THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
0000, 00 (0), 000–000

Psychology Press
Taylor & Francis Group

Slowing down an internal clock: Implications for accounts of performance on four timing tasks

J. H. Wearden
Keele University, Keele, UK

An experiment investigated the potential effects of lowering arousal on performance on time perception tasks. Four participant groups received different tasks: Normal and episodic temporal generalization, bisection, and verbal estimation, all involving judgements of the duration of visual stimuli. Self-rated arousal during the experimental session was lowered by spacing experimental trials approximately 10 s apart. Between the early and late blocks of the experiment, performance changed on normal temporal generalization and verbal estimation, but not on episodic temporal generalization and bisection. The changes were consistent with the idea that the pacemaker of the participant's internal clock had been slowed down by the slow trial spacing. Results suggested that bisection was based on a criterion that adjusted during the experiment, whereas verbal estimation was based on preexisting standards, or those established early in the experiment.

The idea that humans possess an internal clock-like mechanism that they can use to perform some timing tasks is not new, and research in this area dates back to the 1920s (see Wearden, 2005, and Wearden & Penton-Voak, 1995, for reviews). In the last two decades, new life has been breathed into this notion by the notable success of scalar timing (or scalar expectancy) theory (SET: Gibbon, Church, & Meck, 1984) as an account of human performance on some timing tasks (see Allan, 1998, and Wearden, 2003, for reviews). SET is a complex multiprocess model of performance on tasks with temporal requirements, and its three component levels are an internal clock, working and reference memory mechanisms, and a decision process. Only after operation of all three parts of the model is observed behaviour produced, with the operation of the individual components being individually affected by a range of manipulations (see Wearden & Bray, 2001, and Wearden & Grindrod, 2003, for examples of the manipulation of memory and decision processes in SET).

The focus of interest of the present article is the internal clock. SET supposes that people possess an internal clock of a pacemaker–accumulator type. When an event to be timed begins, a switch connecting the pacemaker to the accumulator closes and this allows the "pulses" of the pacemaker to be accumulated. When the event stops, the switch opens, thus cutting the connection. The accumulator then contains the basic raw

---

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

I am grateful to a "TP" group at Manchester University Psychology Department for assistance in running participants in this experiment.

http://www.psypress.com/qjep
DOI:10.1080/17470210601154610

time representation, but how this is used depends, sometimes in complex ways, on the memory and decision processes that the particular timing task is said to involve.

Important evidence for the existence of a pacemaker–accumulator clock in humans comes from attempts to "speed up" or "slow down" the pacemaker of the clock. This manipulation is usually referred to as a "clock speed" manipulation, and this term is used here. Early studies with humans (inspired by François, 1927, and Hoagland, 1933) attempted to change clock speed by manipulations of body temperature, often using experimental procedures that might give contemporary ethics committees considerable pause for thought, but in general their results supported the view that raised body temperature resulted in faster clock speed, whereas studies of lowered body temperature, which were very rare, resulted in slower clock speed (Wearden & Penton-Voak, 1995, review both sorts of studies).

Fortunately, modern experimental participants can escape the rigours of some of the body temperature experiments thanks to a method introduced by Treisman, Faulkner, Naish, and Brogan (1990). As part of a more complex investigation, Treisman et al. (1990) proposed that accompanying stimuli whose duration had to be judged by trains of repetitive stimulation (such as a train of clicks) resulted in participants behaving as if their internal clock had been sped up, compared with conditions where the clicks were absent. Although the mechanism of the "click trains effect" is unclear (Treisman et al. attributed it to increases in "arousal", although the click trains do not seem subjectively arousing in the everyday sense of this term), the method has proved to be very reliable. Penton-Voak, Edwards, Percival, and Wearden (1996) showed that the click trains could change judgements of subjective time during presentations of auditory and visual stimuli, as well as changing the duration of intervals produced by participants, in a manner consistent with the idea that clock speed had been increased by the clicks. For other confirmatory studies, see Burle and Bonnet (1997, 1999), Burle and Casini (2001), and Wearden, Edwards, Fakhri, and Percival (1998). See also Droit-Volet and Wearden (2002) for effects of repetitive stimulation (flicker) on timing in children as young as 3 years of age.

