# The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

## Is the growth of subjective time in humans a linear or nonlinear function of real time?

Online Publication Date: 01 September 2007
To cite this Article: Wearden, J.H. and Jones, Luke A. (2007) 'Is the growth of subjective time in humans a linear or nonlinear function of real time?', The Quarterly Journal of Experimental Psychology, 60:9, 1289 - 1302
To link to this article: DOI: 10.1080/17470210600971576
URL: http://dx.doi.org/10.1080/17470210600971576

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2007, 60 (9), 1289–1302

Psychology Press
Taylor & Francis Group

# Is the growth of subjective time in humans a linear or nonlinear function of real time?

**J. H. Wearden**
*Keele University, Keele, UK*

**Luke A. Jones**
*University of Manchester, Manchester, UK*

The difficulties of deciding whether subjective time grows as a linear or nonlinear function of real time are discussed, and two experiments are presented to address this question. In Experiment 1, people received a 10-s standard duration and then had to judge what proportion other durations (ranging from 1 to 10 s) were of the standard. Counting was prevented by a concurrent task. The relation between judged and actual proportions was linear. In Experiment 2, people were required to average together three tone durations (mean duration 600 ms) and to judge whether subsequently presented comparisons were or were not the average. The spacing of the tone durations had no effect on judgements, suggesting a linear underlying time scale.

What is the relation between the growth of real, physical, time (as measured by clocks and other manufactured devices) and the internal, subjective, time that passes for people and animals? The question is easily posed, although much less easily answered, but Figure 1 shows three general potential solutions.

In the upper panel of Figure 1, three possibilities are shown. In one, subjective time grows as a convex function of real time, thus subjective time initially passes quickly, but increases more slowly, so that longer real-time values produce smaller increases in subjective time: Some kind of logarithmic function would be an example of this sort. Another curved line (concave) shows the opposite trend: As real time elapses, subjective time grows initially very slowly, but then at an increasing rate: Some type of power function with an exponent greater than 1.0 might produce such an effect. The third possibility shown is linear, where subjective time grows as a linear function of real time. If we indicate subjective time by $t$ and real time by $T$, then the linear function is just

$$t = bT + a,$$

where $b$ and $a$ are constants representing the slope and intercept of the linear function, respectively. The line shown in Figure 1 is both linear and accurate, so $b = 1.0$ and $a = 0$. In this case, subjective time is exactly equal to real time, but the linear case with $b$ and $a$ having other values is more general.

---

Correspondence should be addressed to J. H. Wearden, School of Psychology, Dorothy Hodgkin Building, Keele University, Keele, Staffordshire, ST5 5BG, UK. E-mail: j.h.wearden@psy.keele.ac.uk

Experiment 2 was conducted by the second author as partial fulfilment of the requirements for the degree of PhD awarded by the University of Manchester.

http://www.psypress.com/qjep

DOI:10.1080/17470210600971576

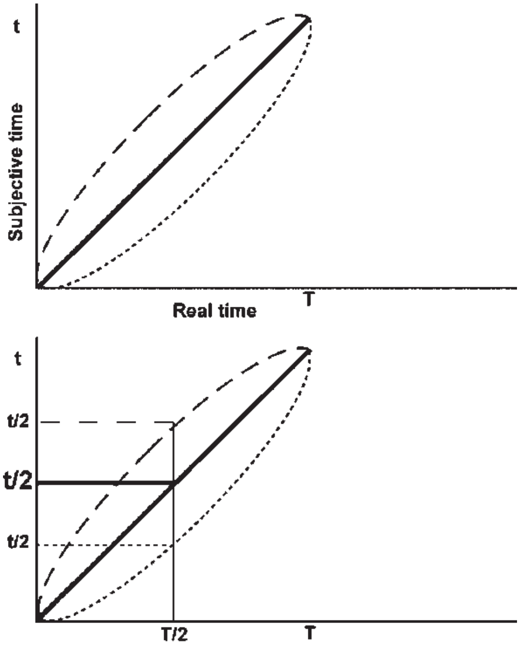

**Figure 1.** *Upper panel: Three potential functions for the growth of subjective time (vertical axis), as a function of real time (horizontal axis). Functions shown are linear (solid line), convex (dashed line), and concave (dotted line). Lower panel: Values of subjective time (t/2) corresponding to half the real time (T/2) for the three functions shown in the upper panel.*

A reader might think that the question of the relation between subjective and real time is easily solved: We just take some measures of judgements based on subjective time, plot them against real time, and observe the resulting function. Many data from recent timing experiments seem suitable for this. For example, Wearden and McShane (1988) asked students to produce time intervals of from 500 ms to 1,300 ms (intervals too short to make chronometric counting useful) and gave accurate feedback after each production. The mean time produced (a putative measure of subjective time, *t*) was almost exactly equal to the target time (*T*). Another example comes from the timing of stimuli. Wearden, Denovan, Fakhri, and Haworth (1997) used a temporal generalization technique. In their experiment, people had to identify a standard duration (which was 2, 4, 6, or 8 s, in different groups, with counting prevented by a secondary task) on the basis of feedback that was given, and then had to decide whether other durations were or were not the standard. The proportion of identifications of a stimulus as the standard was plotted against stimulus duration, and the resulting function (a temporal generalization gradient) was found to peak at the standard value, whatever it was for the group. So, 2 s was identified maximally as the standard in the 2-s group, and so on.

