

# The Remembrance of Times Past: Interference in Temporal Reference Memory

Ruth S. Ogden  
The University of Manchester

J. H. Wearden  
Keele University

Luke A. Jones  
The University of Manchester

Six experiments examined human performance on a modified temporal generalization task when either 1 or 2 standard durations were encoded. In most conditions, participants were presented with a 1st standard duration (A), then judged whether a number of comparison stimuli had the same duration as A. They were then presented with a 2nd standard (B) and again judged whether other comparison stimuli had the same duration as B. Then, after a delay period of 0–45 s, further comparison stimuli were presented, and participants judged whether those stimuli had the same duration as A, without A being represented. A was either the same length as B or shorter or longer than it, so potential retroactive interference effects of B on A could be examined. After a short delay before retesting of A comparisons, the peak of the temporal generalization gradient shifted toward the shortest of the comparisons when  $A < B$  and the longest when  $A > B$ . The results suggest that certain combinations of delay and interference might render the memory of A unusable, so that a new standard is constructed on the basis of the remembered relationship between A and B, a kind of “false memory” for duration.

*Keywords:* reference memory, reference, timing, interference, temporal generalization

A central aspect of the leading contemporary theory of time perception, scalar expectancy theory (SET; Gibbon, Church, & Meck, 1984; for applications to human timing, see Allan, 1998, and Wearden, 2003), is the importance of memory processes. SET is a multiprocess theory with three distinct levels. The first level involves a pacemaker–accumulator internal clock, a mechanism proposed in earlier work by Creelman (1962) and Treisman (1963). This clock converts the physical duration of events into an internal representation in terms of clock “ticks” or “pulses,” but this internal clock is only the starting point for the process, which eventually leads to time judgments or timed behavior. For example, consider the apparently simple task of temporal generalization (Wearden, 1992), used in a variant form in the present article. Typically, the experimental session begins with a small number of presentations of a temporal standard (e.g., a tone 400 ms long). Following this, comparison stimuli of varying durations are presented (e.g., ranging from 100 to 700 ms in 100-ms steps), and the participant is asked to judge whether each comparison had the same duration as the standard, using a *yes* or *no* response. Feedback may or may not be given after each response.

Even if an internal clock provides representations of the durations of all the stimuli used in this sort of experiment, memory and

decision processes are proposed to also intervene critically to determine performance. For example, SET supposes that the standard duration of the task initially enters into a working memory, then passes to a (possibly more permanent) reference memory, which contains memories of “important” times (Church & Gibbon, 1982). When each comparison stimulus is delivered, it is considered to reside temporarily in working memory and is then compared by a decision process with a sample of the standard drawn from reference memory. If the comparison stimulus and the sample of the standard are judged to be close enough by the decision mechanism, the participant makes a *yes* response; otherwise, a *no* response is given. For a more detailed discussion of decision processes in temporal generalization and other tasks, see Church and Gibbon (1982), Wearden and Grindrod (2003), and Wearden (2004).

In spite of the fact that the operation of memory processes is critical to SET’s account of how time judgments are generated, the properties of both the working and the reference memories remain underresearched relative to the properties of the internal clock, and very few studies have attempted to discover just how the memory processes proposed operate in practice. Only a handful of studies of working memory for duration have been conducted with human participants (e.g., Droit-Volet, Wearden, & Delgado-Yonger, 2007; Lieving, Lane, Cherek, & Tcheremissine, 2006; Wearden & Ferrara, 1993; Wearden, Foran, & Goodson, 2007; Wearden, Parry, & Stamp, 2002), although a larger literature derived from experiments with pigeons and rats exists (see Miki & Santi, 2005, for a recent example). In these experiments, the focus of interest is on the properties of working memory for duration per se.

---

Ruth S. Ogden and Luke A. Jones, School of Psychological Sciences, The University of Manchester, Manchester, England; J. H. Wearden, School of Psychology, Keele University, Keele, England.

