Ferrara, 1995, 1996; Wearden, Denovan et al., 1997; Wearden, Rogers et al., 1997). The present data not only provide rare examples of bisection at or below geometric mean from our laboratory, but also show that the bisection point depends on the absolute values of *S* and *L*. Why does this effect occur? A possible, albeit speculative, explanation uses the idea of *subjective shortening*. This is an unusual form of forgetting found in memory for duration, which was popularized by Spetch and Wilkie (1983). In brief, subjective shortening suggests that as a duration memory ages, it does not degrade randomly but instead become progressively shorter. Although subjective shortening was originally proposed as an explanation of data from experiments with pigeons, Wearden and Ferrara (1993) demonstrated that subjective shortening also occurs in humans' memory for stimulus duration.

Consider a trial on the episodic bisection task as a list of three durations, S_1, S_2, and *t*, two standard durations (*S* and *L*), followed by a value that must be compared with them. The total duration of the trial events is determined by the gap between the consecutive stimuli, in our case an average of 750 ms, but also by the duration of the stimuli themselves. As S_1 and S_2, the standard values, become absolutely longer, then the time between S_1 and the comparison, *t*, increases, with the possibility of increased subjective shortening of S_1. If subjective shortening is absolute (i.e., a shortening that increases with increasing retention interval but that does not depend on stimulus length), then such shortening will have a particularly marked effect on trials where S_1 is *S*, the *short* standard. The effect of this may be to make the relatively unshortened comparison stimulus, *t*, more similar to the *long* standard than to the *short* one, thus biasing the response function towards responding *long*, and relatively reducing the bisection point.

More specifically, suppose that S_1, S_2, and *t* are the stimuli on the trial, and that their retention intervals are measured from their offset. Thus, at the end of *t*, this comparison stimulus is compared with S_1 and S_2, which will have subjectively shortened in the time since their offset, a time that is composed of the interstimulus intervals, the length of any intervening stimuli, and the length of *t* itself. Consider trials in which $S = 300$ ms and $L = 1{,}200$ ms, and *S* comes before *L* on the trial (as it does on 50% of the experimental trials). With the interstimulus intervals (average 750 ms) and *t* values used in Experiment 2, the average time since *S* will range from 3,000 to 3,900 ms, and the time from *L* from 1,050 to 1,950 ms. Thus, on trials where *S* precedes *L*, there is obviously more opportunity for subjective shortening of *S* than of *L*. When *L* precedes *S*, delays from *S* range from 1,050 to 1,950 ms, and those from *L* from 2,100 to 3,000 ms, so there will be more subjective shortening of *L* than *S*, but notice that the *difference* in times between the offsets of *S* and *L* and the putative bisection judgement is not the same for both trial types. When *S* precedes *L*, this difference is 1,950 ms in favour of *L*, but when *L* precedes *S* it is only 1,050 ms in favour of *S*, so when the two trial types are intermixed, subjective shortening will exert a more powerful effect on *S* than *L*, thus potentially making the unshortened comparison stimulus, *t*, more like *L* than *S* overall.

When the stimuli are shorter (e.g., 100 and 400 ms), the opportunity for subjective shortening of either stimulus will be lessened, and indeed it is even possible that subjective shortening only occurs when the retention interval exceeds the capacity of some storage mechanism, such as the phonological loop (Baddeley, 1986). Wearden and Culpin (1998) discuss some possible relations between subjective shortening of tone stimuli and the phonological loop, and it is perhaps relevant that when the standards on the trial are 100 and 400 ms, the maximum length of the events, S, L, and t on the trials (assuming the maximum t of 400 ms, and an average interstimulus interval of 750 ms) is only 2,400 ms, about the same duration as the estimated span of the phonological loop. If subjective shortening only operates in humans at longer retention intervals, then the 100/400-ms condition would be unaffected, although the 200/800 and 300/1,200-ms conditions would be biased towards *long* responses to some degree, as argued earlier.

This interpretation is speculative, but it at least consistent with previous work on retention of duration representations in humans (e.g., Wearden & Ferrara, 1993) and is offered here in the absence of any competitor. The account would predict that bisection performance would be different when S came first on the trial, rather than when L did but, unfortunately, although the order of S and L was randomized in Experiment 2, the order in which the two stimuli occurred on the trial was not stored. The account would, in addition, predict little or no effect of stimulus order in Experiment 1, as in the worst case the delay from offset of the first stimulus on the trial to offset of the second is 2,000 ms (1,400 ms for a comparison stimulus plus 600-ms interstimulus interval), so no subjective shortening would be found if this effect only occurs when the phonological loop time span is exceeded.

