THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56B* (4), 321–343

More is not necessarily better: Examining the nature of the temporal reference memory component in timing

Luke A. Jones and J. H. Wearden
University of Manchester, Manchester, UK

Three experiments compared the timing performance of humans on a modified temporal generalization task with 1, 3, or 5 presentations of the standard duration. In all three experiments subjects received presentations of a standard duration at the beginning of a trial block and then had to judge whether each of a number of comparison stimuli was or was not the standard. The duration of the standard changed between blocks. The three experiments varied the experimental design (between or within subjects), task difficulty (how closely the comparison stimuli were spaced around the standards), and presence or absence of feedback on performance accuracy. Number of presentations of the standard never affected the proportion of identifications of the standard when it was presented, nor other features of the temporal generalization gradients observed. The implications for the operation of reference memories within the scalar timing system were explored via models that made different assumptions about how the individual presentations of the standard were stored and used.

Scalar expectancy theory (SET: Gibbon, 1977) and its associated information-processing framework (Church, 1984, 1989; Gibbon & Church, 1984; Gibbon, Church, & Meck, 1984) have enjoyed much success in explaining both animal and human timing behaviour (Allan, 1998). At the heart of the information-processing framework are three major components: a clock process consisting of a pacemaker and an accumulator, a memory process consisting of short-term and reference memory stores, and a comparator process where decisions are made that lead to behavioural output. Using these three components operating together to produce timed behaviour, the SET framework has achieved much success in fitting experimental findings. However, SET's goodness of fit to data is based partly on the flexibility of its overall structure, and properties of its different parts can be altered as necessary to fit data (although these alterations are often plausible), resulting in problems of falsifiability (Wearden, 1999). If

Requests for reprints should be sent to Luke A. Jones, Department of Psychology, University of Manchester, Manchester, M13 9PL, UK. Email: luke.a.jones@stud.man.ac.uk

The perturbation model discussed in the text was based on ideas initially developed with Dr. André Ferrara, as part of a separate project supported by a Joint Project Grant from the Royal Society. We are grateful to Katherine Hill for running the subjects in Experiment 1 for us.

http://www.tandf.co.uk/journals/pp/02724995.html DOI:10.1080/02724990244000287

the properties, in particular, of the memory and decision components of the SET model remain so flexible, then no data can disprove, or even modify, SET. How can this situation be improved? One way (Wearden, 1999) is to try to better specify the operation of the individual components of the SET model (clock, memory, decisions) by manipulating each one separately, leaving everything else constant.

A number of studies have manipulated the clock component of the model both in animals (Maricq, Roberts, & Church, 1981; Meck, 1983) and in humans (Droit-Volet & Wearden, 2002; Penton-Voak, Edwards, Percival, & Wearden, 1996) and have provided evidence for a pacemaker–accumulator clock of the type that SET proposes. Much less attention has been devoted to memory and decision components of the model. Some recent studies have explored timing in the probable absence of reference memory (Allan & Gerhardt, 2001; Rodriguez-Girones & Kacelnik, 1999; Wearden & Bray, 2001), but little research (apart from that of Meck, 1983, with rats) has attempted to manipulate the contents of the reference memory itself.

The reference memory component of the SET model serves two functions. The first is to provide a temporal reference for behavioural stability by storing "important" times. In animal studies the content of reference memory will typically be the remembered time of reinforcement; in human studies the reference memory is assumed to contain memories of standard durations important for the task in hand.

The second role of reference memory is to generate the scalar property of timing: the requirement that the standard deviation of internal "estimates" of time is a constant fraction of the mean. This scalar property is a form of Weber's law and assumes a timing process with constant sensitivity as the interval timed varies. The scalar property is generated in reference memory by the introduction of K^*, a memory transform process intervening between the working memory/accumulator system of the SET model, which reflects the output of the internal clock directly, and reference memory. The concept of K^* has been developed mainly in a series of experiments with rats by Meck and associates (Meck, 1983,1991; Meck & Angell, 1992; Meck & Church, 1987a, 1987b; Meck, Church, & Olton, 1984). The proposed mechanism for generating the scalar property in reference memory is that some presented duration (s) is timed perfectly veridically by the pacemaker/switch/accumulator clock mechanism, but when the contents of the accumulator/short-term memory (Λs) are transferred to reference memory they are multiplied by a memory storage constant, K^*, which varies as a Gaussian distribution with a mean of 1.0. Thus a Gaussian distribution of reinforced times is built up in reference memory by multiple presentations of the standard, s, which is multiplied by K^* every time it occurs. Given that K^* has a mean of 1.0 and some coefficient of variation, c, different s values will effectively be represented in reference memory as a distribution with mean s and coefficient of variation c, so the scalar property of constant coefficient of variation is obtained from the multiplicative nature of K^*.

Although K^* is usually represented as a Gaussian distribution with a mean of 1, this mean value can be experimentally altered (as in Meck, 1983) or differ between individuals or groups. If K^* is equal to 1 then the time stored in reference memory is on average equal to its real time value; if $K^* > 1$ then stored time will be longer than real time; if $K^* < 1$ then stored time will be less than its real time value, on average. Non-veridical values of mean K^* can apparently occur as a result of drug manipulations: For example, physostigmine has been found to decrease the remembered duration of reinforcement times (Meck, 1983), while atropine has been found to

increase them (Meck, 1983). Gibbon, Malapani, Dale, and Gallistel (1997) provide a review of drug effects on K^*.

The age of the subject (rat or human) may also affect the mean value of K^*. Meck, Church, and Wenk (1986) found that old rats (aged 27–30 months) showed a permanently maintained rightward shift in peak time on a peak-interval procedure task, consistent with an increase in K^* (i.e., the time of reinforcement was being stored in reference memory as consistently longer than its real-time value). Meck et al. (1986) also found that repeatedly administering vasopressin (which increases blood pressure and had previously been found by Meck, 1983, to decrease K^*) to young rats (aged 10 months) prevented this age effect from manifesting itself Lejeune, Ferrara, Soffié, Bronchart, and Wearden (1998) found a similar age effect in rats' temporal performance. In their study; 24-month-old and 4-month-old rats were trained on a peak-interval procedure with a time of reinforcement that varied twice between 20 and 40 s. The peak times from the older rats were consistently longer than the reinforcement time and longer than the younger rats' peak time, which tracked the varying time of reinforcement more closely.

More recently, the importance of age as a modulating factor on K^* has been highlighted by work on young children. In a study by Droit-Volet, Clément, and Wearden (2001), temporal generalization performance of three different age groups of children, 3, 5, and 8 years old, was studied. One finding that is relevant here is that the 3-year-olds appeared to consistently misremember the duration of the standard as being shorter than it actually was (see also McCormack, Brown, Maylor, Darby, & Green, 1999). The data from 3-year-olds could be successfully modelled by assuming that their K^* value was on average 0.83–0.88 (depending on whether the standard duration in the temporal generalization task was 4 or 8 s). Furthermore this tendency to have K^* values less than 1.0 was smaller in 5-year-olds whose data could be modelled by assuming a K^* of 0.83–0.90. By 8 years old the K^* value needed to model data was close to that used for normal adults: between 0.9 and 1. Droit-Volet et al.'s (2001) work fits the general hypothesis that K^* is generally <1 for young children and approximately 1 for adults and children over the age of 8 years, and this result complements work with aged rats, showing $K^* > 1$ in these subjects.