If body temperature manipulations and the use of click trains can "speed up" the clock, can it also be "slowed down"? As mentioned above, the very few studies that have lowered core body temperature supported the view that lower temperature resulted in lower clock speed, but lowering body temperature requires special equipment and careful medical supervision and will never become a routinely useful laboratory procedure. However, Wearden, Philpott, and Win (1999a) used the Treisman et al. (1990) click train method to provide an example of *relative* slowing down of the clock. In the "speeding up" experiments, the general logic was that behaviour in a control condition was contrasted with that in the putatively speeded-up condition: For example, judgements of a standard duration learned with a "normal" clock can be contrasted with judgements of the same duration made with a "speeded-up" clock. Wearden et al. reversed the usual procedure, and participants learned temporal standards in a bisection task (described more fully later) when the standards were presented after clicks (thus putatively timed with a "faster" clock), and then comparison stimuli were delivered without clicks (thus timed with a "normal" clock, i.e., relatively slowed down). The results were as expected: When people learned standards with a speeded-up clock, they behaved as if normally timed stimuli were timed with a *relatively* slower clock.

Wearden et al. (1999a) could not, however, address the question of what happens when the clock is *absolutely* slowed down. How might such absolute slowing down be achieved? As Wearden and Penton-Voak (1995) noted, a common general idea has been that "arousal" and clock speed are positively related, with higher states of arousal being associated with faster clock speed. This position concurs with anecdotes suggesting that in emergency situations such as car crashes "time seems to stand still"—in other words, external events seem to last much longer than they normally do, exactly the effect that would be predicted from a greatly increased clock speed. Obviously, we cannot ethically put our participants in life-threatening situations, or even pretend to do so, so producing states of extremely high arousal in the laboratory is unlikely to be realized. However, the other end of the arousal spectrum, states of low arousal, might be routinely explored without any ethical or practical problems.

Wearden, Pilkington, and Carter (1999b) conducted a study of this sort, which is the starting point for the present work. The task used was "normal" temporal generalization (Wearden, 1992). In this task, the participant receives a standard duration at the start of the experiment (e.g., a tone or visual stimulus 400 ms long) and then receives comparison stimuli, some of which have the same duration as the standard intermixed with others that are shorter or longer. The participant indicates after each comparison stimulus whether they judged it to be equal to the standard (a YES or NO response). The proportion of YES responses plotted against stimulus duration yields a temporal generalization gradient that is usually (a) peaked at or close to the standard and (b) slightly asymmetrical so that stimuli longer than the standard tend to be more confused with it than those shorter by the same amount (Wearden, 1992; for potential explanations of gradient shape, see Wearden, 2004).

In most temporal generalization experiments, participants receive accurate feedback after each response, but in the study of Wearden et al. (1999b) no feedback was given. The general result was that, as the experiment proceeded, longer and longer comparison stimuli were judged to be the standard (that is, the temporal generalization gradient shifted to the right). Why might this occur? Wearden et al. suggested that the effect was due to a progressive decrease in clock speed caused by decreasing arousal as the experiment proceeded. Suppose that the standard was initially encoded as *n* "ticks" of the internal clock. As the clock speed decreases in the experiment, longer and longer stimuli are needed to produce these *n* ticks, so longer and longer stimuli are identified as the standard. In support of the arousal explanation Wearden et al. (1999b) found independent evidence for changes in arousal using a rating scale derived from Thayer (1967), as shifts in the temporal generalization gradient in the predicted direction were associated with rated decreases in arousal (e.g., see their Experiment 3).

The present article reports data from four different experimental procedures using a particularly boring experimental arrangement, which was intended to produce a fall in participants' arousal as the experiment proceeded. Although lowering arousal in this way might not cause ethical problems, low arousal states may change behaviour for a number of reasons, not only because of putative decreases in clock speed. Obviously, the participant's attention may decline, and in general performance may deteriorate because of decreased motivation to perform tedious tasks for periods of tens of minutes. How can a slowing down effect be distinguished from others?

A potential answer to this lies at the core of the rationale of the present study. In it, I not only introduce a simple method for apparently decreasing clock speed, but also use that method to discover something about how different timing tasks are performed. As mentioned above, SET is a model involving a number of interacting components, as well as the internal clock. The theory has been developed to provide quantitative models of performance on at least some timing tasks, including three of those used in the present article, "normal" and "episodic" temporal generalization, and bisection. An unusual problem for SET is that the fact that a quantitative model consistent with it fits performance on some particular task to a high degree of precision is no guarantee that the model specified is the correct one (see Wearden & Ferrara, 1995, for discussion), but manipulation of one of the components of the model (here the internal clock) can provide insight into the mechanisms actually used by participants.

Four tasks are used here, but it is convenient theoretically to group them into two pairs. The first two tasks both involve temporal generalization. In "normal" temporal generalization, as mentioned above, a standard duration is presented at the start of the experiment, and then later stimuli are compared with it. Only one stimulus is presented on each trial, and the usual model of performance on this task (Wearden, 1992) supposes that the just-presented comparison stimulus duration is held in a working-memory store and compared with a "sample" of the memory of the previously learned standard, retrieved from a reference memory store. If the two duration representations are "close enough", according to a specified decision rule (the one usually proposed comes from Wearden, 1992), then the participant responds YES; otherwise they respond NO.