The two examples seem to prove that subjective time appreciates not only linearly but accurately: For example, if people are trained to produce 700 ms, the mean time that they produce is almost exactly 700 ms and not some shorter or longer value; if they are trained with 4 s as a standard in temporal generalization, they maximally identify this duration as the standard and not some shorter or longer value. The answer to the question posed in the first sentence of this article seems clear. Indeed, it is true that data from Wearden and McShane (1988) and Wearden et al. (1997) are consistent with people having a linear and accurate subjective time scale but, unfortunately, the data do not prove that people have such a scale. In fact, behaviour like that observed in Wearden and McShane's (1988) experiment is consistent with any underlying relation between subjective and real time and tells us nothing about how internal time passes.

This conclusion may seem strange: Surely, if timed behaviour varies as a linear function of real time, then subjective time must also appreciate linearly? This conjecture, although reasonable, is completely false. All that is needed to perform as participants do in the two experiments quoted is that they possess a set of subjective internal states, $s_1$, $s_2$, … $s_n$, each member of which corresponds to a real time $T_1$, $T_2$, … $T_{\mathrm{n}}$, and that the subjective internal states can be reliably distinguished. To perform, the participant needs only to learn to respond in the presence of some state $s_x$ when the real-time requirement (or real-time stimulus duration) will be $T_x$. The relations between the various *s* states and their corresponding real-time *T* values could be anything: It is even possible, although unlikely, that $s_a < s_b$ when $T_a > T_b$ (i.e., that the relations between the *s* states is not even ordinal so, e.g., 400 ms could "feel longer" than 800 ms).

To illustrate just how nonlinear an underlying time representation might be, we might consider a type of representation of duration instantiated in a "picture clock". This imaginary device is like a digital watch-face, but with pictures instead of numbers. For simplicity, consider that the picture clock times only in seconds, so one picture (a dog) is present for the first second, then another (a book) for another second, and so on, with different seconds being represented by different pictures. A person could learn to produce any arbitrary duration just by learning to respond when some particular picture was visible and could produce temporal generalization gradients that peak at arbitrary standards merely by identifying the presented stimulus as the standard when a particular picture is shown. If, in the picture clock, the different pictures represent the internal "time" states then it is obvious that people can learn to produce time intervals with average accuracy even when the relation between the different internal states corresponding to different real times is completely arbitrary.

This reasoning leads to the conclusion that data like those of Wearden and McShane (1988) and Wearden et al. (1997), while consistent with linear time growth, do not in any way conclusively support this notion. However, the converse is also true: Data exhibiting nonlinear relations between measures of behaviour and imposed time constraints are likewise inconclusive. Wearden (2002) discusses some of these in detail.

Is the task of deciding whether subjective time grows linearly or nonlinearly with real time impossible to accomplish? Probably not. The lower panel of Figure 1 shows a potential method of attack, which involves comparing the whole of a duration with parts of the same duration. For example, suppose that time grows linearly and accurately, and consider the timing of some real-time interval, $T$. In this case, when the real time is $T/2$, half the subjective time has elapsed, whereas with the convex function more than half the subjective time has elapsed at $T/2$, and with the concave one less than half of the subjective time has elapsed. The same arguments apply to other fractions of $T$, so if we ask people to make judgements that relate parts of an interval to the whole interval, then the nature of the underlying time scale might be revealed.

Comparing the whole of one interval with parts of another one is the basis of the controversial "time-left" procedure, developed by Gibbon and Church (1981) as a way of distinguishing between linear and logarithmic time growth in animals. To simplify, the time-left procedure offers animals a choice between a fixed time to food ($S$) and a constantly diminishing time to food ($C$). For example, suppose $S$ is 30 s, and $C$ is initially 60 s. In this case, animals should initially prefer $S$, but as the trial elapses, $C$ decreases, so there comes a point where the "time left" in $C$ is less than $S$, so preference should shift.

Gibbon and Church (1981) did find systematic shift of preference for $C$ as the time in the trial elapsed, but they reported a "bias" towards the time-left alternative, $C$, so the elapsed time in the trial when $S$ and $C$ were equally preferred was not when the time left in $C$ and $S$ were equal. For example, with $C = 60$ s, and $S = 30$ s, the animals were indifferent between the two at a time less than 30 s into the trial. However, the predicted behavioural difference between linear and logarithmic growth of time lies not in the exact time value when the animal is indifferent between $S$ and $C$, but in the fact that this point of indifference varies with the absolute values of $S$ and $C$, as it did in Gibbon and Church's study, thus supporting linear time growth.

Wearden (2002) developed an analogue for humans of the time-left procedure, based on imaginary journeys by train. In one condition, a participant could receive three sorts of trial. On one (a "normal" train trial), the journey took 12 s; on another (a "special" train trial), the journey only took 6 s. In the remaining trials, from which the data were collected, the journey started on the normal train, but after some time had elapsed the "special" train was offered, and the participant could accept it or reject it, but was instructed to try to minimize journey time.

If the "special" train was accepted, the rest of the journey took 6 s; if it was rejected, the rest of the journey took the "time left" on the "normal" train. The choice point was varied over a range of 1 to 11 s, and the choice at each time was measured. People showed an increasing tendency to reject the "special" train the later it was offered (i.e., when the "time left" was shorter) and were indifferent at roughly the time at which the "special" train duration and the "time left" were equal.