A key question in relation to reference memory is how the temporal representations develop, and the aim of the series of studies reported here is to attempt, for the first time, to manipulate the development of temporal representation in a precise way. In the experiments below, we study the effect, if any, on task performance of the number of presentations of a standard in a temporal generalization procedure. In temporal generalization in humans (a procedure developed from a method used with rats by Church & Gibbon, 1982), participants are initially presented with a standard duration (e.g., a 400-ms tone) and then receive comparison stimuli longer than, shorter than, or equal in duration to the standard. After each comparison stimulus presentation they respond YES or NO (standard or not), and feedback as to performance accuracy is usually given (e.g., Wearden, 1991, 1992; Wearden & Towse, 1994). In this "normal" temporal generalization, the value of the standard duration remains constant for the entire experimental session.

The usual theory of temporal generalization assumes (1) that representations of the standard (s) enter into a reference memory, after being transformed by K^*, as described above, (2) that the comparison duration presented on a particular trial, t, is timed without error, and (3) that the YES/NO response is generated by a comparison between t and a sample drawn from

the reference memory on that particular trial, s^*. In temporal generalization with animals (Church & Gibbon, 1982), extensive training with the standard duration is given, so the idea that reference memory accumulates (e.g., in the form of a Gaussian distribution with accurate mean and scalar variance) as a result of this extensive experience is a very natural one. In studies with humans, however, much less "training" is given, and, as is seen below, humans can perform on this task when just one example of the standard has been given. The obvious question is what effect multiple presentations of the standard duration have, if they have any.

A priori, successive temporal memories could be stored in at least two distinct ways. In one of these, the successive individual examples could be aggregated together in some way (e.g., averaged into some mean time value), so increasing numbers of successive presentations of a standard would tend to make the aggregate value more representative or accurate than individual examples were. In this case, the contents of temporal memory would be transformed versions of the individual standard values provided. In contrast, another possibility is that the different examples of the standard are stored completely separately and then sampled. In this case, the individual identity of each previously presented standard would be maintained in memory, and the contents of reference memory might not statistically change with increasing numbers of standard presentations. Depending on how these separately stored examples of the temporal reference are used, increasing numbers of presentations of the standard may not necessarily change the temporal representation used. In the present article, we present three experiments, which examine the effect of changing the number of standards in temporal reference memory empirically; we then subsequently present results derived from computer modelling of some of the possibilities outlined above to explore various potential structures and operations of temporal reference memory.

EXPERIMENT 1

To explore the effect of number of standard presentations on temporal generalization, a procedural variant of the normal method had to be developed. In a typical temporal generalization task for humans (e.g., Wearden, 1992) a standard duration is presented a few times (3 to 5) at the start of the experimental session, and this standard remains in force for the entire session. The subject then receives trials in which comparison durations (a mix of durations shorter than, longer than, or equal in duration to the standard) are presented, and must judge whether each duration was the standard or not, with accurate performance-related feedback being given. However, with this method, the number of standard presentations cannot be effectively controlled, as subjects will use the feedback they are given after different comparison durations (both presentations of the standard and other durations) to regulate their representation of the standard. One possibility was to use the normal method but without feedback, but this has two drawbacks. First, there is evidence that in temporal generalization without feedback the representation of the standard changes systematically throughout the experimental session (Wearden, Pilkington, & Carter, 1999); second, the fact that the standard remains constant throughout the session may encourage the subject to "construct" the standard using judgements of some of the nonstandard durations (e.g., "the standard is shorter than this one, but longer than this one"), or by using as a standard the mean of all the comparison stimuli. If this were done, then our aim of manipulating the number of presentations of the standard would be obviously compromised.

The method we developed was a "changing standard" temporal generalization method, where the standard in force changed for every experimental block. Each experimental block contained a small number of comparison stimuli (but more than one, cf. "episodic" temporal generalization: Wearden & Bray, 2001). The subject was informed of this change so, presumably, he or she only responds on the basis of the standard currently in force for the block. Using this method, we manipulated the number of standard presentations in the block, so the subject received 1, 3, or 5 identical presentations of a standard, and then responded to a small number of comparison durations (which included durations shorter than, longer than, or equal in duration to the standard). If the different presentations of the standard were aggregated together in some way, then there might be some observable effect of the number of presentations of the standard.

Method

Participants and apparatus

Twenty-four Manchester University undergraduates were randomly assigned to three equal sized groups. A Hyundai 386 (IBM compatible) computer controlled all experimental events. The computer speaker produced the stimuli to be judged, and the keyboard measured participant's responses. The program used to run the experiment and record data was written in the MEL language (Micro-Experimental Laboratory: Psychology Software Tools, Inc.), thus providing millisecond accuracy for timing of stimuli and responses.

Procedure

The procedure for the three experimental groups (1-standard, 3-standard, 5-standard) was almost identical except for the number of presentations of the standard. Consider first the procedure for the 1-standard group. Participants received 10 blocks each consisting of one standard presentation and seven test trials. The standard was drawn from one of two equally likely uniform distributions: a distribution running from 200–400 ms (short-duration range) and another running from 400–800 ms (long-duration range). Order of presentation of different duration ranges was randomized for each participant. The participant received a single presentation of the standard duration (e.g., 400 ms) in the form of a 500-hz tone, following a display stating that a single standard duration would be given. Following standard presentation, participants received comparison tones whose duration was the standard duration (whatever it was on that block) multiplied by 0.25, 0.5, 0.75, 1, 1.25, 1.5, or 1.75, with the order randomized. Each comparison stimulus presentation followed a response to a "press spacebar for next trial" prompt by a delay that was a random value picked from a uniform distribution running from 750 to 1250 ms. After the 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. No feedback as to response correctness was given. Following presentation of all seven comparison stimuli, a new standard was randomly generated for the next block and so on. Participants had been informed by previous instructions that the standard would change for each block.

The procedure for the 3-standard and 5-standard groups was identical to that for the 1-standard group except that the standard duration was presented 3 and 5 times, respectively, at the beginning of each block. Between each presentation of the standard there was an interval drawn from a uniform distribution running from 1500–2000 ms, offset to onset.

Results

Figure 1 shows temporal generalization gradients—the mean proportion of YES responses (identification of a presented comparison duration as the standard)—plotted against comparison/standard ratio. The upper, middle, and bottom panels show data for the 1-, 3-, and 5-standard groups respectively. In each panel data are shown separately for comparisons from the short and long duration ranges.

Inspection of the data in Figure 1 suggests that the maximum proportion of YES responses occurred at the standard duration in some of the conditions only: In others, the temporal generalization gradients peaked at the comparison duration which was 75% of the standard. There seemed little obvious effect of either duration range or number of presentations of the standard on the proportion of YES responses that occurred when the standard was actually presented, which ranged from .55 to .82, in different conditions. Another suggestion was that data from the short and long duration ranges did not superimpose, which is highlighted by the fact that the plots shown in Figure 1 display data from the two ranges on the same relative scale.

These suggestions were confirmed by statistical analyses. Consider first the proportion of YES responses occurring when the standard duration was presented. Kruskal–Wallis tests found no effect of number of presentations on the mean proportion of YES responses to the standard when it was presented, for the short, $X^2(2) = 0.35$, $p = .84$, or the long, $X^2(2) = 0.16$, $p = .93$, duration ranges.

Consider next analyses of the whole temporal generalization gradients. A repeated measures analysis of variance (ANOVA) used duration range (short or long), and comparison/standard ratio (effectively the duration of the comparison) as within-subject factors, and number of standard presentations (1, 3, or 5) as the between-subjects factor. There was no effect of number of standard presentations, $F(2, 21) = 0.22$, $p = .808$, but there was a significant effect of comparison/standard ratio, $F(6, 12) = 57.25$, $p < .001$, indicating that the subjects were sensitive to comparison duration. There was no number of standard presentations by comparison/standard ratio interaction, $F(12, 126) = 0.64$, $p = .658$, suggesting that the number of presentations of the standard did not significantly affect the shape of the temporal generalization gradients. There was a significant main effect of duration range, $F(1, 21) = 29.72$, $p < .0001$, and a significant duration range by comparison/standard ratio, $F(6, 126) = 5.32$, $p < .001$, indicating a failure of superimposition of temporal generalization gradients from the short and long duration ranges.