It is the proposed difference between the representation of the standard and the comparison that is the basis of a predicted effect if the clock is "slowed down". The standard is learned at the beginning of the experiment, thus presumably timed with a "normal" clock, yet as the experiment proceeds the comparison stimuli are timed by a slower and slower clock, and thus the generalization gradient should shift to the right. To illustrate this, suppose that a 400-ms standard duration is used, and this is encoded as 40 "ticks" of the internal clock—that is, initially the clock "ticks" every 10 ms. Early in the experiment, stimuli 400-ms long will be encoded in terms of 40 "ticks", but if the clock slows down (e.g., to 1 "tick" every 12 ms), then stimuli nearly 500 ms long are needed for 40 "ticks" to accumulate so, in general, stimuli longer than the standard will tend to be confused with it to a greater extent late in the experiment, when arousal has fallen, than earlier. The figures given for clock speed and its potential change are invented, of course, but serve to illustrate the principle.

In contrast, "episodic" temporal generalization (Wearden & Bray, 2001) presents the participant with two stimuli on the trial, and in most examples with this method the stimuli change from one trial to another. The stimuli can have the same duration, or different durations (i.e., they can have some ratio to one another other than 1.0), but there is no initially learned "standard". Presumably, in this case, both stimuli in the trial are held briefly in working memory for comparison, only to be replaced on the next trial by two others. Now, even if the internal clock slows down, no effect should be noted, as both stimuli on the trial are timed with the same clock speed, whatever that is, and the basis for their judgement is the potentially different numbers of "ticks" accumulated during each stimulus, not the absolute number of clock "ticks". On the other hand, if the experimental procedure generates a decline in attention or some other general performance-reducing factor, then performance on episodic generalization should change as the experiment proceeds.

The second pair of tasks is bisection and verbal estimation. In the bisection method used here and in some other studies (e.g., Wearden, 1991; Wearden & Ferrara, 1995; Wearden, Rogers, & Thomas, 1997), the participant initially receives two standard durations, a standard SHORT and a standard LONG value (e.g., tones 200 and 800 ms in duration). Then a series of comparison values is presented (e.g., 200 to 800 ms in 100-ms steps), and the participant's task is to classify each one in terms of its judged similarity to the SHORT and LONG standards. Superficially, the bisection task seems to have much in common with "normal" temporal generalization, except that there are two standards instead of one, and, indeed, early theories of bisection performance (Allan & Gibbon, 1991; Wearden, 1991) assumed that the comparisons were judged in terms of their similarity to each of the standards learned initially. However, later work (Allan & Gerhardt, 2001; Wearden & Bray, 2001; Wearden & Ferrara, 1995) has cast doubt on this, and a recent idea is that each comparison is compared not with the standard SHORT or LONG, but with some intermediate "criterion" duration (Wearden & Ferrara, 1995, suggested the arithmetic mean of all the durations presented, but this is not the only possibility). So, for example, if a comparison duration is longer than this criterion the participant classifies the stimulus as "long", if shorter as "short": The original standards have no special status except insofar as they contribute to the criterion.

A persistent problem is that different models of bisection will all fit normal bisection performance to a high degree of precision, so experimental manipulations rather than modelling are needed to distinguish different possibilities. The putative slowing-down operation used here can help. For example, suppose that people are using the initially learned SHORT and LONG standard durations as the basis for classifying later comparison stimuli.

To use our earlier example, suppose that a 200-ms duration is presented as the SHORT standard and an 800-ms one as the LONG one, and that these are initially encoded as 20 and 80 "ticks", respectively. Now, as in the "normal" temporal generalization case, performance should shift as the experiment proceeds as the comparison stimuli are being judged using a slower clock than that used to establish the standards. More stimuli should be judged as "short" as the experiment proceeds. For example, a 400-ms stimulus initially gives rise to 40 clock "ticks", but later in the experiment to only 33 (to use the illustrative clock-speed values used earlier), so its representation is more similar to the SHORT standard later in the experiment than earlier and thus gives rise to more "short" responses as the experiment proceeds. This will also be true if the "criterion" is established early in the experiment and then maintained. On the other hand, if the "criterion" is more local, then it might itself shift as the experiment proceeds, resulting in no effect of the "slowing-down" manipulation. As in the episodic generalization case, however, a general decline in attention might result in a degradation of performance at all stimulus durations, rather than a predictable shift.