EXPERIMENT 3

Experiments 1 and 2 appeared to demonstrate good superimposition, except when the very shortest time values were employed in episodic temporal generalization, using procedures that were intended to avoid the formation of reference memory by frequently varying stimuli that might be used as standards. However, the possibility remains that the fact that certain stimuli were repeated from time to time in the experiment allowed subjects to encode these in reference memory (albeit unconsciously, as no subject reported even noticing that stimuli were repeated, let alone deliberately using them as standards). Can the formation of reference memory be absolutely prevented? Rodriguez-Girones and Kacelnik's (1995, 1998) roving bisection probably achieved this by never repeating stimuli, except presumably by chance, throughout their experiment. However, the avoidance of repetition means that the presence or absence of the scalar property could not be observed directly in their data and had to be inferred by fitting a theoretical model.

Experiment 3 used an episodic temporal generalization method that, likewise, never repeats stimulus durations, yet in a design that potentially allows the scalar property to be observed in data by testing superimposition. In outline the method is as follows. On each trial a *sample* stimulus duration is selected from a uniform distribution spanning one of three, equally likely, ranges: SHORT, 300–500 ms; MEDIUM, 450–750 ms; LONG, 600–1,000 ms. Thus all durations ranging from 300 to 1,000 ms can potentially occur as sample values. A *comparison* stimulus is also selected on each trial, and this is the sample duration, whatever it is on the trial, multiplied by 0.25, 0.5, 0.75, 1, 1.25, 1.5, and 1.75, with all values being equally likely. The order of presentation of sample and comparison is randomly varied between trials, and the two stimuli are presented with a random gap between them. The subject has to judge whether the stimuli have the same duration, and no feedback is given. Superimposition is tested by superimposing the generalization gradients from the three ranges, SHORT, MEDIUM, and LONG.

In Experiment 3, it is difficult to imagine how any reference memory could be developed by the subjects. For one thing, the same stimuli are never presented twice in the experiment, except by chance, and all sample durations between 300 and 1,000 ms could occur. For another, the average duration of all the stimuli presented is 600 ms, so even if a subject did average all these durations to produce some kind of reference (as proposed for bisection by Wearden & Ferrara, 1995, for example) it is unclear how this reference would aid task performance, particularly with samples from the SHORT and LONG ranges.

Method

Subjects

A total of 16 Manchester University psychology undergraduates participated.

Apparatus

The apparatus was the same as that used in Experiment 1.

Procedure

The experiment consisted of two conditions, auditory and visual, and all subjects were tested under both conditions (half with auditory first, half with visual first), in a single experimental session lasting about 30 min. The auditory and visual conditions were identical except for the stimuli timed, which were 500-Hz tones in the auditory condition and 14 × 14-cm light-blue squares in the visual condition. These were the same stimuli as those used in Experiment 1.

All trials consisted of the presentation of two stimuli, separated by a 400-600-ms gap (offset to onset). Following presentation of the second stimulus the subject was required to judge whether the two stimuli had the SAME duration, pressing the Y (YES) or N (NO) keys, but no feedback as to performance accuracy was given. On each trial, one of the two stimuli (the *sample*) was selected randomly from one of three equally likely duration ranges: SHORT, 300–500 ms; MEDIUM, 450-750 ms; LONG, 600–1,000 ms. The duration of the other stimulus presented on the trial (the *comparison*) was determined by the sample duration on the trial multiplied by one of the following seven equally likely values: 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75. The order of presentation of sample and comparison stimuli within the trial was randomly varied between trials. A block of trials thus consisted of 21 different trials: the seven multipliers (0.25 to 1.75) combined with the three duration ranges (SHORT, MEDIUM, LONG). Within each block, the order of the 21 trials was randomized for each subject and for each block. Four blocks of trials (84 trials) were given in total. All other procedural details were as for Experiment 1.

Results

Figure 4 shows the temporal generalization gradients from the auditory condition (upper panel) and visual condition (lower panel). Within each panel, the proportion of SAME responses is plotted against comparison/sample ratio. Within each panel, data points from the SHORT, MEDIUM, and LONG ranges are shown separately.