Discussion

Experiment 1 failed to find any effect of the number of presentations of the standard on accuracy of identification of the standard when it was presented, overall level of YES responses, or the shape of the temporal generalization gradients. Our data did, however, have a number of features unusual for temporal generalization performance assessed by the "normal" method (e.g., Wearden, 1992), as in many cases the peak of YES responses was not at the standard, and generalization gradients from the short and long standard ranges did not superimpose. Because of these differences from the results usually obtained with the normal temporal generalization method we decided to repeat the experiment with a number of changes.

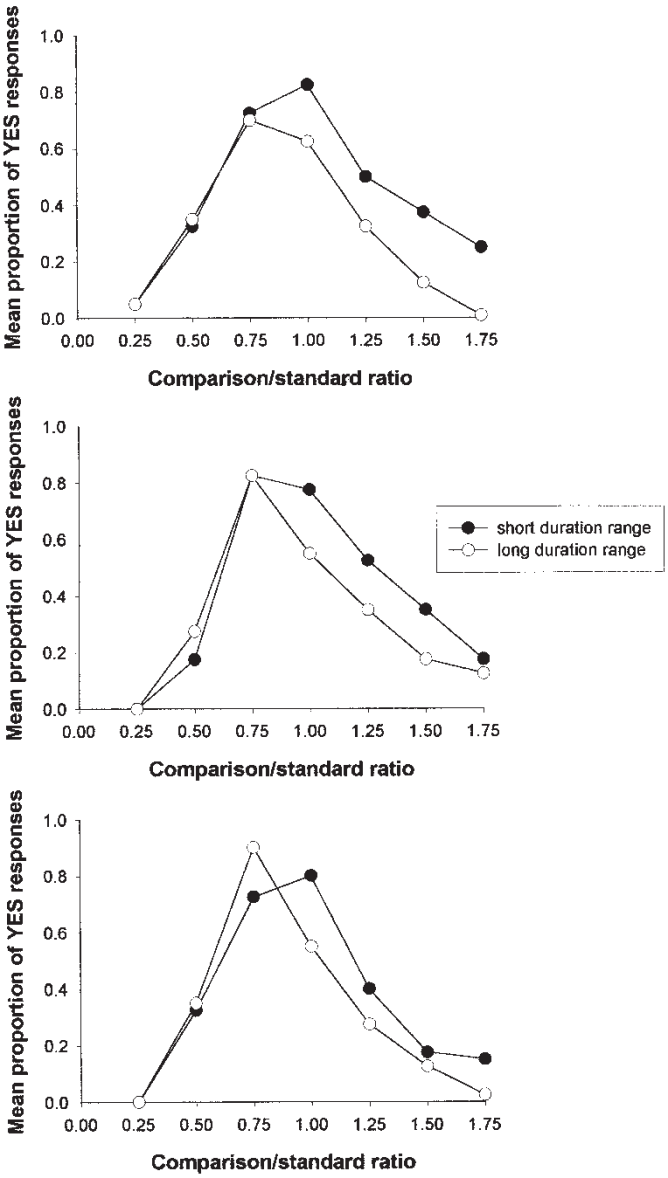

Figure 1. Temporal generalization gradients (proportion of identifications of a duration as the standard, i.e., YES responses, plotted against comparison stimulus durations expressed as fraction of the standard duration) for each of the three conditions (1-, 3-, 5-standards; top, centre, and bottom panels, respectively) and for each duration range (short or long) of Experiment 1.

For Experiment 2, we used 24 subjects and employed a within-subjects design to increase sensitivity and the number of subjects in each condition. Second, we increased the difficulty of the task, in the hope of increasing subjects' temporal acuity, as it has been shown in previous experiments (Ferrara, Lejeune, & Wearden, 1997) that increasing task difficulty results in higher temporal acuity. Difficulty was increased by reducing the spacing of the comparison durations around the standard, and in Experiment 2 the comparison/standard ratios were 0.4, 0.6, 0.8, 1, 1.2, 1.4, and 1.6.

EXPERIMENT 2

Method

Participants and apparatus

Twenty-four Manchester University undergraduates participated, and the apparatus was that used in Experiment 1.

Procedure

Each participant received a single experimental session lasting approximately 45 min. This consisted of three conditions (1, 3, 5-standard presentations) with 10 blocks in each condition and seven trials in each block. The standard was drawn from one of two equally likely uniform distributions; short range running from 300–500 ms and a long range running from 600–1000 ms. Order of presentation of the duration ranges was randomized for each participant. The order of the three conditions was counterbalanced with 4 participants being pseudorandomly assigned to each of the six possible orders. The procedure for the 1-standard condition was almost identical to that for the 1-standard condition of Experiment 1. The participants received the standard presentation, followed by comparison durations, which were the standard duration multiplied by 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, and 1.6. All other experimental details were as those for Experiment 1. The procedure for the 3-standard and 5-standard conditions was identical except that the standard duration was presented 3 and 5 times, respectively, at the start of each block.

Results

Figure 2 shows the temporal generalization gradients from Experiment 2. The upper, middle, and bottom panels show data for the 1-, 3-, and 5-standard conditions, respectively. In each panel data are shown separately for the short and long duration ranges.

Inspection of the data in Figure 2 shows that the maximum proportion of YES responses occurred after presentation of the standard for most conditions. Proportions of YES responses at the standard ranged from .76 to .85, but showed no obvious effects of duration range or number of presentations. Once again generalization gradients from the short and long duration ranges did not appear to superimpose.

These suggestions were confirmed by statistical analysis. Consider first the proportion of YES responses occurring when the standard duration was presented. Taking the short duration range first, a Friedman test found no effect of number of presentations on the mean proportion of YES responses to the standard when it was presented, when comparing the 1-, 3-, and 5-standard conditions, $X^2(2) = 0.56$, $p = .756$. For the long duration range there was also no significant effect of number of presentation on the mean number of YES responses to the standard when it was presented, although it did approach significance, $X^2(2) = 5.56$, $p =$