The final task used is verbal estimation of duration, where people assign verbal labels (in ms, i.e., "1,000 = 1 second") to the duration of stimuli presented. It is known that verbal estimation is very sensitive to the "speeding-up" operations used in some studies (e.g., Penton-Voak et al., 1996), but these involved the intermixing of "speeded-up" and "normal" stimuli, so some implicit comparison of the two types was possible. The predictions for performance on a verbal estimation task of an absolute slowing-down manipulation are difficult to specify for certain. If the participant uses some implicit standard (e.g. "one second feels like *this* . . .) which is derived from experience outside the current experiment, or is established at the start of the experiment, then a slowing-down effect should be expected: Stimuli should be judged as shorter as the experiment proceeds. If, on the other hand, any standard used is locally constructed (e.g., every trial, or every few trials), then there will be little or no slowing down effect observable, as the standard used will also itself be subject to slowing down.

It can be seen from the above that a putative slowing-down manipulation can help uncover the mechanism of performance on different timing tasks. An effect on performance in normal but not episodic generalization would be an experimental confirmation of the existing theoretical orthodoxy with respect to such tasks (specified in detail in Wearden, 2004). An effect on bisection involving the general "shortening" of stimulus durations would suggest a constant criterion, whereas no effect would imply a shifting criterion. Likewise, a shortening effect on verbal estimation would suggest a pre- or early established "standard", whereas no effect would suggest a changing one.

The basic experimental procedure used here involved spacing out experimental trials so that one occurred approximately every 10 s. As the stimuli used were all short (to prevent chronometric counting), and the timing decisions fairly simple, the actual work of the trial took around just a few seconds, so the participant spent a lot of time faced with a "please wait" prompt, and the trial pace was subjectively very slow (see below for details). Arousal was measured before the procedure started, and immediately after it had finished, using a scale derived from Thayer (1967). Given the fact that the procedure was deliberately developed to be boring, rated arousal was expected to fall between the two (before and after) measures. Performance early in the task (when arousal was presumably higher) was always contrasted with late performance (when arousal would have been expected to fall), so this comparison shows the behavioural effect for each task that accompanies the expected arousal change.

Method

Participants

A total of 60 Manchester University undergraduates participated for course credit; 14 were arbitrarily allocated to the "normal" temporal generalization condition, 16 to the bisection group, and 15 each to the episodic temporal generalization and verbal estimation conditions. All received a single experimental session lasting about 30 minutes.

Apparatus

An Opus 16-X IBM-compatible PC controlled the experiment. The stimuli to be judged were light-blue 5 × 5-cm squares located in the centre of the computer screen. The keyboard registered responses. The experimental conditions were programmed in MEL (Micro-Experimental Laboratory).

Procedure

The experimental session began with administration of a miniquestionnaire derived by Wearden et al. (1999b) from Thayer (1967). The questionnaire consisted of eight questions comprising two subscales, and participants rated their state on a 4-point scale going from "definitely feel" to "definitely do not feel". The two scales were *activation* (energetic, jittery, lively, stirred-up), and *deactivation* (drowsy, calm, sleepy, relaxed). For more details see Wearden et al. (1999b, p. 34), and below. The questionnaire was also administered after the experimental procedure described below. As well as filling in the questionnaire twice, each participant received one of the procedures described below.

All the trials of the tasks used were self-paced, and the participant pressed a spacebar to initiate trial events. A delay then ensued, and the stimuli were presented, followed by the opportunity to respond. Prior to the procedures reported below, a pilot study using 3 participants (whose data are not reported here) measured how long it took, on average, for the participant to carry out the trial events described below, but without the "please wait" period that spaced out the trials. The time from the spacebar press to the response never took more than 5 s for the normal and episodic generalization and the verbal estimation task, and never more than 4 s for the bisection task. The procedures reported below spaced out trials by approximately 10 s, so the "please wait" period (see below) was set at 5 s for three of the tasks and 6 s for bisection. Self-pacing of trials was used in preference to experimenter-controlled pacing, as visual stimuli, some of which were very brief, were employed, so there was a risk with experimenter-controlled pacing that the participant may not have been looking at the screen when the stimulus was presented.

Normal temporal generalization. The experimental session began with five presentations of a 400-ms blue square that served as the standard for the task, with each presentation spaced from the previous one by a random value picked from a uniform distribution running between 2,000 and 3,000 ms. Then comparison stimuli were presented. The participant received a "Press spacebar for next trial" prompt, and a response followed by a delay ranging from 2,000–3,000 ms and then the stimulus presentation. The comparison stimuli were arranged in blocks of 10, with the presentation order randomized within and between blocks. Stimulus durations were 100, 200, 300, 350, 400 (presented twice per block), 450, 500, 600, and 700 ms. After each comparison stimulus presentation, the participant pressed the "Y" or "N" key to indicate their judgement as to whether or not the comparison stimulus duration was the standard. No feedback was given after responses, and 10 blocks of comparison stimuli were presented. After each response, the participant received a "Please wait" prompt, followed after 5 s by the "Press spacebar for next trial" prompt.