Correspondence concerning this article should be addressed to Luke A. Jones, School of Psychological Sciences, The University of Manchester, Manchester M13 9PL, England. E-mail: luke.jones@manchester.ac.uk

In contrast, a somewhat larger number of studies have examined the influence of working memory processes in general for performance on timing tasks. For example, a series of experiments by Fortin (1999; Fortin & Breton, 1995; Fortin & Rousseau, 1987, 1998; Fortin, Rousseau, Bourque, & Kirouac, 1993) investigated the effects of nontemporal tasks, such as visual search and mental rotation, on the reproduction and production of time intervals. The researchers consistently found that performing a task that imposed a load on short-term memory while reproducing a previously learned duration increased reproduction duration, and this effect was found to be greater at longer durations and with more complex tasks (Fortin & Breton, 1995). Reproductions and productions are thought to increase in duration because the switch or accumulation process is controlled by short-term memory, so if accumulation is interrupted by a concurrent short-term memory task, then fewer pulses will be accumulated per unit time. This will result in a longer amount of real time passing before the required number of pulses are accumulated, therefore lengthening the time produced (Fortin et al., 1993). Notice that in these experiments, the focus of interest was not on memory for duration per se but was on the way short-term memory is involved in the timing system more generally.

Studies of the properties of reference memory for time have been less common, with articles by Jones and Wearden (2003, 2004); Penney, Gibbon, and Meck (2000); Rattat and Droit-Volet (2005); and Grondin (2005) constituting virtually the entire literature. Within these few studies, the focus of interest has been somewhat diverse (e.g., differences in duration judgments of auditory and visual stimuli; Penney et al., 2000; developmental effects; Rattat & Droit-Volet, 2005), so the answers to many fundamental questions about properties of reference memory for duration remain uncertain.

In two articles, Jones and Wearden (2003, 2004) explored some properties of the reference memories people apparently use when they perform timing tasks. In both articles, variants of the temporal generalization method described earlier were used. Jones and Wearden (2003) investigated whether temporal generalization performance was affected by the number of identical standards presented. People received one, three, or five presentations of a standard duration, then had to decide whether subsequent comparison durations were or were not the same as the standard. In fact, the number of standards presented never affected performance significantly. Jones and Wearden (2003) used computer modeling to show that such an effect was consistent with a temporal memory storage system in which the individual presentations of the standard were stored separately, rather than being averaged. In fact, such a proposal was implicit in the way that temporal reference memory was conceived in the original exposition of SET by Gibbon et al. (1984), even though the lack of effect of repeated presentations of a standard may seem counterintuitive and had not previously been emphasized as a consequence of the theory. Jones and Wearden (2003) also developed a model of how reference memory changes when new instances of a standard are presented, but the details are too complicated for discussion here, and the interested reader is referred to the original article.

In their second article, Jones and Wearden (2004) explored the central issue of whether remembering more than one time interval produces changes in performance compared with remembering only one. Both Jones and Wearden (2004) and Grondin (2005)

found evidence for relative performance decrements when two times were remembered, even in situations in which only one stimulus whose duration had to be judged was presented on each trial, which thus renders the results unlikely to have been caused by attentional division between multiple temporal targets (as might be the case in earlier work, e.g., Brown & West, 1990). This suggests that memories for duration may interfere with one another, much like memories for other sorts of material, and this potential interference is the subject of the present work. Jones and Wearden (2004) also showed that for any interference effect to be observed, both the standards presented needed to be used to make judgments about comparison stimuli. If two standards were presented but only one was used to judge comparisons, then no interfering effect could be observed. However, it should be noted that in both Jones and Wearden (2004) and Grondin (2005), interference effects were small in magnitude.

Why should interference paradigms for temporal memories be of any interest? Interference paradigms are commonly used to investigate memory contents, when these are potentially of different types (e.g., verbal and spatial working memory; Hale, Myerson, Rhee, Weiss, & Abrams, 1996). The rationale for the paradigm is that material that is "similar" to other material, or encoded a similar manner, interferes with its memory, whereas dissimilar material interferes to a lesser extent. Thus, the development of reliable interference paradigms for temporal memory might enable us to understand more about how reference memory for duration is encoded and decoded. Such information might then shed light on what are currently some very difficult issues in the study of time perception, such as whether the durations of stimuli presented in different modalities (e.g., auditory or visual) are stored as a "common code"—for example, clock ticks—independently of the original stimulus modality, or in some other way. For example, if the durations of auditory and visual stimuli are translated into a common code, then interference effects between the representations of the durations of stimuli presented in the auditory or visual modalities ought to be strong, whereas if stimuli in different modalities are encoded by separate systems, then interference ought to be weaker or even absent. The development of methods that reliably produced interference when stimuli were in the same modality would enable these possibilities to be explored more precisely than is possible at present.