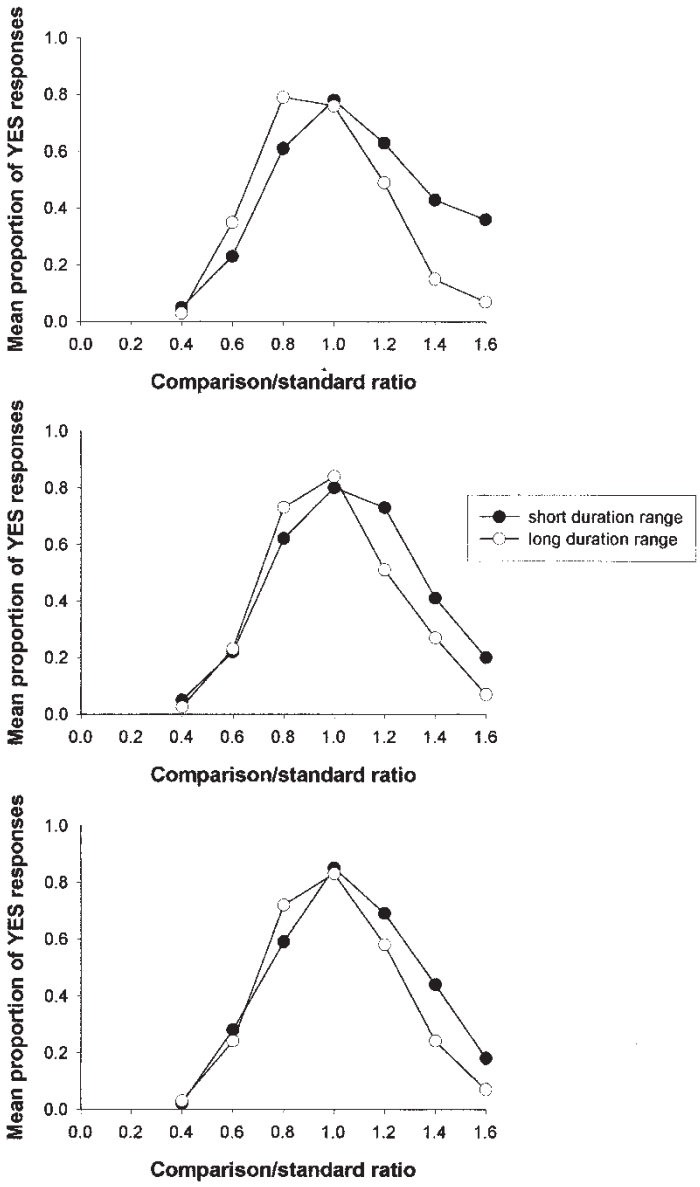

Figure 2. Temporal generalization gradients for each of the three conditions (1-, 3-, 5- standards; top, centre, and bottom panels, respectively) and for each duration range (short or long) of Experiment 2.

.062. As this result was approaching significance we conducted a Wilcoxon repeated measures test, but no significant difference was found between any of the pairwise comparisons—for example, when 1 standard was compared to 3 standards, $Z = -1.77$, $p = .08$, when 1 standard was compared to 5 standards, $Z = -1.27$, $p = .20$, or when 3 standards were compared to 5 standards, $Z = -0.28$, $p = .78$.

Consider next analyses of the whole generalization gradients. A repeated measures ANOVA used the duration range (short or long), comparison/standard ratio, and number of standard presentations (1, 3, 5) as within-subject factors. There was no effect of number of standard presentations, $F(2, 46) = 0.62$, $p = .54$, but there was a significant effect of comparison/standard ratio, $F(6, 138) = 125.28$, $p < .001$. There was no number of standard presentations by comparison/standard ratio interaction, $F(12, 276) = 1.22$, $p = .27$, suggesting that the number of presentations did not significantly affect the shape of the temporal generalization gradient. There was a significant main effect of duration range, $F(1, 23) = 28.09$, $p < .01$, and a significant duration range by comparison/standard ratio interaction, $F(6, 138) = 8.56$, $p < .01$, suggesting that the temporal generalization gradients for the short and long duration ranges did not superimpose.

Discussion

Experiment 2 replicated the principal findings of Experiment 1 in that there was no evidence of an effect of number of presentations of the standard on identification of the standard when presented, overall level of responding, or the shape of the temporal generalization gradient. In Experiment 2, most temporal generalization gradients peaked at the standard, although superimposition of gradients from the short and long duration ranges was still elusive. In Experiments 1 and 2 we did not use feedback to try to discourage subjects from "constructing" the standard using information gained after their responses to the comparison stimuli, but it may be that the lack of feedback was wholly or partly responsible for the rather unusual failure of superimposition in our data. To test this idea, we performed Experiment 3, which was essentially a replication of Experiment 2 with added performance-related feedback after the response to each comparison stimulus.

EXPERIMENT 3

Method

Participants and apparatus

Twenty-four Manchester University undergraduates participated, and the apparatus was that used in Experiments 1 and 2.

Procedure

Each participant received a single experimental session lasting approximately 45 min. The procedure was identical to that of Experiment 2 except that accurate feedback was given after the response to each comparison stimulus. After the participants feedback 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 2000 ms, the participants were prompted to press the spacebar to receive the next trial.

Participants received the three different conditions (1-, 3-, and 5-standards) in an order that was counterbalanced over the group as a whole, with four participants assigned to each of the six possible orders of conditions.

Results and discussion

Figure 3 shows the mean proportion of YES responses (identification of a presented stimulus duration as the standard) plotted against comparison/standard ratio. The upper, middle, and bottom panels show data for the 1-, 3- and 5-standard conditions, respectively. In each panel data are shown separately for the short and long duration ranges.

Inspection of the data in Figure 3 shows that the maximum proportion of YES responses occurred after presentation of the standard duration in all conditions. The proportion of identifications of the standard ranged from .61 to .71, but showed no obvious effects of the number of presentations of the standard, nor of the duration range. Although the temporal generalization gradients from Experiment 3 all peaked at the standard duration, in general it seemed that superimposition was usually absent when the short and long duration ranges from a particular number of standard presentations were compared.

These suggestions were confirmed by statistical analysis. Consider first the proportion of YES responses occurring when the standard duration was presented. Taking the short duration ranges first, a Friedman test found no effect of number of standard presentations on the mean proportion of YES responses to the standard when it was presented, when comparing the 1-, 3- and 5-standard conditions, $X^2(2) = 4.33$, $p = .11$. For the long duration ranges there was also no significant effect of number of presentation on the mean proportion of YES responses to the standard when it was presented, $X^2(2) = 0.75$, $p = .96$.

Consider next analyses of the whole generalization gradients. A repeated measures ANOVA used the duration range (short or long), comparison/standard ratio, and number of standard presentations (1, 3, 5) as within-subject factors. There was no effect of number of standard presentations, $F(2, 46) = 0.12$, $p = .89$, but there was a significant effect of comparison/standard ratio, $F(6, 138) = 97.98$, $p < .001$. There was no number of standard presentations by comparison/standard ratio interaction, $F(12, 276) = 0.30$, $p = .97$, suggesting that the number of presentations did not significantly affect the shape of the temporal generalization gradient. There was a significant main effect of duration range, $F(1, 23) = 6.06$, $p < .05$, and a significant duration range by comparison/standard ratio interaction, $F(6, 138) = 4.88$, $p < .01$, suggesting that the temporal generalization gradients for the short and long duration ranges did not superimpose.

The aim of Experiment 3 was to replicate the findings of Experiments 1 and 2. In addition we wished to ascertain whether the lack of superimposition found in Experiments 1 and 2 was due to a lack of feedback so we introduced feedback into the procedure. The findings of Experiments 1 and 2 were replicated in that there was no evidence of an effect of

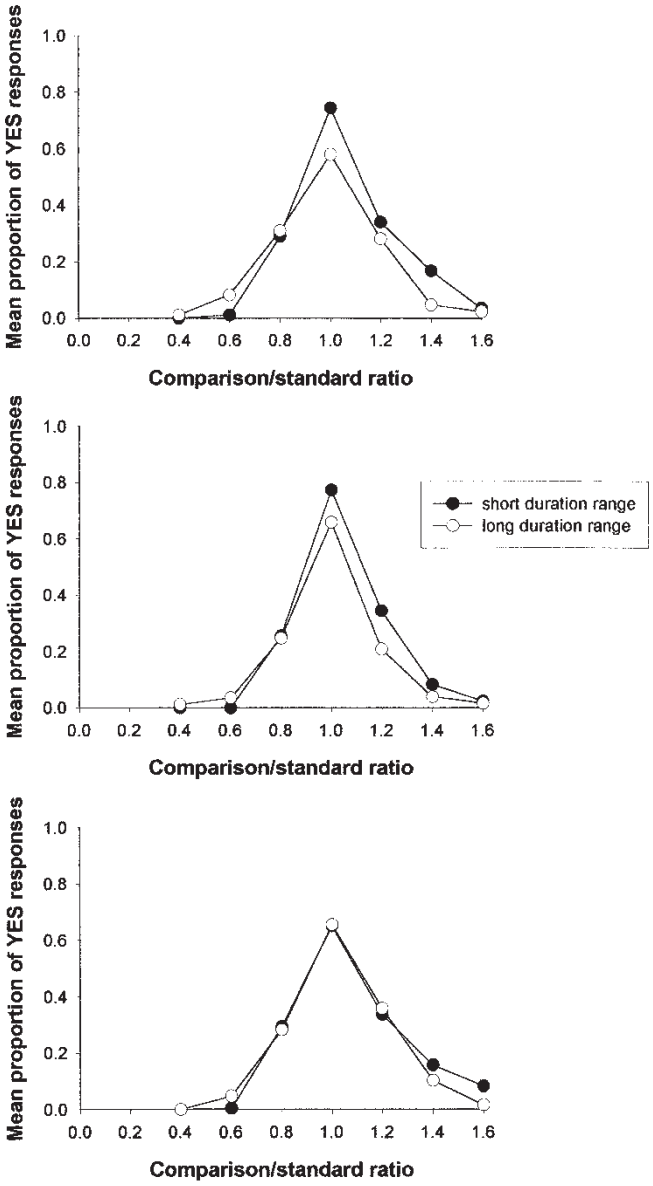

Figure 3. Temporal generalization gradients for each of the three conditions (1-, 3-, 5-standards; top, centre, and bottom panels, respectively) and for each duration range (short or long) of Experiment 3.