Episodic temporal generalization. On each trial, after the participant responded to a "Press spacebar for next trial" prompt, a delay ranging from 2,000–3,000 ms ensued, and then two blue squares were presented with a gap ranging from 400 to 600 ms between them. After the second stimulus had been presented, the participant pressed the "Y" or "N" keys to indicate whether or not they judged the stimuli to have the same duration. No feedback was given. The response was followed by a "Please wait" prompt, lasting 5 s. The stimuli were generated as follows. On each trial, a "standard" was randomly selected from one of two equally likely ranges: 300–500 ms, and 600–1,000 ms. Once the "standard" had been selected, it was multiplied by one of the following ratios to produce a "comparison": 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75. Once the "standard" and "comparison" had been generated, their order of presentation was randomized with the "comparison" coming first on half the trials, and the "standard" on the other half. Stimulus durations were never repeated except by chance. A block consisted of 16 trials, the seven ratios above (with 1.0 presented twice) used with "standard" stimuli from the 300–500-ms and 600–1,000-ms ranges. Seven blocks were given in all.

Bisection. The experiment began with four presentations each of the standard SHORT and LONG durations, which were blue squares with presentation durations of 200 and 800 ms, respectively, with the SHORT presented first, and random delays ranging from 1,000 to 2,000 ms between presentations. Then 10 blocks of comparison stimuli were presented. A response to a "Press spacebar for next trial" prompt was followed by a delay ranging from 2,000 to 3,000 ms, then stimulus presentation. A block of stimuli consisted of the duration values 200, 300, 400, 500, 600, 700, and 800 ms, with order randomized between and within blocks. After each stimulus presentation, the participant was required to classify it in terms of its similarity to the standard SHORT and LONG durations presented at the start, by pressing the "S" and "L" keys. The response was followed by a "Please wait" lasting 6 s, and then the "Press spacebar for next trial" prompt was presented again.

Verbal estimation. A response to a "Press spacebar for next trial" prompt was followed by a random delay ranging from 2,000 to 3,000 ms and then presentation of a single blue square. After the square had been presented, the participant was required to type in its estimated duration, using a scale of 1,000 = 1 second. Participants were informed that all durations were between 50 and 1,500 ms and were asked not to use values outside that range. Stimuli to be estimated were presented in blocks of nine, and 10 blocks were given in all. Six of the duration values in each block were 77, 203, 461, 707, 958, and 1,183 ms. The other three values in the block were randomly selected from a range of 50 to 1,500 ms. These were included to prevent the participant realizing that stimulus durations were repeated. The order of presentation of stimuli was randomized between and within blocks. Typing in the verbal estimate was followed by a "Please wait" lasting 5 s.

Results

Arousal measures

The scale used generated two scores: *activation* (associated with high arousal) and *deactivation* (associated with low arousal). The miniquestionnaire, derived from Thayer (1967) and used by Wearden et al. (1999b), in the present study used a scale from 1 (definitely feel) to 4 (definitely do not feel), so higher values meant less of the measure assessed. The overall activation and deactivation measures were derived from 4 items measuring activation and 4 measuring deactivation, so the maximum score on each was 16, the minimum 4. Median values for deactivation and activation taken before the timing procedure was initiated were: activation: normal generalization group, 11.5; episodic generalization group, 12.0; bisection group, 11.5; verbal estimation group, 11.0; deactivation: normal generalization group, 9.5; episodic generalization group, 8.0; bisection group: 10.5; verbal estimation group: 10.0. Kruskal–Wallis tests were used to examine potential between-group differences in activation and deactivation before the timing procedures started, but no significant differences were found: activation, $\chi^2(3) = 3.11$, $p = .38$; deactivation, $\chi^2(3) = 6.25$, $p = .10$.

Scores on both measures from before and after the experimental procedure were analysed with Wilcoxon tests. For all groups, one or both of these measures changed significantly in the predicted direction (i.e., decrease in activation, increase in deactivation): normal temporal generalization (activation: *ns*; deactivation: $p < .05$); episodic temporal generalization (activation: *ns*; deactivation: $p < .05$); bisection (activation: $p < .05$; deactivation: $p < .05$); verbal estimation (activation: $p < .01$; deactivation: $p < .05$). The two nonsignificant changes were in the predicted direction, so overall the procedure changed levels of rated arousal in all groups. The possibility remained, however, that the arousal changes were more marked in one group than another, so to test this differences in deactivation and activation were calculated by subtracting the values obtained at the start of the experimental session from those obtained at the end. This yielded a deactivation and activation difference score for each participant in each group, and a Kruskal–Wallis test was used to examine potential between-group differences. No significant difference between the groups was found either in deactivation difference scores, $\chi^2(3) = 3.63$, $p = .30$, or activation difference scores, $\chi^2(3) = 0.63$, $p = .89$.