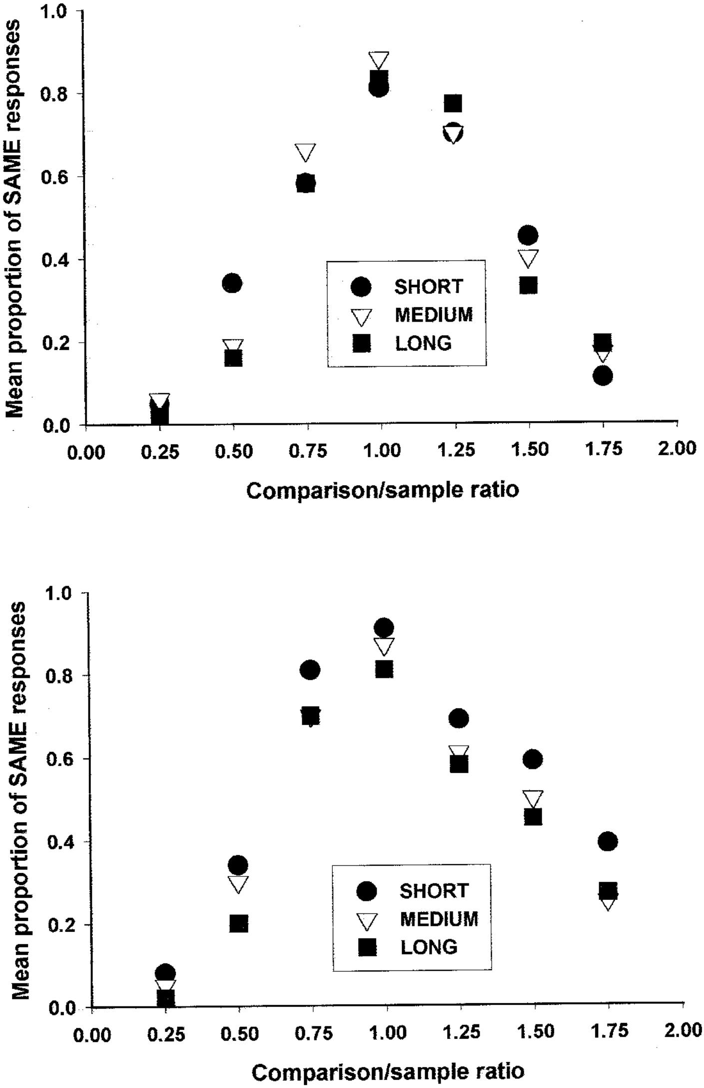


Figure 4. Temporal generalization gradients from Experiment 3. The proportion of SAME responses (identifications of the two stimulus durations as being the same) is plotted against comparison/sample ratio. Upper panel shows data from the auditory condition; lower panel shows data from the visual condition.

Inspection of the data in Figure 4 suggests several conclusions. First, generalization gradients from all conditions were orderly, peaking at the comparison/sample ratio of 1.0 (i.e., when the two stimuli actually had the same duration) and decreasing in an orderly way as the ratio deviated from 1.0 in both directions. As far as superimposition is concerned, inspection of data from the auditory condition suggests that superimposition was good, with data from the SHORT, MEDIUM, and LONG duration ranges overlapping. In the case of the visual stimuli, however, it appears that the SHORT duration range resulted in a higher proportion of SAME responses than did the other two ranges, which themselves seemed to superimpose.

These suggestions were confirmed by statistical analysis. Consider first data from the auditory condition. There was no significant effect of duration range, $F(2, 30) = 0.35$, which suggested superimposition, nor was there any duration range by comparison/sample ratio interaction, $F(12, 180) = 1.15$, but there was an effect of comparison/sample ratio, $F(6, 90) = 83.26, p < .001$, confirming the obvious change in proportion of SAME responses as this variable changes.

In the visual condition, there was a significant effect of duration range, $F(2, 30) = 6.63, p < .01$, and the usual effect of comparison/sample ratio, $F(6, 90) = 51.69, p < .001$, but no interaction between duration range and comparison/sample ratio, $F(12, 180) = 0.27$. Given the effect of duration range, simpler ANOVAs were conducted to identify the source of this effect. Apart from the effect of comparison/sample ratio, which was significant in all comparisons, then only significant effects were duration range effects in comparison of SHORT and MEDIUM ranges, $F(1, 15) = 9.50, p < .01$ and SHORT and LONG ranges, $F(1, 15) = 7.62, p < .01$, whereas the comparison between the MEDIUM and LONG ranges was not significant, $F(1, 15) = 1.72$. Another way of identifying the source of deviation from superimposition was to use the same t-test analysis for that in Experiment 1. When this was done, it was found that the SHORT duration range produce more SAME responses than did either the MEDIUM range, $t(14) = -3.08, p < .01$, or the LONG range, $t(14) = -2.76, p < .02$. These analyses confirm the earlier suggestion that the failure of superimposition in the visual condition was due to the fact that the SHORT duration range produced more SAME responses than did the other ranges, although the generalization gradient from this condition did not otherwise differ in shape from those obtained with the MEDIUM and LONG ranges, as interactions between range and time were never significant.