presentation number on identification of the standard when presented, overall level of responding, or shape of the temporal generalization gradient. In addition we also replicated the failure of superimposition in that there was a significant effect of duration range overall. However, inspection of Figure 3 suggests that data from the short and long duration ranges did superimpose in the 5-standard conditions, but not when 1 and 3 standards were given.

GENERAL DISCUSSION

Although differing in some minor respects, all our three experiments tell a consistent story: Increasing the number of presentations of the standard from 1 to 5 had no significant effect either on the proportion of correct identifications of the standard when it was presented or on other features of the temporal generalization gradients. This was true whether the standard came from the short or long duration ranges in all the experiments, whatever the spacing of the comparison durations around the standard, and whether or not feedback was given. On the other hand, subjects were highly sensitive to the duration of the comparison stimuli relative to the standard in all conditions of the modified temporal generalization procedure that we used in this article. A subsidiary finding was a persistent failure of superimposition when duration range was varied.

The fact that repetition of temporal generalization standards does not "improve", or in any way seem to markedly affect, temporal generalization performance seems, at least at first sight, intuitively surprising: Most people would suppose that repeatedly presenting an "item" to be remembered would certainly improve memory performance for that item. However, whether this is true or not depends on how the repeated presentations of the item are processed, as is seen below.

To explore the effect of the number of presentations of the standard theoretically, we conducted a number of computer simulations of temporal generalization performance, which, in essence, varied the way that the presented standards were stored and used, keeping all other features of the simulation constant. The two models that we explored initially differed only in the way that the presentations of the standard were stored. In one of the models (*average* or AVE), the repeated presentations of the standard were averaged together to produced a mean value (the arithmetic mean of all the standards presented), and this mean was used to make judgements about the relation between the comparison durations and the (mean) standard. In the other model (*sampling* or SAM), the repeated presentations of the standard were stored separately in an array, rather than being averaged, and on each trial a value from this array was randomly sampled and compared with the comparison duration on the trial.

The SAM model is actually an embodiment of the standard ideas incorporated into SET about how reference memory is formed. The exact nature of reference memory has received relatively little discussion in treatments of SET, but Gibbon, Church, Fairhurst, and Kacelnik (1988), and Brunner, Fairhurst, Stolovitsky, and Gibbon (1997) provide two sources where the contents of reference memory are discussed explicitly. In a condition where reinforcement is available for responses at a single time, s, which remains constant throughout the experiment, each presentation of s is assumed to be transformed by K^* and then added to the reference memory. The transformed values are merely aggregated together to form a distribution: They are not averaged or otherwise transformed. For simplicity, the reference memory is

assumed to be represented as a Gaussian distribution, but this is a mathematical convenience only, and the continuous Gaussian distribution arises because of the accumulation of a large number of instances of presentations of reinforcers at s. To produce behaviour on any particular trial, a single sample, s^*, is drawn randomly from the Gaussian distribution and used as the effective standard on the trial. The fact that statistical distributions and mathematical analysis are used to derive predictions from the model (see Brunner et al., 1997, and Gibbon et al., 1988, for example) partly conceals the fact that the reference memory is built up by accumulation of separate instances of experience of reinforcement at time s. Our SAM model uses exactly the same principles, except that each instance of s is transformed by K^* and then stored separately as an array element.

In all our simulations, the standard duration was 400 ms, and comparisons were 160, 240, 320, 400, 480, 560, and 640 ms. If the standard in force on each comparison trial was s^* (and this was generated in a different way in the two models), then the model responded YES when $|s^*-t|/t < b^*$, where s^* is the sample drawn from reference memory, t is the just-presented duration (comparison), and b^* is a threshold value, which is variable from trial to trial. The focus of interest of the present simulations was solely on the effect of number of presentations of the standard on temporal generalization performance, so other parameters of the model were kept constant at values similar to those needed to accurately simulate temporal generalization performance in humans in other work (Ferrara et al., 1997; Wearden, 1992; Wearden, Denovan, Fakhri, & Haworth, 1997). The threshold mean used, b, was 0.22, and the coefficient of variation of the threshold was 0.5, so the threshold varied randomly from trial to trial, taking a value b^*, which was a value sampled randomly from a Gaussian distribution with a mean of 0.22 and a standard deviation of 0.11. The comparison duration, t, on the trial, was assumed to be timed without error (i.e., to have its real-time value).

When a standard, s, was presented, this was transformed into s' by being multiplied by a memory constant, K^*. K^* was assumed to have a mean of 1.0 and a coefficient of variation, in our simulations, of 0.18, and it was represented as a Gaussian distribution of values. The standard, s (400 ms) in our simulation, was thus transformed into different s' values every time s was presented. The AVE and SAM models differed in the subsequent fate of the s' values. In the AVE model, the n s' values were just averaged together, and their mean was used as the s^* for all judgements of the relation between s^* and t on that block, using the decision rule and other details as specified above. For the SAM model, the s' values were stored separately in an array, and, for each comparison stimulus presented on that block, a value was randomly sampled from the array and used as s^* on that trial.

The different simulations employed different numbers of presentations of standard, from 1 to 100, although only values up to 10 are reported, as increasing the number of presentations above 10 had no further effect. The simulations were run in two ways. In the first (10,000 sample), 24 independent runs of the model each made 10,000 judgements of the equality of the standard and each different t value used. This showed the results of simulations with all sorts of random variability averaged out over many trials, so shows "asymptotic" performance of the models. In addition, we also ran 24 independent runs of the model where 10 comparisons of the standard with each comparison stimulus were made (10 sample)—approximately the same amount of results per run as observations collected from subjects in the experiments reported above. The 10,000 sample simulations thus showed what the performance of the models was like at a limit; the 10 sample simulations showed

what the performance of the models was like when a realistically sized sample of results was simulated.

We looked at two principal measures of performance, which relate to the data analysis conducted earlier. The first was how the proportions of identifications of the standard when it was presented changed when the number of presentations of the standard was varied. Figure 4 shows the results, for the AVE and SAM models with the 10,000 sample and the 10 sample. The number of presentations of the standard varied from 1 to 10 (and results from 100 presentations, not shown in the Figure, were essentially the same as those with 10).

Inspection of Figure 4 shows the main result of interest. With the averaging model (AVE), the proportion of YES responses to the standard duration systematically increased with increasing numbers of standard presentations, and the increase was marked between 1 and 5 presentations both in the 10,000 sample and in the 10 sample. This suggests that such an effect would be readily observable in data (recall that the 10-sample simulations produced about the same number of observations as did the subjects tested). On the other hand, the SAM simulation showed no such change in proportion of YES responses with increasing number of standard presentations, either in the 10 or in the 10,000 samples.