Such behaviour is consistent with an underlying linear time scale, as the whole of the "special" train interval was judged as equal to the "time left" when they were in fact physically equal, suggesting that time in the trial was elapsing linearly. Unfortunately, interpretation of the time-left method with animals is complicated by a number of procedural and theoretical considerations (see Cerutti & Staddon, 2004; Dehaene, 2001; Gallistel, 1999; Preston, 1994; Staddon & Higa, 1999), which make conclusions as to the underlying time scale used difficult to draw. Many of these apply with less force to the human time-left analogue (Wearden, 2002) than to the procedure used with animals, but one difficulty remains.

In the human time-left analogue, people were never told whether or not their choice was the correct one (i.e., whether it produced the shortest journey time on the trial). If such feedback had been provided, then people could learn (perhaps in a single trial), that a certain choice is "correct" (i.e., minimizes journey time) at some elapsed time, $t_a$, and that another choice is correct at some other elapsed time, $t_b$. Elapsed time in the trial alone could then control judgements, without any need for the "time left" to be used. The task would therefore become like the production or generalization procedures described above and could be performed even if the relations between the different *t*s were arbitrary.

Although no feedback was used in the human time-left analogue, S. Dehaene (personal communication, January 2002) correctly pointed out that people actually experienced the consequences of their choice, so one choice led to a longer trial duration, and overall session duration, than another. The time-left analogue was boring even by the standards of laboratory experiments on timing, so participants were presumably highly motivated to reduce session length and thus possibly sensitive to the consequences of their choice. This potential "self-reinforcement" means that participants might learn to use elapsed time in the trial as a cue for response choice, even without explicit feedback.

One way of overcoming this problem is to develop a consequence-free version of the human time-left, where the trial ends once the participant has made the choice. However, a simpler way of addressing relations between parts of an interval and the whole of another one, and a method that avoids many of the complexities of the time-left procedure, was developed in Experiment 1.

## EXPERIMENT 1

The basic principle of Experiment 1 was that participants initially learned a standard duration and then had to make estimates of the percentage that other durations were of this standard. Experiment 1 consisted of two subexperiments (Experiments 1a and 1b), which were procedurally almost identical. The procedure for Experiment 1a was as follows. The participant initially received presentations of a standard duration, which was actually 10 s, although the participant was not told this. Counting was prevented by a concurrent digit-shadowing task (e.g., as in Wearden, 2002, and Wearden et al., 1997). The initial presentations of the standard were followed by presentation of comparison durations (which actually ranged from 1 to 10 s and were presented in a random order). When the comparison duration terminated, the participant estimated the percentage that the comparison was of the standard, using a scale from 0 to 100. Occasional trials, interspersed among the comparisons, re-presented the standard.

Experiment 1b, which was actually conducted first, was identical to Experiment 1a except that, as a result of a programming error, all durations were 500 ms shorter than intended. So, the standard was 9.5 s, and all comparisons were 0.5 s shorter than those in Experiment 1a. Given that no numerical information about the durations was ever given to participants, the programming error did not actually change the experimental situation markedly by comparison to Experiment 1a, so in fact Experiment 1b represents a replication of Experiment 1a with slightly different duration values.

The focus of Experiment 1 is obvious. If people are required to judge what fraction of some standard duration has elapsed is the resulting function linear or nonlinear? The two different forms of nonlinear time growth shown in Figure 1 (concave and convex) might be expected to manifest themselves in the data of Experiment 1, if present.

Experiment 1 avoids some problems associated with the human time-left procedure (Wearden, 2002). For one thing, the instructions were easy for people to understand, unlike those of the time-left method, which some participants apparently never mastered. For another, the verbal responses (judgements of the percentage of the standard elapsed after each comparison) had no consequences. No feedback was given, nor did one response or another shorten or lengthen the trial, or the session.

## Method

### *Participants*

A total of 15 Manchester University psychology undergraduates served in Experiment 1a and a different 17 in Experiment 1b.

### *Apparatus*

An Opus (IBM-compatible) computer with a colour screen controlled all experimental events. The computer keyboard was used for registering responses. The experimental programs were written in the MEL (Micro Experimental Laboratory: Psychology Software Tools, Inc.), which assured millisecond accuracy for timing of stimuli and responses. Stimuli were defined by tones produced by the computer speaker.

### *Procedure*

*Experiment 1a*. Participants received a single experimental session lasting about 30 minutes. All stimulus durations were defined by the interval between two 1,000-Hz tones, 50 ms long, and durations were timed offset to onset. While all durations were elapsing, participants were required to verbalize aloud two-digit numbers presented in the centre of the computer screen. Digits were randomly chosen from the range 10–99 and were displayed for 150 ms. The times (offset to onset) between random digit displays were drawn from two equally likely uniform distributions, one running from 150 to 850 ms, the other from 750 to 1,250 ms. Participants' verbalizations were recorded, and all were found to shadow the digits as instructed. The session began with three presentations of the standard (10-s) duration, which was identified as the standard, although its actual duration was not given. Following this, participants received five blocks of 13 trials, with the 13 trials in each block arranged in a random order, which differed between participants and blocks. Of the 13 trials, 3 presented durations 10 s long, which on termination were followed by the display "That was the standard time". The other 10 durations were 1, 2, 3 . . . 10 s, and after termination the participants were prompted to type in the percentage estimate, using the numeric keypad of the keyboard, by the display: "Estimate the percentage of time elapsed. Values from 0–100." The initial standard-duration presentations and presentations of all test durations were cued by a "Press spacebar for next trial" prompt, and this response was followed by the tone that started the duration to be presented, after a delay drawn from a uniform distribution running from 2,000 to 4,000 ms.