Discussion

Experiment 3 confirmed and extended the findings of Experiments 1 and 2. In an episodic temporal generalization procedure designed to prevent the formation of reference memory, superimposition was once again generally found. Data from the three duration ranges from the auditory condition superimposed well, as did data from the MEDIUM and LONG ranges from the visual condition. The sole failure of superimposition was the SHORT visual condition. However, even though performance in this condition differed significantly from that obtained with the other visual duration ranges, inspection of Figure 3, as well as the lack of interaction effects between duration range and time, suggests that behaviour in the SHORT visual condition differed only from behaviour in others only in having a higher proportion of SAME responses. Shape of generalization gradient and other features of data, such as the way that SAME responses changed with time, did not differ markedly from other conditions.

GENERAL DISCUSSION

Our three experiments taken together strongly suggest that the scalar property of timing, such a consistent feature of human timing in situations where people do not use chronometric counting or other uniquely human time-measuring devices (Allan, 1998), can be found when procedures discourage the formation of reference memories of standard durations. This conclusion holds over both temporal generalization (Experiments 1 and 3) and bisection (Experiment 2) tasks.

If the scalar property of timing is present in data obtained with our episodic procedures, where does it come from? One possible answer is that the source of scalar variability is in the internal clock itself. Although the pacemaker of the internal clock is usually supposed to be a Poisson pacemaker (Gibbon, 1992), which by itself would produce non-scalar variability in behaviour, Gibbon et al. (1984) discuss variants of Poisson timing (for example, varying average inter-pulse interval from one trial to another), which will produce scalar timing in "raw" time representations, so a pacemaker producing scalar properties is perfectly possible. The clock is not the only possible source of scalar variability; for example, the scalar property could arise somewhere in the decision process, the part of the scalar timing system that remains the least explored (Wearden, 1999). However, it is perhaps easier to conceive of simple ways in which the scalar property could arise from the variability of the "ticking" of the internal clock than by other means (see, however, Killeen & Taylor, 2000, for discussion of how variance can arise in the accumulator processes of internal clocks). Furthermore, consideration of how the internal clock might operate may also address other features of our data, as shown later.

In our Experiments 1 and 3, superimposition sometimes failed with the shortest durations used. However, it has long been known that scalar-type timing can fail at very short durations. Friberg and Sundberg (1995) provide clear examples from a very different task from those employed here, that of discrimination of "isochronous sequences". A series of periodic clicks is presented, and occasionally one of these clicks is slightly displaced in time relative to the others. The subject's task is to detect this "deviant" click. The just-noticeable difference (JND) for this task was expressed as a percentage of the inter-click interval, and many studies reviewed showed scalar-type timing (i.e., constant percentage JND) at inter-click intervals between 200–300 ms and 1,000 ms, but increasing JND at the shortest values used, indicating higher relative variability for the judgements of the shortest durations.

If internal clock processes are an important source of variance in human timing, then such a result, and the general failure of scalar timing at very short durations, is probably only to be expected. According to the concept of the pacemaker–accumulator internal clock that underlies SET, the onset of a stimulus to be timed closes a switch connecting the pacemaker and accumulator, and subsequent offset of the stimulus opens the switch, cutting the pacemaker–accumulator connection. The switch that controls the start and end of timing need not operate either instantaneously with the physical onset and offset of the to-be-timed stimulus, or with zero variance from trial to trial, and some arguments have implicated switch processes as a potentially important source of variance in cross-modal timing (e.g., Wearden et al., 1998). More generally, the starting and stopping of some interval to be timed may be associated with some absolute variance, independent of the interval duration (the variance of which may change relative to its duration, via the medium of a Poisson or scalar pacemaker). If there is some absolute start/stop variance, then this will surefly make a larger relative contribution at shorter intervals than at longer ones, eventually coming to dominate sources of variance at very short intervals. When relative measures of variability, such as the coefficient of variation, are calculated, this absolute start/stop variance may inflate relative variability estimates of short durations considerably by comparison with longer ones, where the main source of variance comes from pacemaker processes or is derived from "later" parts of the SET system, such as reference memory. The "failure" of SET predictions like superimposition at very short durations therefore seems, to us at least, theoretically untroubling and, in fact, completely consistent with the conception of the pacemaker–accumulator internal clock as proposed by SET. A problem is to identify where "very short" intervals begin and end, and the possibility that very short intervals are timed by different mechanisms from longer ones remains viable (e.g., see Rammsayer, 1997), but in general such deviations from scalar timing at short durations may do as much to confirm SET as a robust model of many aspects of human timing as to challenge it.