Comparison of the simulated results with data suggests strongly that the data were much more compatible with the SAM than the AVE model, as the proportion of YES responses at the standard duration was never significantly affected by the number of standard presentations in any of our experiments.

Figure 5 shows temporal generalization gradients obtained from the simulations. The upper two panels show results from the 10,000 samples, and the lower two panels show results from the 10 samples (AVE and SAM, left and right panels, respectively). It is clear that, in the AVE model, increasing the number of standard presentations sharpens the generalization gradient, which becomes more peaked (see Figure 5) and produces systematically fewer YES

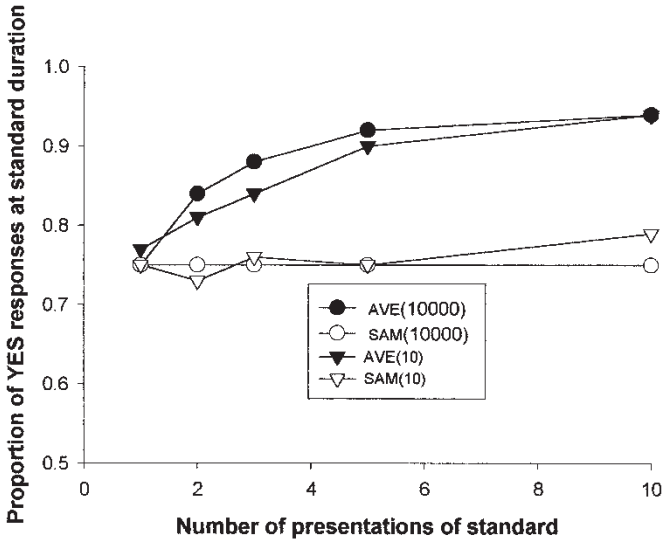

Figure 4. Proportion of YES responses at standard duration plotted against number of presentations of standard for the AVE and SAM models, and the 10,000 and 10 sample simulations output for each.

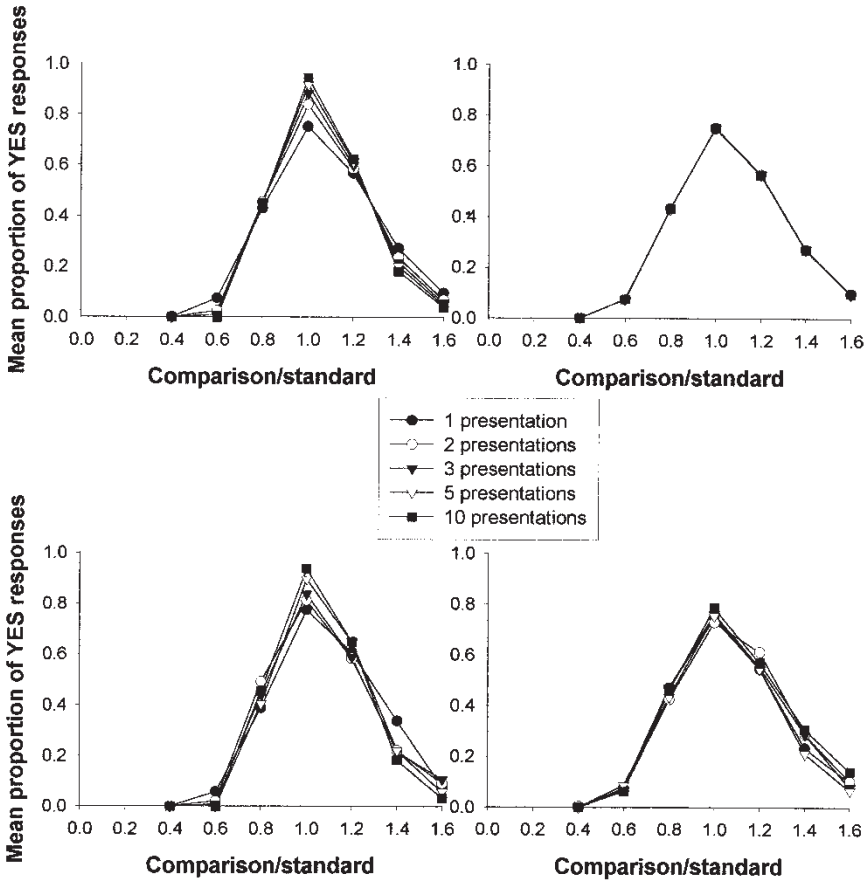

Figure 5. Temporal generalization gradients obtained from the model simulations. Upper panels: Results from 10,000 samples. Lower panels: Results from 10 samples. Left panels: AVE model; right panels: SAM model.

responses at stimulus durations remote from the standard. The effect is, obviously, clearest in the 10,000 sample, but is observable in the 10 sample. On the other hand, the SAM model shows no effect of the number of presentations of the standard duration (see the 10,000 sample, upper right panel of Figure 5). Once again, comparison with the data suggests that the SAM model fits better, as in our experiments there was never a significant effect of the number of presentations of the standard on the shape of the temporal generalization gradient overall.

We have described the AVE and SAM models in parallel and highlighted that the sole difference between them is in the way that the presentations of the sample are stored and used after transformation. However, the SAM model is logically much simpler than our exposition would initially imply: In fact, the SAM model is always equivalent to using a single s' (i.e., transformed s) value as s^*, no matter how many are presented and accumulated, so the number of presentations of s has no systematic effect. If this seems intuitively surprising, consider the following argument. Some number of presentations of s, n, is presented and stored separately in an array of n items. On each trial, one of these n items is

picked as s^* and used for the current decision. But this s^* is just a random sample from an array that is itself a random sample of values generated from transformation of s by the memory constant, so it is equivalent to just taking a single s value and transforming it using K^* into s'. The number of items stored is irrelevant.

As mentioned above, the SAM model, by showing no effect of the number of presentations of the standard on behaviour, is more compatible with the data obtained in our three experiments than assuming that presentations of the standard are averaged together. If K^* is used to transform some presented standard, s, into a value used on a particular subsequent trial (s^*), then the random model suggests that this process by itself may be sufficient to account for many properties of observed timing behaviour, without the need for an extensive reference memory containing the individual s^* values accumulated over previous trials. If these values are merely sampled randomly, then storing a single one that is different on every trial has the same consequences as storing tens or hundreds and sampling randomly amongst them, as our simulations with the SAM model show. Thus the temporal reference memory might be supposed to contain just a single item, the s^* used on the current trial, which is just the s value (transformed by K^*) from the previous trial. Simulations we have conducted of performance on some other timing tasks, such as the peak–interval procedure (Roberts, 1981), suggest that, as in the present case, data from steady-state conditions can be simulated accurately by the assumption that only a single item is stored in reference memory.

The SAM model embodies the orthodox position of SET about the formation and content of reference memory, so what is true of the SAM model is also true of the mathematical embodiments of reference memory in SET (e.g., Gibbon et al., 1988). In one sense, there is no need for more than one item to be stored in reference memory: The use of a Gaussian K^* to transform each s into an s^*, which is then used to control behaviour on the trial, would in fact be statistically equivalent over a large number of trials to storing the transformed items in a reference memory distribution and then sampling a value from this distribution as s^* on each trial. The trial-by-trial variability in the representation of s is in fact controlled by K^* and not by any storage mechanism.