*Experiment 1b*. This was identical in all respects except that all durations were 0.5 s shorter than intended, as the result of a programming error. Thus the standard was 9.5 s long, and test durations were 0.5, 1.5, 2.5 . . . 9.5 s.

## Results

Figure 2 shows the mean estimated elapsed percentage of the standard plotted against the actual

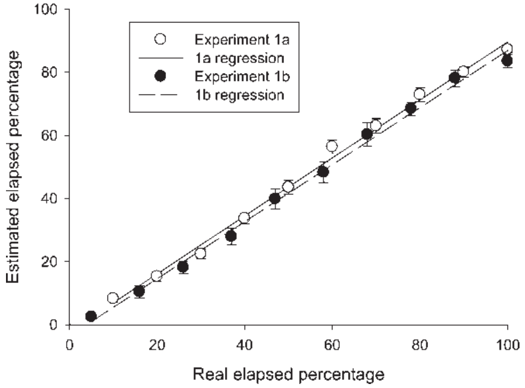

**Figure 2.** *Data from Experiment 1a (open circles) and Experiment 1b (filled circles). Values shown are mean percentages that presented durations (1 to 10 s, in Experiment 1a; 0.5 to 9.5 s in Experiment 1b) were of some standard (10 s or 9.5 s for the two subexperiments, respectively), plotted against the actual percentage. Standard errors of the means are shown as vertical bars, and the best fitting linear regression lines for data from the two experiments are shown.*

elapsed percentage from Experiment 1a (open circles) and Experiment 1b (filled circles). Standard errors of the mean are also shown but these are usually too small to be visible in the figure. Data from one participant in Experiment 1a were lost because of a storage error.

Inspection of the estimated versus real percentage functions from both experiments suggests a very strong linear relation. Linear regression of mean estimate against real percentage found significant effects ($r^2 = .995$ in Experiment 1a and .998 in Experiment 1b, $p < .001$). Linear regression slopes were .919 and .945 for the two subexperiments, respectively, and both slopes were significantly different from zero; the intercepts were $-1.65$ and $-3.15$, respectively, and the second one was significantly different from zero. The regression was repeated with the actual elapsed time being log-transformed, but this regression produced smaller $r^2$ values (.845 and .901) than those for the normal linear regression.

Consistent with the results of the linear regression, analysis of variance (ANOVA) of estimated percentages against real percentage revealed highly significant effects of real percentage, $F(9, 117) = 305.53$, $p < .001$, for Experiment 1a, $F(9, 144) = 257.02$, $p < .001$, for Experiment 1b.

## Discussion

Data from both subexperiments of Experiment 1 show that average estimates of the percentage that some stimulus duration is judged to be of a standard duration is a linear function of the real percentage. Our data show no evidence whatsoever of any sort of nonlinear curvature (see Figure 2). The results support the contention that subjective time grows as a linear, rather than nonlinear, function of real time, at least over the durations used. This result is consistent with data from the human time-left study (Wearden, 2002), but cannot be attributed to any possible differential consequences of making one estimate rather than another.

The method used in Experiment 1 bears some relation to the ratio-setting method popular in time perception studies during the last century (see Allan, 1978, for discussion). In a typical study (Allan, 1978) people receive two successive stimulus durations; the first is the standard, and the second a comparison the duration of which can be adjusted by the participant. The participant's task is to set the comparison as some fraction of the standard—for example, as equal to the standard, half of its duration, or twice its duration. Our procedure does, however, differ from ratio setting in a number of ways. For one thing, the "standard" duration is not presented on each trial, but is presumed to reside in memory. For another, the comparison stimulus has a fixed length, and verbal estimation, rather than adjustment, is used to judge the comparison's relation to the standard. A final difference is that our method uses more comparison values (10) rather than the number usually used in ratio-setting studies (e.g., Allan, 1978). Nevertheless, in spite of many differences in method, our results supporting a linear time scale are in good agreement with those obtained earlier. For example, Allan found that three of her participants "do remarkably well in setting durations which bear the appropriate relation to the standard duration" (p. 449).

Although at first sight results from our Experiment 1 seem to support a linear time scale overwhelmingly, there is a potential objection to them. This derives from the use of a verbal estimation method, where numerical values are assigned to subjective quantities. Some results on the judgments of number have indicated that, although the number scale might be linear for arithmetic purposes (i.e., a person knows that the difference between 20 and 10 is the same arithmetically as the difference between 90 and 80), it is not linear in terms of psychological scaling. Buckley and Gillman (1974) proposed a logarithmic compression of the psychological "number line", and Brysbaert (1995) discusses more recent evidence relating to this idea. One of the most compelling results supporting logarithmic compression of numbers is that the time to make judgements about numbers is best predicted by the logarithm of the number value (e.g., Brysbaert, 1995). Although this might suggest a logarithmically compressed number line, other workers prefer to account for these effects using a linear number line but one that has scalar variance (i.e., a standard deviation that is proportional to the mean). Thus decision times vary with number size not because of logarithmic compression but because larger numbers are more "fuzzily" represented than shorter numbers. Brannon, Wusthoff, Gallistel, and Gibbon (2001) discuss attempts to distinguish between these interpretations.