In general, the present article, like some other recent work (Grondin, 2005; Jones & Wearden, 2003, 2004), seeks to explore properties of reference memory for duration and thus develop our understanding of how such memories operate. This increase in knowledge may enable some of the less well-developed areas of formal theories such as SET to be specified more precisely, thus allowing the derivation of detailed predictions. In the present article, SET is used in two ways. First, it provides a general background, via its emphasis on the way that reference and working memories operate in timing tasks. Second, as we show later, a computer model consistent with SET can be developed that can deal with many aspects of the effects obtained in the experiments reported below.

To begin, we developed a new technique that might be useful in trying to produce interference between standards in reference memory. In most of the experiments presented below, people learned successive temporal standards in the form of tones (designated A and B, which usually differed in duration) and made

judgments of comparison stimuli using these standards in a temporal generalization variant. They were then required to make judgments using the A standard as the basis for the comparisons, without this standard being presented again. The obvious focus of interest is how the use of the B standard might affect deferred performance with the A standard: a sort of potential retroactive interference effect.

### Experiment 1

The basic design for Experiment 1 was as follows. The experimental session was arranged in the form of repeated blocks. Each block began with three presentations of a standard duration, A, which was chosen randomly from a uniform distribution running from 400 to 600 ms, so each block began with a different standard. Following the three standard presentations, participants received comparison stimuli, the duration of which we calculated by multiplying the A duration by values from 0.25 to 1.75 in increments of 0.25 and presenting them in a random order. We call these stimuli *A comparisons*. After presentation of each comparison, the participants were required to indicate whether they judged it to have the same duration as the standard (by making *yes* or *no* responses), but no performance-related feedback was given. Then another standard, B, was presented, and in Experiment 1 the duration of B was either the same duration as A or the duration of A plus either 100 ms or 400 ms, in different experimental groups. After three presentations of B, stimuli that were multiples of B were presented and had to be compared with B (*B comparisons*). After this had been done, the A comparisons were again presented, but the standard A was not presented again, and participants were required to judge whether each comparison had the same duration as A. This procedure constituted one block. Then another A value was randomly generated for the next block (and this led, in turn, to another value of B), and so on for 18 blocks.

The aim of the procedure is obvious. The main focus of interest was performance with the comparison stimuli just after A had been presented (called an *immediate judgment*) and later, after B had been presented (called a *deferred judgment*). Any difference in performance between immediate and deferred judgments could be due to (at least) two causes. One is the delay interposed between the immediate (immediately after A) and deferred (immediately after B) A comparisons. The other is the potentially interfering effect of B and the B comparisons. As demonstrated by Jones and Wearden (2004), merely presenting a standard (e.g., B) after another one (A) would not necessarily produce any evidence of interference, unless B was used as the basis for judgments of other stimuli, but our use of B comparisons in the above design ensured that this was the case, so some interfering effect of B on later judgment of A comparisons should be expected.

### Method

**Participants.** Forty-five University of Manchester undergraduates participated for course credit and were arbitrarily allocated to three groups (*close*, *far*, and *same*).

**Apparatus.** An IBM PC-compatible computer recorded all experimental events. The stimuli whose durations were judged were always 500-Hz tones produced by the computer speaker. The keyboard monitored the participant's responses. The program that

ran the experiment and recorded all the data was written in Micro Experimental Laboratory language (Schneider, 1988).

**Procedure.** A variant of the temporal generalization method was used. The participants received all trials in one session lasting approximately 45 min. Instructions were presented to the participants on the computer screen and on paper prior to the start of the test.

The procedure for the close group was as follows. At the start of each block, participants were presented with the Standard Duration A three times. This standard was a tone with a duration randomly sampled from a uniform distribution running from 400 to 600 ms, and the same duration was used for all three presentations within a block. Between presentations there was a delay drawn from a uniform distribution ranging from 1,500 to 2,000 ms. Participants then received seven comparison stimuli that were Standard A multiplied by 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, or 1.75, with stimuli presented in a random order (A comparisons). After each comparison tone, participants had to indicate whether the comparison had the same duration as Standard A by pressing the Y key for yes or the N key for no. No feedback was given. We refer to this as *immediate testing*.