In conclusion, our three experiments taken together suggest that a close approximation to scalar timing can be obtained in both temporal generalization and bisection procedures, except at the very shortest intervals used, when "episodic" variants of the tasks, which appear to discourage, or even prevent, the use of reference memory, are employed. We conjecture that the source of scalar timing in these procedures is the internal clock itself.

REFERENCES

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

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

Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.

Bovet, P. (1968). Echelles subjectives de durée obtenues par une méthode de bisection. *L'Année Psychologique, 68,* 23–36.

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.

Ferrara, A., Lejeune, H., & Wearden, J.H. (1997). Changing sensitivity to duration in human scalar timing: An experiment, a review, and some possible explanations. *Quarterly Journal of Experimental Psychology, 50B,* 217–237.

Friberg, A., & Sundberg, J. (1995). Time discrimination in monotonic, isochronous sequences. *Journal of the Acoustical Society of America, 98,* 2524–2531.

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

Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation, 22,* 3–38.

Gibbon, J. (1992). Ubiquity of scalar timing with a Poisson clock. *Journal of Mathematical Psychology, 36,* 283–293.

Gibbon, J., & Church, R.M. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 87–108.

Gibbon, J., Church, R.M., & Meck, W. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences, 423: Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.

Killeen, P.R., & Taylor, T.J. (2000). How the propagation of error through stochastic counters affects time discrimination and other psychophysical judgements. *Psychological Review, 107,* 430–459.

Lejeune, H., & Wearden, J.H. (1991). The comparative psychology of fixed-interval responding: Some quantitative analyses. *Learning and Motivation, 22,* 84–111.

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

Nichelli, P., Alway, D., & Grafman, J. (1996). Perceptual timing in cerebellar degeneration. *Neuropsychologia, 34,* 863–871.

Penney, T.B., Allan, L.G., Meck, W.H., & Gibbon, J. (1998). Memory mixing in duration bisection. In D.A. Rosenbaum & C.E. Collyer (Eds.), *Timing of behaviour: Neural, psychological, and computational perspectives* (pp. 165–193). Cambridge, MA: MIT Presss.

Penton-Voak, I.S., Edwards, H., Percival, A., & Wearden, J.H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 307–320.

Rammsayer, T.H. (1997). Are there dissociable roles of the mesostriatal and mesolimbocortical dopamine systems on temporal information processing in humans? *Neuropsychobiology, 35,* 36–45.

Rodriguez-Girones, M.A., & Kacelnik, A. (1995). Interval bisection with and without reference memory. In M. Richelle, V. DeKeyser, G. d'Ydewalle, & A. Vandierendonck (Eds.), *Time and the dynamic control of behavior* (pp. 47–72). Liège: Université de Liège P.A.I.

Rodriguez-Girones, M.A., & Kacelnik, A. (1998). Response latencies in temporal bisection: Implications for timing models. In V. DeKeyser, G. d'Ydewalle, & A. Vandierendonck (Eds.), *Time and the dynamic control of behavior* (pp. 51–70). Gottingen: Hogrefe and Huber.

Spetch, M.L., & Wilkie, D.M. (1983). Subjective shortening: A model of pigeons' memory for event duration. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 14–30.

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

Wearden, J.H., & Culpin, V. (1998). Exploring scalar timing theory with human subjects. In V. DeKeyser, G. d'Ydewalle, & A. Vandierendonck (Eds.), *Time and the dynamic control of behavior* (pp. 33–49). Gottingen: Hogrefe and Huber.

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

Wearden, J.H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights". Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology, 51B,* 97–120.

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., & 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., Rogers, P., & Thomas, R. (1997). Temporal bisection in humans with longer stimulus durations. *Quarterly Journal of Experimental Psychology, 50B,* 79–94.

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

Manuscript received 21 April 1999
Accepted revision received 5 November 2000