The assertion that the number line might be logarithmically compressed, although not uncontested, raises the possibility that the relation between the judged and actual proportions found in Experiment 1 arises because both time and number are logarithmically compressed. To produce the behaviour shown in Figure 2, the correspondence between potential curvature of the time scale and that of the number line must be very close over the time (1–10 s) and number (0–100) ranges studied in Experiment 1, which may seem a very unlikely coincidence. However, the persisting possibility that the number scale may not be subjectively linear means that attempts to decide whether the time scale is linear or not should not depend exclusively on methods using verbal estimation. Accordingly, Experiment 2 addressed the problem of linear versus nonlinear time growth in a different way, using a technique of temporal generalization (Wearden, 1992) that does not depend on participants assigning verbal labels to stimulus durations.

## EXPERIMENT 2

Suppose that a person is posed the following problem: They receive three different stimulus durations and then are required to compare subsequently presented durations with the mean of the three previously presented. Can people do this at all, and what does their behaviour look like? Suppose that the averaging process can be performed (and some accounts of duration discrimination in humans assume that it can, see Wearden & Ferrara, 1995, for example). If time elapses linearly and accurately, then the three different duration values will be able to be averaged to their arithmetic mean, if a veridical averaging process could be performed. If, on the other hand, time elapsed in some radically nonlinear way, then the average of three different time values could not be located at their arithmetic mean, even if the "raw" duration values could be averaged together. Thus, evidence that three time values can be veridically averaged together implies not only an accurate averaging process, but also linear and accurate representations of the underlying durations. Of course, if results suggested that performance was not based on an arithmetically accurate average, then this could be because (a) time elapses in a nonlinear way even though the averaging process is arithmetic, (b) time elapses linearly but the averaging process is nonarithmetic, or (c) both the growth of subjective time and the averaging process are nonarithmetic.

Experiment 2 examined performance on a temporal averaging task by using two subexperiments that were procedurally almost identical. Both used the "changing standard" variant of temporal generalization (Jones & Wearden, 2003, 2004). Changing standard temporal generalization procedures are arranged in blocks. At the start of each block, a number of examples of a standard duration (e.g., a tone *x* ms long) is presented.

Then, the participant receives a number of comparison stimuli, presented in a random order, with durations that are multiples of *x* (e.g., 0.2*x*, 0.6*x*, 0.8*x*, 1.0*x*, 1.2*x*, etc.). After each comparison has been presented, the participant must judge whether or not it had the standard duration, and feedback may or may not be given. Then another standard value, *y* ms, is chosen, and then comparisons that are multiples of *y* are presented, and so on. Thus the standard in force changes for each block of trials and is valid for only one block of comparison stimuli, with the participants being informed before the experiment begins that this is the case.

Experiments 2a and 2b used this method, but with the difference that the three tone presentations at the start of each block had to be averaged together to produce a "standard" for the block. In Experiment 2a one of the durations was actually exactly equal to the mean of all three, but in Experiment 2b none of the durations presented actually was the mean value. The different conditions of each subexperiment also differed in terms of the spacing of the three durations presented at the start of each block. In Experiment 2a, one condition involved the three stimuli at the start of the block actually being the *same*, and in two other conditions the temporal spacing between the stimuli was varied between a *near* and a *far* spacing. In Experiment 2a, a stimulus that had a duration that was actually equal to the average duration was presented at the start of each block either once (near and far spacing) or three times (same spacing). Experiment 2b used an almost identical procedure, except that none of the three stimuli presented at the start of each block had a duration that was exactly equal to the average, except by chance. Experiment 2b also varied the spacing of the stimuli that had to be averaged together over three values (*very close*, *close*, and *far*).

## Method

### *Participants*

A total of 16 Manchester University undergraduates participated. Participants served in both Experiments 2a and 2b, with 8 receiving Experiment 2a first and 8 receiving Experiment 2b first. Both experiments were presented in a single session lasting approximately 20 minutes.

### *Apparatus*

The apparatus was the same as that in Experiment 1.

### *Procedure*

*Experiment 2a*. All the stimuli used in the Experiment were 500-Hz tones generated by the computer speaker. The experimental procedure consisted of three conditions (same, near, and far) with five blocks in each condition and nine trials in each block. The order of the different types of blocks was randomized for each participant.

**Condition 1 (*same* condition)**. Participants were first given task instructions presented on the computer screen, which instructed them that they should use the average of the durations presented at the beginning of each block to make subsequent comparisons. The participant received a "press space bar" prompt, and this response was followed by a delay drawn from a uniform distribution running from 1,500 to 2,000 ms; then the standard duration in the form of a 500-Hz tone was presented three times, with a random interval drawn from a uniform distribution running from 1,500 to 2,000 ms between each presentation. The standard duration was chosen randomly from a uniform distribution running from 500–700 ms, and the standard value chosen differed for each block of trials. Following standard presentation, the participant received comparison tones for which the duration was the standard multiplied by 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, or 1.6. with the order randomized for each block. The standard duration (i.e., multiple = 1.0) was presented three times as a comparison duration, thus making nine trials in all in each block. Each comparison trial followed the response to a "press spacebar for next trial" prompt by 1,500 ms. After each comparison stimulus presentation, the participant judged whether or not the stimulus had the same duration as the standard, making a "Y" (Yes) or "N" (No) response on the keyboard. After the participant had made their response they were given the following feedback dependent on their response and the comparison duration:

> Hit–"CORRECT, that WAS the standard duration"
> Miss–"INCORRECT, that WAS the standard duration"
> False positive–"INCORRECT, that WAS NOT the standard duration"
> Correct rejection–"CORRECT, that WAS NOT the standard duration".