Then participants were presented with Standard B three times: The duration of B was 100 ms longer than A, whatever A had been on that block. They then received seven comparison stimuli that had durations defined by the B duration multiplied by 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, or 1.75, and comparison stimuli were presented in a random order (B comparisons). After each comparison, participants had to indicate whether it had the same duration as B by pressing Y for yes or N for no. Finally, the A comparisons were presented again, with instructions to compare them to the original A standard, but this was not presented. We refer to this second presentation of the A comparisons as *deferred testing*. The above steps were repeated 18 times, with a different A duration being sampled for each block. No feedback as to performance accuracy was ever given after responses to the comparison durations.

The procedure for the far and same experimental groups was the same except for the relation between the durations of B and A. For the far group B was 400 ms longer than A, and for the same group B and A were actually identical, although the participants were not told this.

### Results

Figure 1 shows temporal generalization gradients in the form of the mean proportion of *yes* responses (judgments that the comparison duration was the standard) plotted against the comparison/standard ratio. Data are shown from the immediate and deferred tests with the A comparisons. The upper, middle, and bottom panels show the data from the close, far, and same groups, respectively.

Inspection of the data in all panels of Figure 1 shows that the maximum number of *yes* responses occurred when the comparison/standard ratio was 1.0 in all cases, suggesting that the standard was stored veridically in reference memory. Gradients appeared slightly flatter in deferred than in immediate testing of the A comparisons, possibly reflecting poorer task performance in the former case.

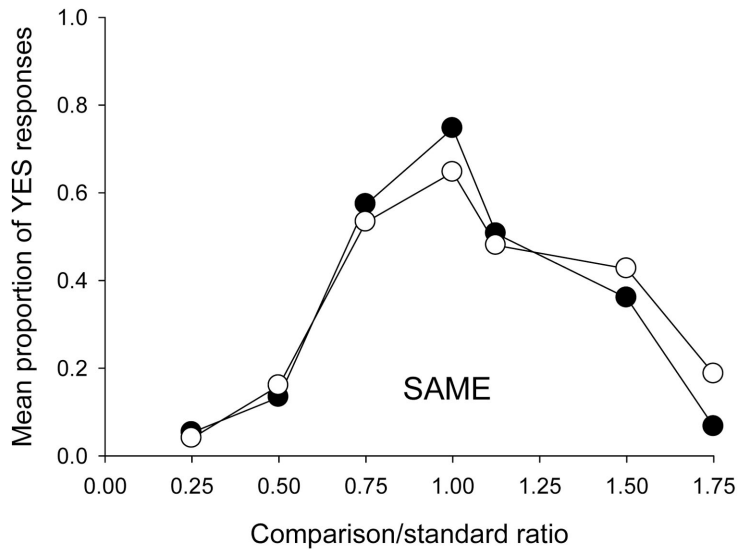
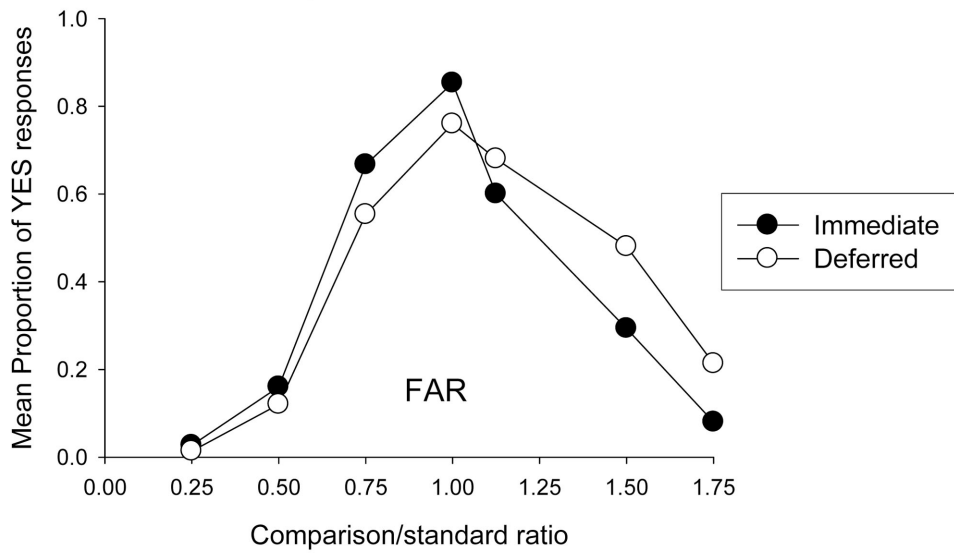
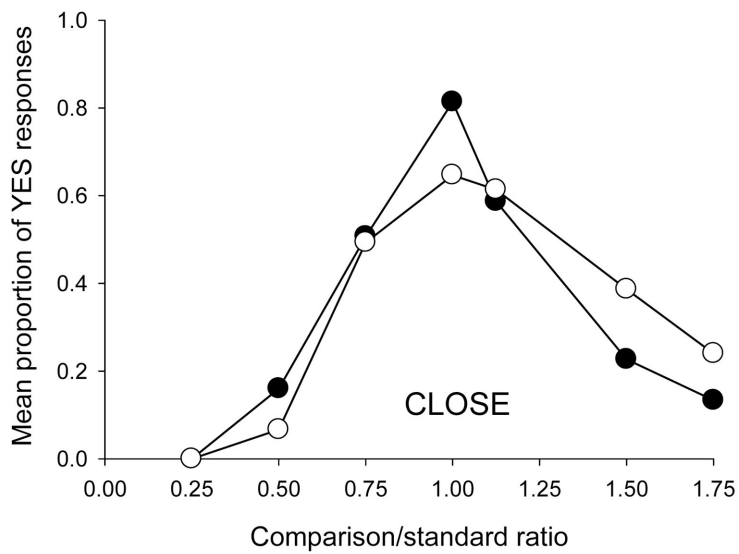