Performance on the timing tasks

The upper panel of Figure 1 shows data from the normal temporal generalization group, with data coming from the first (early) and last (late) three blocks of the experiment. Inspection of the generalization gradients suggests a change during the experiment, with the gradient more skewed towards the right in the last three blocks than in the first. Analysis of variance (ANOVA) confirmed this suggestion. There was a significant effect of early versus late, $F(1, 13) = 7.73$, $p = .015$, stimulus duration, $F(8, 104) = 14.46$, $p < .001$, and a significant Early/Late × Duration interaction, $F(8, 104) = 3.86$, $p < .01$. The stimulus duration effect merely indicates that participants were sensitive to stimulus duration, but the other two indicate a significant change in performance as the experiment proceeded, with the latter confirming the shift in the gradient shape.

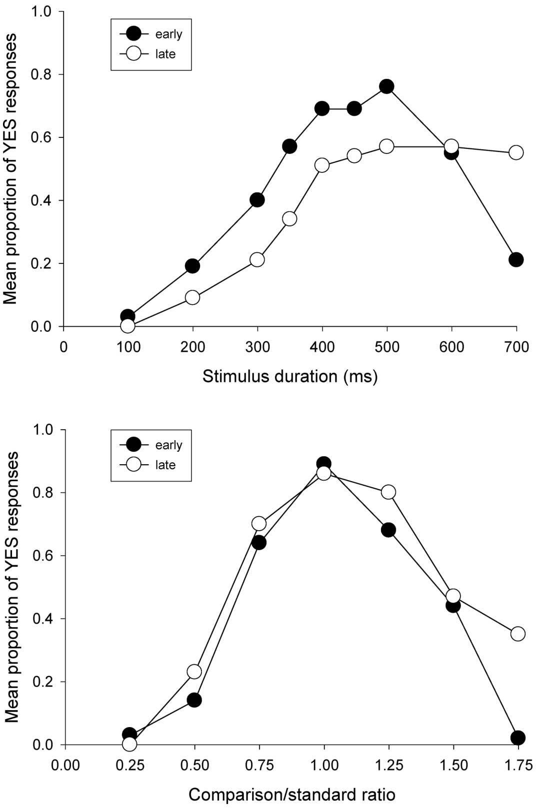

Figure 1. *Upper panel: Mean proportion of YES responses (judgements that a comparison stimulus had the same duration as the standard) plotted against comparison stimulus duration from the normal temporal generalization group. Data are shown separately from early (first three blocks: filled circles) and late (last three blocks: unfilled circles) phases of the experiment. Lower panel: Mean proportion of YES responses (judgements that the two stimuli on the trial had the same duration) plotted against comparison/standard ratio, from the episodic temporal generalization group. Data are shown separately from early (first two blocks: filled circles) and late (last two blocks: unfilled circles) phases of the experiment.*

The lower panel of Figure 1 shows data from the episodic temporal generalization group. Inspection of the data suggests that effects of arousal (early and late) in this condition were much less marked. The gradients were peaked at 1.0 (i.e., when the two durations presented actually were equal) in both the first and last two blocks. There was no effect of early versus late, although the results approached significance,