After the feedback message had been displayed for 2,000 ms the participant was prompted to press the spacebar to receive either the next trial or the next block. Following presentation of all nine comparison stimuli, a new block corresponding to one of the three conditions began, with a different standard duration being generated.

**Condition 2 (*near* condition).** The same procedure as that for Condition 1 was used except for presentation of the standard duration. At the beginning of each block a single standard was generated (randomly drawn from a uniform distribution between 500 and 700 ms), and then two other durations, which were the standard plus 100 ms and the standard minus 100 ms, were generated. The three stimuli were then presented (as in the same condition) with order of the presentation randomized for each block. The rest of the block followed the same procedure as that for the normal condition.

**Condition 3 (*far* condition).** The procedure was identical to that for the near condition except that the three stimuli presented were the standard, the standard plus 300 ms, and the standard minus 300 ms.

In both Conditions 2 and 3 the standard duration differed randomly for each block, which ensured that all three stimulus durations presented at the start of the block also differed.

*Experiment 2b*. The procedure for Experiment 2b was very similar to that for Experiment 2a. The important difference was that in all the three conditions tested (very close, close, and far), although the average of the three stimuli presented at the start of each block was presented as a comparison stimulus, none of the three stimuli presented at the start of the block actually was this average, except by chance. This was accomplished by generating an initial duration, as in Experiment 2a, and then adding and subtracting variable quantities to it to make up the other two durations presented. Thus, it was unlikely that any of the three durations presented was the average value.

**Condition 1 (*very close* condition).** At the beginning of each block a duration was generated from a uniform distribution running from 500 to 700 ms. The three durations presented were (a) the generated duration, (b) the generated duration with a random value drawn from a uniform distribution running from 25 to 75 ms added to it, and (c) the generated duration with a random value drawn from a uniform distribution running from 25 to 75 ms subtracted from it. The order of presentation of these three durations was randomized for each block.

Following the three initial presentations, the participant received comparison tones that had a duration equal to the average of the durations presented at the start of the block multiplied by 0.4, 0.6, 0.8, 1, 1.2, 1.4, or 1.6, with the order randomized for each block. The average of the initially presented durations (multiple = 1.0) was presented three times as a comparison, thus making nine trials in all.

**Condition 2 (*close* condition).** This was identical to the very close condition except that the durations presented at the beginning of each block were (a) the generated duration, (b) the generated duration increased by a random value from 50 to 150 ms, and (c) the generated duration decreased by a random value drawn from a uniform distribution running from 50 to 150 ms. The order of presentation of these three durations was randomized for each block.

**Condition 3 (*far* condition).** This was the same as the very close condition except that the three durations presented at the start of each block were (a) the generated duration, (b) the generated duration plus a random value from 150–450 ms, and (c) the generated duration minus a random value from 150–450 ms.

All other experimental details, such as trial randomization and feedback, were the same as those in Experiment 2a.

## Results

Figure 3 shows temporal generalization gradients, in the form of the proportion of YES responses (judgements that a comparison was the standard average), plotted against the comparison/standard average ratio. The upper panel shows data from Experiment 2a, the lower panel data from Experiment 2b.

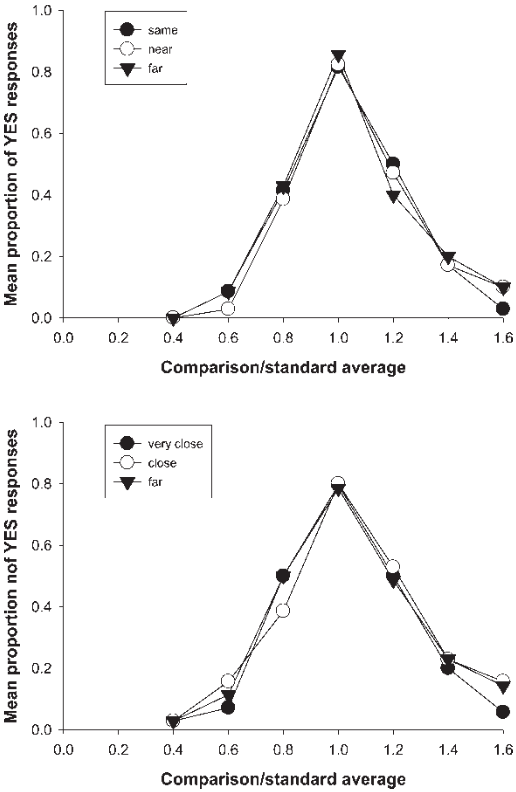

**Figure 3.** *Temporal generalization gradients in the form of proportion of YES responses (judgements that a presented comparison duration was the average of the three presented standards), plotted against comparison/average ratio. Data from Experiment 2a are in the upper panel, data from Experiment 2b in the lower panel. The different spacing conditions (same, near, far, etc.) are indicated by the different symbols.*

Inspection of the results in both panels shows that, in both experiments and all conditions, the proportion of YES responses peaked at the comparison value that actually was the standard average. Comparing the different spacing conditions (close, far, etc.) from the different experiments likewise showed no obvious effect of condition on the proportion of YES responses at the average of the standards.