Figure 1. Temporal generalization gradients (proportion of *yes* responses [identifications of a comparison duration as the standard] plotted against comparison/standard ratio) from Experiment 1. Within each panel, data from the immediate and deferred comparisons are presented. Upper panel: *close* comparisons; center panel: *far* comparisons; lower panel: *same* comparisons.

An analysis of variance (ANOVA) was carried out, with testing condition (immediate or deferred) and comparison/standard ratio (using the seven comparison values) as within-subject factors and group (close, far, same) as the between-subjects factor. This showed a significant main effect of comparison/standard ratio,  $F(6, 252) = 107.25, p < .001$ ; and a significant Comparison/Standard Ratio  $\times$  Delay interaction,  $F(6, 252) = 7.78, p < .001$ ; but no significant effect of delay,  $F(1, 44) = 0.84, p = .365$ . There was no significant effect of group,  $F(2, 42) = 1.21, p = .309$ , nor were there any significant interactions involving group as a factor. To look at the experimental effects in more detail, we conducted repeated-measures ANOVAs for each of the three conditions separately.

For the close group, there was no significant effect of testing condition (immediate vs. deferred),  $F(1, 13) = 0.04, p = .852$ ; but there was a significant effect of comparison/standard ratio,  $F(6, 78) = 38.27, p < .001$ ; and also a significant Comparison/Standard Ratio  $\times$  Testing Condition interaction,  $F(6, 78) = 4.50, p < .001$ , with the latter effect indicating that the temporal generalization gradients in the conditions compared had slightly different shapes. Inspection suggests that the gradient from the deferred condition was slightly displaced to the right relative to that from the immediate condition and perhaps also slightly flatter.

For the far group, there was no significant effect of testing condition,  $F(1, 14) = 1.18, p = .295$ , but there was a significant effect of comparison/standard ratio,  $F(6, 84) = 49.43, p < .001$ . There was also a significant Testing Condition  $\times$  Comparison/Standard Ratio interaction,  $F(6, 84) = 3.73, p < .01$ , once again showing flattening and slight rightward displacement of the temporal generalization gradient between immediate and deferred testing.

For the same group, there was no significant effect of testing condition,  $F(1, 14) = 0.06, p = .807$ , but there was a significant effect of comparison/standard ratio,  $F(6, 84) = 24.13, p < .001$ . There was no significant interaction between testing condition and comparison/standard ratio,  $F(6, 84) = 1.36, p = .241$ , so in the same group the generalization gradient observed in the immediate and deferred conditions appeared more statistically similar than in the far and close groups.