Is there, then, any evidence that the reference memory contains more than one item? In fact, almost the only such evidence comes from the study of transitions from one time of reinforcement to another in studies with animals. Various studies by Staddon and colleagues (e.g., Higa, 1996; Higa, Wynne, & Staddon, 1991), as well as other work by Lejeune, Ferrara, Simons, and Wearden (1997), suggest that transition from one time of reinforcement to another in animals is very rapid, taking place within a very small number of experiences with the new reinforcement time, possibly even as few as one. On the other hand, manipulations of remembered times of reinforcement by drugs (e.g., Meck, 1983) appear much slower, taking a number of experimental sessions (involving tens of experiences of the reinforcement time in the drug state). We might consider these two manipulations to be equivalent: In one case the time of reinforcement is changed by actually changing the reinforcement time; in the other the time of reinforcement is changed by the time being "misremembered" as shorter or longer than it really is, because of the effect of the drug. One difference between the sorts of experiments is that the actual changes of reinforcement time (e.g., Lejeune et al., 1997) are often substantial (e.g., from 20 to 30 s), whereas the effects of drugs are smaller, in the order of 10%. An implication of this is that large changes in times of reinforcement are reacted to rapidly, whereas smaller changes take longer to change behaviour.

Models that assume that experiences of previous times of reinforcement are either averaged together to produce some sort of mean (as our AVE model), or stored individually and sampled (our SAM model), have difficulty in dealing with these transition effects. Any kind of averaging would tend to imply that all transitions would be slow, as new instances contribute more or less to an average, depending how many items are averaged together, or how much the most recent item are weighted compared with previous items. On the other hand, if many instances of reinforcement times are stored, and one is sampled to be used on the current trial, the rapidity of transition will depend on the number of items stored and sampling biases (such as a tendency to pick more recent values). At a limit, if a single item is stored, then the transition (in response to both small and large shifts in reinforcement time) will be rapid.

One resolution of these problems might reside in a slightly different model of temporal reference memory from those presented above, which we will only sketch out here: the *perturbation model*. The perturbation model operates as follows. Some standard duration is presented on a trial (s), and this is transformed by K^* into an effective value for behaviour on a subsequent trial, s^*. The value s^* is stored not as a single value, but as a distribution having upper and lower bound; so, for example, the limits of the distribution might be proportional to s^* and run from $0.9s^*$ (lower bound) to $1.1s^*$ (upper bound, with 10% limit), or from $0.8s^*$ to $1.2s^*$ (20% limit), and so on. On the next trial, another s^* value is generated (s^*_{new}), and if s^*_{new} is within the upper and lower bounds of the current s^*, then effectively nothing happens, and the old s^* value (and its upper and lower bounds) is maintained. If, on the other hand, s^*_{new} is *outside* the bounds of the old s^*, then there is "perturbation" of the temporal reference memory, and in the simplest form of the model the s^*_{new} just displaces the old s^* and creates new upper and lower bounds (e.g., 10% above and below the new s^*).

The perturbation model economically achieves stability of temporal reference when conditions remain stable, but at the same time can react flexibly to changes in the time of reinforcement. Furthermore, large changes in the time of reinforcement (occasioned by shifts in delivered reinforcement times) will be reacted to quickly (as the new time value will almost always be outside the bounds of the old one), whereas small effective changes in reinforcement time (as produced by drug manipulations) will take much longer to produce a consistent change in the temporal memory, as many instances will be within the limits of the previous time reference and will therefore leave the reference memory "unperturbed".

The full ramifications of this sort of approach to temporal reference memory are beyond the scope of the present article, and tests of the model are probably best conducted under conditions where there are transitions from one temporal standard to another, but the perturbation model is broadly consistent with the results obtained in the experiments above. Presenting repeated temporal standards may have no effect in our studies, not because only the most recent time value is stored, but because each successive temporal standard falls within the upper and lower bounds of the representation of the previous one and thus effects no change.

A simple form of the perturbation model can be applied to the present data. All rules for generating behaviour were as those in the models discussed above, but the effective standard for the seven comparisons in the block (s^*) was generated differently. When the first standard was presented, this was transformed into s^*, and upper and lower bounds were set (in our simulations either 10% or 20% above and below s^*). If a subsequent s^* value, resulting from

another presentation of the standard, fell within these limits there was no change in s^*; if it was outside the current bounds, the new s^* value was substituted for the old, and 10 or 20% limits around this new value were generated. If s^* remained constant over a number of blocks, then variance was generated not only by the threshold (as above), but also by sampling randomly from a uniform distribution running from the lower to the upper bound of the s^* limit. For the simulations shown in Figure 6, K^* was 1.0, but with a coefficient of variation of 0.18, b was 0.22 with a coefficient of variation of 0.5, and the limits for the perturbation were either 10% or 20% of the current s^*. The 10,000 trials were simulated at each comparison stimulus value

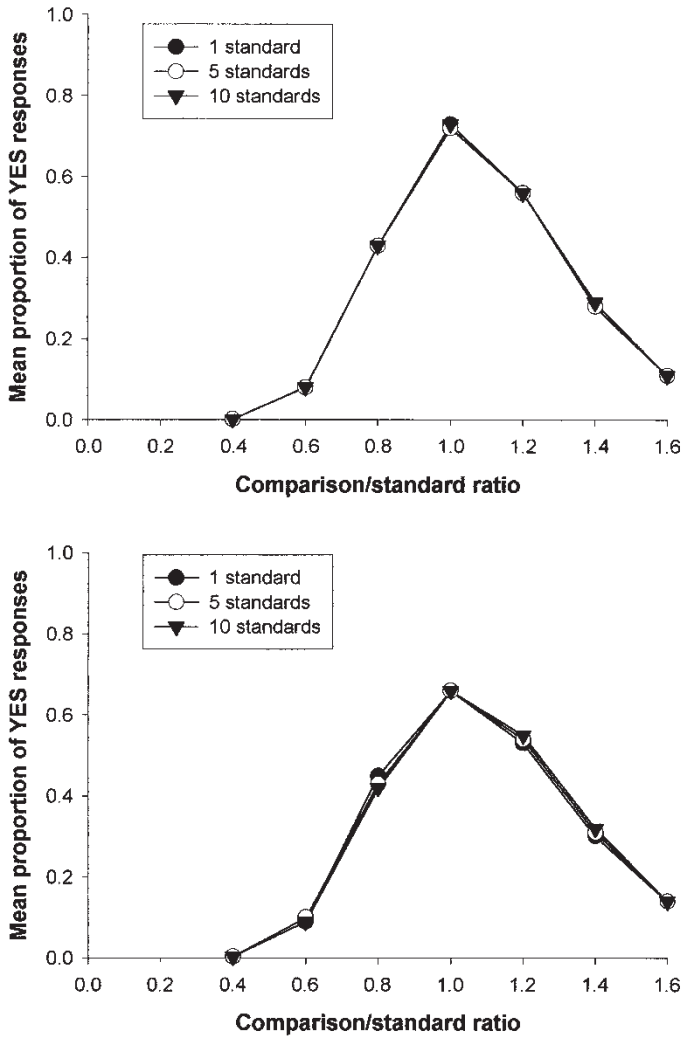

Figure 6. Temporal generalization gradients from the perturbation model with 10,000 trials simulated at each comparison stimulus value, with 1-, 5-, and 10-standard presentations. Upper panel: Results from simulation with 10% perturbation limit: Lower panel: Results from simulation with 20% perturbation limit.

(160 ms to 640 ms, with a 400-ms standard, as the other simulations, above), and results from 1-, 5-, and 10-standard presentations are shown in Figure 6.

The upper panel of Figure 6 shows results with a 10% limit, the lower panel results with a 20% limit. In neither case was there any effect of the number of presentations of the standard on any aspect of the temporal generalization gradient (and this was true even with 100 standard presentations, although results from these conditions are not shown). Not only is the perturbation model compatible with the data from the present experiments (as well as aspects of results obtained in transition states), but it may also explain why the SAM model fits our data. In effect, in conditions where the time of reinforcement is kept constant, the SAM model and the more sophisticated perturbation model may be effectively the same. If a series of similar standard durations is presented (so the K^* transformation produces values that are usually within the limits of previously stored standards), then effectively the standard used on a trial may just vary randomly from one trial to another, much as the SAM model predicts. The SAM model may seem counterintuitive: If humans are presented with a series of examples of the standard it may seem strange to suppose that they do nothing at all with these, almost as if all but one passes unnoticed. However, the perturbation model would suggest that *all* standards presented are processed, but if a standard falls within the limits of the current s^*, then no perturbation and hence no change in the effective standard duration occurs. As in some models of conditioning (e.g., Pearce & Hall, 1980; Rescorla & Wagner, 1972), only "surprising" events (in our case standards falling outside the limit of the current standard) produce behaviour change.