These suggestions were confirmed by statistical analysis. Consider first the data from Experiment 2a (upper panel of Figure 3). A Freidman test found no effect of number of standard presentations on the mean proportion of YES responses to the standard average when it was presented, when comparing the same, near, and far conditions, $\chi^2(2) = 1.51$, $p = .470$. Values ranged over .82 to .86 across conditions. Consider next data from Experiment 2b (lower panel of Figure 3). The proportion of identifications of the standard ranged from .80 to .81, and a Friedman test found no effect of spacing condition on the mean proportion of YES responses to the standard when it was presented, when comparing the very close, close, and far conditions, $\chi^2(2) = 0.255$, $p = .880$.

Consider next the analyses of the whole temporal generalization gradients from Experiment 2a (upper panel of Figure 3). A repeated measures ANOVA used averaging condition (same, near, and far), and comparison/standard average ratio (effectively the duration of the comparison) as within-subject factors. There was no effect of condition, $F(2, 30) < 1$, indicating that condition did not affect the overall level of responding. There was a significant effect of comparison/standard average ratio, $F(6, 90) = 107.232$, $p < .001$, indicating that the participants were sensitive to comparison duration. There was no significant interaction between condition and comparison/standard average ratio, $F(12, 180) < 1$, indicating that condition did not affect the shape of the temporal generalization gradients.

Similar results were found when data from Experiment 2b (lower panel of Figure 3) were analysed. A repeated measures ANOVA used averaging condition (very close, close, and far), and comparison/standard average ratio as within-subject factors. There was no effect of condition, $F(2, 30) = 2.404$, $p = .111$, indicating that condition did not affect the overall level of responding. There was a significant effect of comparison/standard average ratio, $F(6, 90) = 56.769$, $p < .001$, indicating that the participants were sensitive to comparison duration. There was no significant interaction between condition and comparison/standard average ratio, $F(12, 180) < 1$, indicating that condition did not affect the shape of the temporal generalization gradients.

## Discussion

The results of Experiment 2 show that people can use the average duration of three presented stimuli as a standard. Perhaps more surprisingly, they show that the spacing of the three stimuli makes no difference to temporal generalization performance: Whether the stimuli are closely spaced, more widely spaced, or even repeated (as in the same condition of Experiment 2a), performance was unaffected. This was also true even if the duration that was actually the average was not presented at the start of the block (as in Experiment 2b).

Such efficient arithmetical averaging of durations strongly supports the idea of an underlying linear time scale and seems inconsistent with any nonlinear time scale. One possible objection is that the temporal generalization technique is insufficiently sensitive to detect a deviation from arithmetical averaging caused by an underlying nonlinear time scale, although there is previous evidence against this. Consider the same and far conditions of Experiment 2a. If we assume that subjective time is, for example, logarithmically scaled, then in the same condition the geometric mean of the three (identical) stimuli would be 600 ms on average over the session, whereas for the far condition the three durations, on average, would be 300, 600, and 900 ms. The geometric mean of these three durations is 545 ms. Previous work using temporal generalization (Droit-Volet, Clément, & Wearden, 2001; Penton-Voak, Edwards, Percival, & Wearden, 1996; Wearden, 1992) suggests that such different average values would certainly produce significantly different temporal generalization gradients. For example, Wearden (1992) showed that the temporal generalization gradient shifted when the standard varied over values of 500, 600, and 700 ms, Penton-Voak et al. (1996) used a "speeding up the clock" manipulation in temporal generalization and found a shift of gradient when the "speeding up" effect was around 5–10%, and Droit-Volet et al. (2001) showed that temporal generalization gradients were sensitive to a tendency in young children to remember standard durations as being slightly shorter than they really are (see also McCormack, Brown, Maylor, Darby, & Green, 1999). From these results is seems probable that a shift in the average standard occasioned by the different duration spacings, if some nonlinear time scale was being used, would have been detectable.

The ability of our participants to average three, sometimes very different, durations together to produce an arithmetic average may seem remarkable, but such an ability would arise very naturally if participants possessed a stopwatch-like accumulation mechanism. Early work using the framework that later became scalar expectancy theory (SET; e.g., Roberts & Church, 1978, p. 334) introduced a stopwatch metaphor for timing, so organisms were assumed to possess a stopwatch-like mechanism, which could be started and stopped, and which could also automatically accumulate durations between different start and stop instances. Such a mechanism would obviously facilitate the kind of arithmetic averaging found in our Experiment 2 and would produce our results if an underlying linear timescale was being used.

Our Experiment 2 is not in fact the first to pose participants the problem of averaging presented durations. In their "successive presentation" experiment Curtis and Rule (1977) gave participants two successive durations on each trial. The eight duration values used ranged from 0.5 to 10 s, and all 64 combinations were presented. People were required to classify the average of the two durations presented using an 11-category scale, running from shortest (1) to longest (11). In general, their responses provided good evidence for a linear time scale (see Curtis & Rule, 1977, p. 583, for discussion), consistent with the data from our Experiment 2.

## GENERAL DISCUSSION

Both experiments provided evidence that subjective time grows as a linear, rather than nonlinear, function of real time and thus support conclusions derived from the human time-left analogue (Wearden, 2002). No evidence for nonlinear time growth was found in either experiment, in spite of the fact that Experiments 1 and 2 used different time ranges (up to 10 s for Experiment 1, usually less than 1 s for Experiment 2) and very different methods (verbal estimation of percentages and temporal generalization).