### Discussion

If we consider our three groups, we see that the procedure for the close and far groups involved the B standard being systematically longer than A (by 100 and 400 ms for the two groups, respectively), so some potential interference might be expected, whereas in the same group A and B were actually the same, so the two standards might have reinforced one another rather than producing interference. Although the close and far groups differed in the potential interfering relation between B and A, they were the same with respect to the performance differences between the immediate and deferred testing conditions. The analysis suggests that the putative interfering conditions (close and far) resulted in a significant flattening and displacement of the temporal generalization gradient (indicated by the significant Testing Condition  $\times$  Comparison/Standard Ratio interaction), whereas this effect was not significant in the same group. This suggests that in the deferred condition the participants' memory of the Standard A was poorer than when this memory was tested immediately. This might have

been due to the imposition of a delay since A per se; the fact that another standard, B, had been presented and tested since presentation of the A and its comparisons; or both. The exact causes of performance changes are explored in more detail in later experiments, although the absence of interaction in the same group suggests that "pure" delay was not the responsible factor in the absence of any difference in duration between A and B.

The results of Experiment 1 therefore provide some prima facie evidence of interference by B on the memory of A, although in all cases performance in the deferred testing condition was reasonably good, with temporal generalization gradients always peaking at the comparison/standard ratio of 1.0 (i.e., in the case where the comparison and the standard were the same). Overall, therefore, although an interference effect appears to have been demonstrated, it seems rather modest, and this result is in line with results from other experiments, such as those of Jones and Wearden (2004) and Grondin (2005). In both these studies, more than one standard was encoded and tested, and statistically significant, albeit rather small, effects on performance of encoding more than one duration were obtained.

Why was the performance difference between immediate and deferred testing so small? One possibility is that the task employed was "too easy" in two ways. One of these is that some comparison/standard ratios (e.g., 0.25) produced comparison stimuli so remote from the standard that they could be correctly classified as being different from it even if the memory of the standard had been degraded to a considerable extent. The other possibility is that the delays used, even in the deferred testing condition, were too short. Experiment 2 changed the procedure of Experiment 1 to explore these possibilities.

### Experiment 2

The procedure of Experiment 2 was a modified version of that used for Experiment 1, with two principal differences. The first was that the comparisons were generated by a smaller range of standard multiples, from 0.625 to 1.375, rather than the 0.25 to 1.75 used in Experiment 1. This produced a more difficult comparison set (see Ferrara, Lejeune, & Wearden, 1997, and Wearden & Grindrod, 2003, for discussion of the effects of this manipulation). A further change was that on half the blocks, the testing with the A comparisons was delayed by 45 s. That is, after the B comparisons had been presented, there was a delay of 45 s, and then the A comparisons were presented again. We therefore distinguish between three testing conditions for the A comparisons: immediate (just after the A standard had been presented, as in Experiment 1), deferred (just after the B comparisons had been presented, as in Experiment 1), and delayed (45 s after the B comparisons had been presented, a new condition of Experiment 2). As in Experiment 1, there were close, far, and same groups, distinguished by the relation between B and A.

### Method

*Participants.* Forty-five University of Manchester undergraduates participated for course credit and were arbitrarily allocated to one of three groups (close, far, and same).

*Procedure and apparatus.* The general procedure was the same as in Experiment 1 except for two differences. One was that

the comparison durations were generated by the appropriate standard (A or B) multiplied by 0.625, 0.750, 0.875, 1.000, 1.125, 1.250, or 1.375, with the comparison stimuli presented in a random order within each block. A further change was that on half the blocks, the representation of the A comparisons was delayed by 45 s after the last response to the B comparisons. We call this *delayed testing*. On the other half of the blocks, the A comparisons were represented immediately after the B comparisons, as in Experiment 1, and, as before, we refer to this as *deferred testing*. *Immediate testing* refers to presentation of A comparisons immediately after the three presentations of the A standard, as in Experiment 1. In keeping with Experiment 1, the three experimental groups were defined by the relation between B and A, and this relation was the same as in Experiment 1 (close: B 100 ms longer; far: B 400 ms longer; same: B the same duration as A). All other procedural details, including apparatus, were the same as in Experiment 1.