Although we presented three experiments with consistent results above, perhaps the main contribution of the present article is that it argues for a reconceptualization of the nature of temporal reference memory, in humans and (most probably) also in animals. This reconceptualization considers the reference memory not as an extensive store of previous experiences, but as a much smaller store, either containing a single item or, as in the perturbation model, as a single item with upper and lower boundaries, which control whether or not change occurs as a result of some new experience. Perhaps the most surprising thing about this view of reference memory is that it is not only compatible with almost all previous results in both humans and animals, but also completely consistent with almost all the theoretical treatment of reference memory provided by Meck and colleagues (Meck, 1983, 1991; Meck & Angell, 1992; Meck & Church, 1987a, 1987b; Meck et al., 1984), in their development of the K^* concept. The principal mechanism controlling variability of temporal references might be the K^* transformation itself, rather than any aspect of the storage process of reference memory.

Brunner et al. (1997) also attempted some reconceptualization of reference memory in SET, which applied specifically to situations in which animals received variable-interval schedules where more than one time (and often many times) could be associated with reinforcement. The standard treatment of such situations (Gibbon et al., 1988) would assume, like our SAM model, that each time of reinforcement was stored separately, and the individual instances aggregated together into distributions, samples from which were then used to control behaviour. An alternative proposed by Brunner et al. was a *minimax* model, where only the longest and shortest times of reinforcement were stored. Discussion of the operation of this model is beyond the scope of the present work, except for two remarks. First, the minimax model fitted some aspects of obtained data well, thus showing that "simplifications" of

reference memory, like our perturbation model, can have considerable power. Second, the minimax model was very different from our perturbation model in many ways; for example, unlike our model it had no special mechanism for keeping the memory stable in some situations and changing in others. Our perturbation model offers a particularly simple solution to the "stability/plasticity" dilemma: in the present case how to keep temporal reference constant in constant situations while allowing it to flexibly change if the situation changes.

As discussed above, transforming individual "important" times by K^*, storing them, and then sampling randomly from the list of stored times may be essentially indistinguishable from using the single transformed time stored on the previous trial only, in steady-state conditions. Only when times of reinforcement are changed is any inertia of the temporal reference memory observable, which might reveal more about its size and nature. Even in these cases, however, the pattern of behavioural changes observed (rapid adjustment to large changes in reinforcement time, slower adjustment to small changes) may be more compatible with a temporal representation like that provided by the perturbation model than by any mechanism assuming either that some weighted average is used, or that an extensive stock of previous examples of times of reinforcement is present.

It should be noted, however, that although we argue that reference memory might be very limited in extent, we do not suggest that there is no distinction to be made between "working memory" for duration (essentially the contents of the accumulator) and "reference memory". On the contrary, the two may be psychologically and possibly even physiologically completely distinct. Further experimentation is needed to decide how many properties reference and working memory have in common, but even if a number of common properties are found, distinctiveness of the two sorts of temporal memory might still be preserved. Among the interesting properties of working memory for duration is the "subjective shortening" reported in data from both animals and humans (see Wearden, Parry, & Stamp, 2002, for a review and some experimental data from humans). Whether temporal reference memory exhibits similar effects remains to be seen.

The present article represents efforts to fulfil part of the programme proposed by Wearden (1999) for deepening our understanding of the clock/memory/decision structure of the timing system in humans and animals as proposed by SET. Essentially, the proposed method is "isolation" of one part of the system whilst leaving everything else constant, in the present case an exploration of the "black box" of reference memory. Far from being some Pandora's box, which when opened reveals a complexity that produces theoretical confusion, our data and modelling suggest that temporal reference memory has a simpler structure than that previously conjectured. Further work with both humans and animals (with transitions from one time of reinforcement being particularly interesting in the latter type of subjects) may enable us to understand better than ever before the role played by temporal reference memory in SET-like models of timing.

REFERENCES

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

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

Brunner, D., Fairhurst, S., Stolovitsky, G., & Gibbon, J. (1997). Mnemonics for variability: Remembering food delay. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 68–83.

Church, R. M. (1984). Properties of the internal clock. *Annals of the New York Academy of Sciences*, *423*, 566–582.

Church, R. M. (1989). Theories of timing behavior. In S. B. Klein & R. R Mowrer (Eds.), *Contemporary learning theories: Instrumental conditioning theory and the impact of biological constraints on learning* (pp. 41–71). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

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.

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

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

Gibbon, J., & Church, R. M. (1984). Sources of variance in an information processing theory of timing. In H. L. Roitblat, T. B. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 465–488). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gibbon, J., Church, R. M., Fairhurst, S., & Kaceinik, A. (1988). Scalar expectancy theory and choice between delayed rewards. *Psychological Review*, *95*, 102–114.

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

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, *7*, 170–184.

Higa, J. J. (1996). Dynamics of time discrimination: II. The effects of multiple impulses. *Journal of the Experimental Analysis of Behavior*, *59*, 529–542.

Higa, J. J., Wynne, C. D. L., & Staddon, J. E. R. (1991). Dynamics of time discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 281–291.

Lejeune, H., Ferrara, A., Simons, F., & Wearden, J. H. (1997). Adjusting to changes in the time of reinforcement: Peak-interval transitions in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 211–231.

Lejeune, H., Ferrara, A., Soffié, M., Bronchart, M., & Wearden, J. H. (1998). Peak procedure performance in young adult and aged rats: Acquisition and adaptation to a changing temporal criterion. *Quarterly Journal of Experimental Psychology*, *51B*, 193–217.

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

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

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

Meck, W. H. (1991). Modality-specific circadian rhythmicities influence mechanisms of attention and memory for interval timing. *Learning and Motivation*, *22*, 153–179.

Meck, W. H., & Angell, K. E. (1992). Repeated administration of pyrithiamine leads to a proportional increase in the remembered duration of events. *Psychobiology*, *20*, 39–46.

Meck, W. H., & Church, R. M. (1987a). Cholinergic modulation of the content of temporal memory. *Behavioral Neuroscience*, *101*, 457-464.

Meck, W. H., & Church, R. M. (1987b). Nutrients that modify the speed of internal clock and memory storage processes. *Behavioural Neuroscience*, *101*, 465–475.

Meck, W. H., Church, R. M., & Olton, D. S. (1984). Hippocampus, time and memory. *Behavioral Neuroscience*, *98*, 3–22.

Meck, W. H., Church, R. M., & Wenk, G. L. (1986). Arginine vasopressin inoculates against age-related increases in sodium-dependent high affinity choline uptake and discrepancies in the content of temporal memory. *European Journal of Pharmacology*, *130*, 327–331.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552.

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.

Rescorla, R. A., & Wagner, A. K (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*: *Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 242–268.

Rodriguez-Girones, M., & Kacelnik, A. (1999). Behavioral adjustment to modifications in the temporal parameters of the environment. *Behavioural Processes*, *45*, 173–191.

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

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

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

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

Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 50–511.

Wearden, J. H., Parry, A., & Stamp, L. (2002). Is subjective shortening in human memory unique to time representations? *Quarterly Journal of Experimental Psychology*, *55B*, 1–25.

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

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

Original manuscript received 16 April 2002
Accepted revision received 16 October 2002