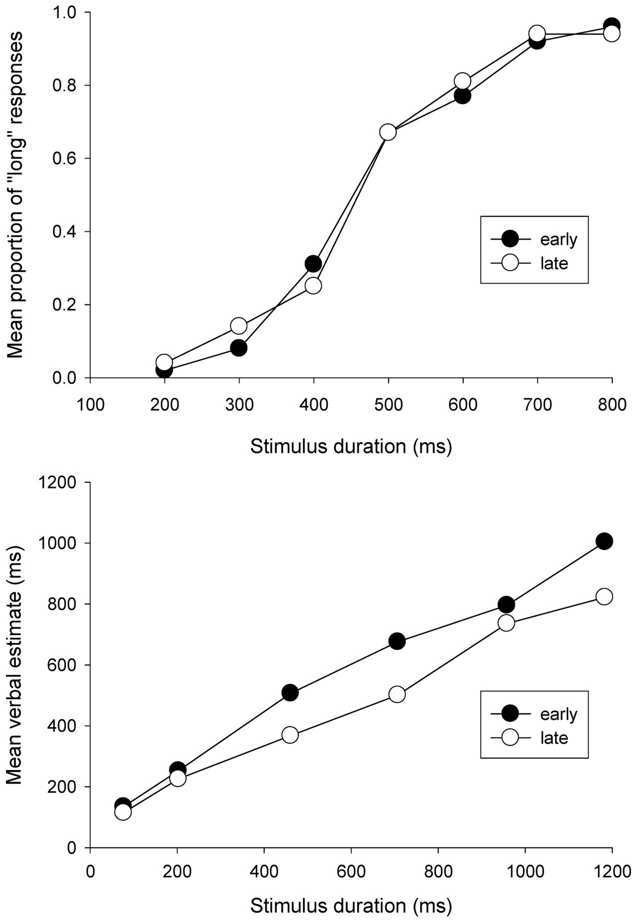

Figure 2. *Upper panel: Psychophysical function (mean proportion of "long" response plotted against comparison stimulus duration) from the bisection group. Data are shown separately from early (first three blocks: filled circles) and late (last three blocks: unfilled circles) phases of the experiment. Lower panel: Mean verbal estimates (in ms) plotted against stimulus duration from the verbal estimation group. Data are shown separately from early (first three blocks: filled circles) and late (last three blocks: unfilled circles) phases of the experiment.*

$F(1, 14) = 4.01$, $p = .065$, nor a Stimulus Duration × Early/Late interaction, $F(6, 84) = 0.95$, *ns*, but there was an effect of relative stimulus duration, $F(6, 84) = 84.65$, $p < .001$.

The upper panel of Figure 2 shows data from the bisection group. The mean proportion of "long" responses (judgements that the presented stimulus was more similar to the LONG standard than to the SHORT one) are plotted against stimulus duration. Obviously, both early and late in the experiment (first and last three blocks), the proportion of "long" responses went from near 0 when the 200-ms duration was presented to near 100% when the 800-ms duration was. There appeared to be no change in behaviour during the experiment. These suggestions were confirmed by statistical analysis. Neither the effect of early and late, $F(1, 15) = 0.04$, nor the Early/Late × Stimulus Duration interaction, $F(6, 90) = 0.34$, was significant or approaching significance, but the effect of stimulus duration was highly significant, $F(9, 60) = 113.18$, $p < .001$.

The lower panel of Figure 2 shows data from the verbal estimation group. Inspection of the panel suggests that participants' mean verbal estimates of duration were highly sensitive to actual stimulus duration and that behaviour changed as the experiment proceeded, with shorter estimates occurring late in the experiment. These suggestions were confirmed by ANOVA. There was a significant effect of stimulus duration, $F(5, 70) = 141.30$, $p < .001$, and of early versus late, $F(1, 14) = 19.16$, $p < .01$, as well as a significant Early/Late × Stimulus Duration interaction, $F(5, 70) = 3.66$, $p < .01$. Inspection of the data and the significant interaction suggests that the early/late effect was manifested in different slopes for the mean estimate versus duration function. As discussed in detail elsewhere (Wearden et al., 1998) such slope effects are expected if clock speed has been changed. To investigate this issue further, the mean estimates from each individual, both early and late, were regressed against stimulus duration. This yielded early and late slopes and intercepts. Mean slope values were: early, .76; late, .63. Mean intercept values were: early, 108.20; late, 81.63. Both measures were analysed two ways, with a Wilcoxon test and with a paired *t* test, even though the normality assumptions of the latter are probably violated. Comparison of early and late values revealed a significant decrease in slope according to both tests ($p = .045$, and .03, respectively), but intercepts did not differ ($p = .233$, and .348).

Discussion

The results can be simply summarized. Spacing experimental trials by approximately 10 s resulted in a significant fall in arousal in all groups, in one or both of the self-rated arousal scales, and between-group comparisons of arousal differences before and after the experiment did not find any significant between-group difference, suggesting that arousal fell in all groups to a similar extent. Performance on normal temporal generalization and verbal estimation changed significantly as the experiment proceeded, whereas performance on episodic temporal generalization and bisection were not significantly affected. This latter result implies that the changes in performance when they occurred were not the result of some general decline in attention or motivation to perform the task: For example, the extreme durations in bisection (200 ms and 800 ms) were almost always "correctly" classified both early and late in the experiment.

How consistent are the effects obtained with the hypothesis that the decrease in arousal slowed down the internal clock? The clearest predictions involve a comparison of normal and episodic temporal generalization. On normal generalization, the standard is assumed to be encoded in terms of a "normal" clock, so the generalization gradient should become skewed to the right as the experiment proceeds, which was the effect found. This also replicates the results in Wearden et al. (1999b). On the other hand, performance on episodic temporal generalization should be unaffected, as both stimuli in the trials are assumed to be timed by the same clock, regardless of its current speed, and this result was again obtained.