It should also be noted that data from paradigms such as temporal generalization (Wearden, 1991a, 1992; Wearden et al., 1997; Wearden & Towse, 1994), temporal bisection (Wearden, 1991b; Wearden & Ferrara, 1995, 1996), and categorical timing (Wearden, 1995) are also compatible with an underlying linear time scale, even though, as noted earlier, their data cannot conclusively prove that linear time growth exists. Given that so much evidence is consistent with linear timing and that tests of linear versus nonlinear timing, like those in the present study and that provided by the time-left analogue (Wearden, 2002), usually support the contention of linear timing, why are alternatives so popular? Any reader who thinks that the idea of linear timing is universally accepted should read Staddon and Higa's (1999) vigorous attack on this notion and their staunch advocacy of logarithmic relations between subjective and real time (e.g., their pp. 219–225).

The main reason for scepticism about linear timing may be that nonlinear time growth may seem a more attractive prospect than linear timing for theoretical rather than empirical reasons. Linear timing follows very naturally from the idea that organisms possess a pacemaker–accumulator internal clock. Such a clock, even if its pacemaker operates in a number of different ways (e.g., Gibbon, Church, & Meck, 1984), will almost always generate linear timing. For example, if the pacemaker produces pulses with complete regularity (Treisman, 1963) then, obviously, the number of pulses produced in time $nt$ will just be $n$ times the number produced in $t$. At the other extreme of pacemaker regularity, this is also true for a Poisson pacemaker, where pulses are emitted at random with some rate that is constant on average. Once again, on average, in a time $nt$ such a pacemaker will produce $n$ times the number of pulses in time $t$, although there will be variability from one example of the timing of an interval to another.

Although a linear time scale follows naturally from the idea that timing is based on a pacemaker–accumulator mechanism, it is much less compatible with other types of potential timing mechanisms. For example, various sorts of oscillatory-based processes (e.g., Brown, Preece, & Hulme, 2000; Church & Broadbent, 1990; Miall, 1989, 1992) may not generate underlying time representations that are linear with real time. The recent instantiation of the idea that timing might be based on memory decay by Staddon and Higa (1999; see also Staddon, Chelaru, & Higa, 2002, for a development of this idea) likewise generates time scales that are not linear in real time, and thus their advocacy of logarithmic timing may come partly from their desire to replace the pacemaker–accumulator clock of SET by another mechanism. So, in general, resistance to the idea of a linear time scale may be based more on objections to a pacemaker–accumulator clock (objections that may be based on presumed physiological plausibility or other considerations, e.g., Matell & Meck, 2000, 2004) and may owe its persistence more to this than to any direct evidence that time growth is nonlinear.

We do not imagine that the results presented in the current article will settle the issue of linear versus nonlinear time growth forever, nor do we expect them to effect a Damascene conversion of those who support models based on nonlinear timing. Nevertheless, our results support linear timing very strongly, in situations where nonlinear timing, if present, would seem to have had a good chance to manifest itself clearly, and thus provide a challenge for the contention that the growth of subjective time is not a linear function of real time.

Original manuscript received 8 August 2005
Accepted revision received 3 August 2006
First published online 23 November 2006

## REFERENCES

Allan, L. G. (1978). Comments on current ratio-setting models for time perception. *Perception and Psychophysics*, *24*, 444–450.

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon. *Psychological Science*, *12*, 238–243.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–181.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, *124*, 434–452.

Buckley, P. B., & Gillman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology*, *103*, 1131–1136.

Cerutti, D. T., & Staddon, J. E. R. (2004). Immediacy versus anticipated delay in the time-left experiment: A test of the cognitive hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 45–57.

Church, R. M., & Broadbent, H. (1990). Alternative representations of time, number and rate. *Cognition*, *37*, 55–81.

Curtis, D. W., & Rule, S. J. (1977). Judgment of duration relations: Simultaneous and sequential presentation. *Perception and Psychophysics*, *22*, 578–584.

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, *12*, 244–246.

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

Gallistel, C. R. (1999). Can a decay process explain timing of conditioned responses? *Journal of the Experimental Analysis of Behavior*, *71*, 264–271.

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: Vol. 423. Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.

Jones, L. A., & Wearden, J. H. (2003). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *Quarterly Journal of Experimental Psychology, 56B*, 321–343.

Jones, L. A., & Wearden, J. H. (2004). Double standards: Memory loading in temporal reference memory. *Quarterly Journal of Experimental Psychology, 57B*, 55–77.

Matell, M., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, *22*, 94–103.

Mattel, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: Coincidence detection and oscillatory processes. *Cognitive Brain Research, 21*, 139–170.

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.

Miall, R. C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation, 1*, 359–371.

Miall, R. C. (1992). Oscillators, predictions, and time. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action, and cognition* (pp. 215–227). Dordrecht, The Netherlands: Kluwer.

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.

Preston, R. A. (1994). Choice in the time-left procedure and in concurrent chains with a time-left terminal link. *Journal of the Experimental Analysis of Behavior*, *61*, 349–373.

Roberts, S., & Church, R. M. (1978). Control of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*, 318–337.

Staddon, J. E. R., Chelaru, I. M., & Higa, J. J. (2002). A tuned-trace theory of interval-timing dynamics. *Journal of the Experimental Analysis of Behavior*, *77*, 105–124.

Staddon, J. E. R., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*, 215–251.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs, 77* (Whole No. 576).

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

Wearden, J. H. (2002). Traveling in time: A time-left analogue for humans. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 200–208.

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., & 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., & Towse, J. (1994). Temporal generalization in humans: Three further studies. *Behavioural Processes*, *32*, 247–264.