If these results support the view that the internal clock has been slowed down, then implications for accounts of performance on bisection and verbal estimation naturally follow. There was no difference in bisection performance early and late in the experiment, which implies that the "criterion" being used to classify stimuli as "short" or "long" itself changes in the course of the experiment. In contrast, the early/late effect on verbal estimation (with stimuli being judged as shorter later in the experimental session than earlier) implies that stimulus durations are being judged according to a scale or standard established early in the experiment. For example, if people were comparing the duration of each stimulus presented with some standard that was established at the outset or preexisted (e.g., what 1 second "feels like"), then the slowing of clock speed later in the experiment would produce the effects obtained. The significant interaction, and the results of the regression-based analysis (showing an effect on slopes but not intercepts) are, of course, consistent with this interpretation.

The idea that some preexisting standard is used by participants in verbal estimation is supported by recent work by Wearden, Todd, and Jones (2006).

We investigated verbal estimation of the duration of auditory (tones) and visual (squares like those used in the present study) stimuli and found that the usual effect that "tones are judged longer than lights" (Wearden et al., 1998) was obtained in participants who received both sorts of stimuli intermixed. However, the same effect was obtained and was of the same magnitude for groups who received only one stimulus type (i.e., the effect was manifest in between-group comparisons as well as within-group ones). The obvious suggestion is that people in the different groups compared the stimuli they receive with some common comparison, and some preexisting "standard" was suggested as a possibility.

In conclusion, the present article not only introduces an ethically unproblematic way of decreasing the speed of the putative "internal clock" in humans, but also shows the utility of this manipulation as a tool for understanding how people actually perform on some timing tasks. The results from the two temporal generalization groups support the analysis of performance that has become standard: On normal temporal generalization, people compare each comparison stimulus with a sample of the standard derived from reference memory, whereas in episodic generalization no reference memory is used (for a quantitative model of performance on episodic generalization, see Wearden, 2004). In addition, results of the manipulation of bisection and verbal estimation performance tell us something about the performance on the tasks that was not known before. On bisection, whatever criterion people are using to classify stimuli as "short" or "long" appears to change as the experiment proceeds, so slowing-down effects are not observed. In contrast, the verbal estimation results suggest some kind of unchanging standard, possibly preexisting the experimental procedure, in accord with other recent suggestions. In general, the slowing-down manipulation used here might be useful in dissecting out the different components making up performance on other timing tasks.

Original manuscript received 22 May 2006
Accepted revision received 17 November 2006

REFERENCES

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

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

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

Burle, B., & Bonnet, M. (1997). Further argument for the existence of a pacemaker in the human information-processing system. *Acta Psychologica*, *97*, 129–143.

Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to the temporal processing of information. *Behavioural Processes*, *45*, 59–72.

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time perception: Implications for clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 195–205.

Droit-Volet, S., & Wearden, J. H. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *Quarterly Journal of Experimental Psychology*, *55B*, 193–211.

François, M. (1927). Contribution à l'étude du sens de temps: La temperature interne comme facteur de variation de l'appréciation subjective des durées [Contribution to the study of the time sense: Internal temperature as a factor in the subjective appreciation of duration]. *L'Année Psychologique*, *27*, 186–204.

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

Hoagland, H. (1933). The physiological control of judgements of duration: Evidence for a chemical clock. *Journal of General Psychology*, *9*, 267–287.

Penton-Voak, E. P., 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.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, *20*, 663–678.

Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705–748.

Wearden, J. H. (1991). 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. (2003). Applying the scalar timing model to human time psychology: Progress and challenges. In H. Helfrich (Ed.), *Time and mind: II. Information-processing perspectives* (pp. 21–39). Gottingen, Germany: Hogrefe & Huber.

Wearden, J. H. (2004). Decision processes in models of timing. *Acta Neurobiologiae Experimentalis*, *64*, 303–317.

Wearden, J. H. (2005). Origines et développement des théories d'horloge interne du temps psychologique [Origin and development of internal clock theories of psychological time]. *Psychologie Francaise*, *50*, 7–25.

Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory: Episodic temporal generalization and bisection in humans. *Quarterly Journal of Experimental Psychology*, *54B*, 289–310.

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

Wearden, J. H., & Grindrod, R. (2003). Manipulating decision processes in the human scalar timing system. *Behavioural Processes*, *61*, 47–56.

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

Wearden, J. H., Philpott, K., & Win, T. (1999a). Speeding up and (...relatively...) slowing down and internal clock in humans. *Behavioural Processes*, *46*, 63–73.

Wearden, J. H., Pilkington, R., & Carter, E. (1999b). "Subjective lengthening" during repeated testing of a simple temporal discrimination. *Behavioural Processes*, *46*, 25–38.

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., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *Quarterly Journal of Experimental Psychology*, *59*, 1709–1724.