## Results

Figure 2 shows temporal generalization gradients, the mean proportion of *yes* responses plotted against the comparison/standard ratio. The upper, middle, and bottom panels show the data from the close, far, and same groups, respectively. Within each panel, data are shown from immediate, deferred, and delayed testing.

Inspection of the data in Figure 2 suggests that the maximum number of *yes* responses occurred when the comparison/standard ratio was 1.0 in some, but not all, generalization gradients. In the close and far groups, deferred and delayed testing appeared to shift temporal generalization gradients to the left, with a particularly marked effect of delayed testing in the far group. In this condition, the proportion of *yes* responses at the smallest comparison/standard ratio (0.625) was 82%, indicating a very marked change in performance compared with other conditions. There was no apparent shift in gradients for either the deferred or the delayed conditions of the same group.

A repeated-measures ANOVA was carried out, with testing condition (immediate, deferred, and delayed) and comparison/standard ratio as within-subject factors and group (close, far, same) as a between-subjects factor. There were significant main effects of testing condition,  $F(2, 82) = 5.32, p < .01$ ; and comparison/standard ratio,  $F(6, 246) = 23.71, p < .001$ ; and a significant Comparison/Standard Ratio  $\times$  Testing Condition interaction,  $F(12, 492) = 10.04, p < .001$ . There was no significant effect of group,  $F(2, 41) = 0.08, p = .927$ . However, there was a significant Group  $\times$  Testing Condition interaction,  $F(4, 82) = 7.23, p < .001$ , and a significant Group  $\times$  Comparison/Standard Ratio interaction,  $F(12, 246) = 3.75, p < .001$ . To look at the experimental effects in more detail, we conducted separate repeated-measures ANOVAs for each of the three groups.

Repeated-measures ANOVAs on data from the different groups separately used testing condition (immediate, deferred, and delayed) and comparison/standard ratio as within-subject variables. For the close group there was no significant effect of testing condition,  $F(2, 28) = 0.33, p = .722$ ; but there were significant effects of comparison/standard ratio,  $F(6, 84) = 11.72, p < .001$ ; and a significant interaction between comparison/standard ratio and testing condition,  $F(12, 168) = 2.94, p < .01$ . For the far

group there were significant effects of testing condition,  $F(2, 26) = 22.19, p < .001$ ; and comparison/standard ratio,  $F(6, 78) = 13.71, p < .001$ ; and also a significant Comparison/Standard Ratio  $\times$  Testing Condition interaction,  $F(12, 156) = 16.66, p < .001$ . For the same group there was no significant effect of testing condition,  $F(2, 28) = 0.84, p = .442$ . There was a significant effect of comparison/standard ratio,  $F(6, 84) = 8.25, p < .001$ , and a significant interaction between comparison/standard ratio and testing condition,  $F(12, 168) = 2.05, p < .05$ .

## Discussion

The results from Experiment 2 clearly demonstrate a systematic interference effect that was dependent on the relationship between the durations of the two encoded standards. When the two standards were of equal length, there was no systematic interference from the second standard, B, although the temporal generalization gradient was flattened in the delayed and deferred conditions compared with the immediate one. However, when B was longer than A (as in the close and far conditions), the interference of B appeared to make the memory of A shorter than its real duration. Furthermore, it appeared that the larger the difference between A and B was, the larger was the reduction in the remembered duration of A, as the shift in the temporal generalization gradients was larger in the far than in the close group. Interposing a delay between the end of testing of the B comparisons and the subsequent retest of the A comparisons appeared to be the critical factor in producing a marked shift in the generalization gradients toward the left, although this effect most clearly occurred in the far condition, in which B was markedly longer than A.

The above experiments, in which potentially interfering stimuli were used (Experiments 1 and 2), have demonstrated various degrees of interference with memory of a standard duration (A) in temporal generalization by another standard that was longer than it (B). The largest and most striking effect obtained was on performance of the far group with delayed testing in Experiment 2, in which 82% of responses that a stimulus duration was equal to a standard were, in fact, given to a stimulus that only had 62.5% of the standard duration, apparently a massive distortion of temporal memory.

Why did the memory of A appear to be distorted so much in the far group with delayed testing? One suggestion is that the apparent distortion of performance was actually a manifestation of an even worse situation, one in which the memory of the Standard A had effectively been lost, in the sense that the participant made no attempt at all to use it. For example, suppose that, in a delayed testing condition, the participant decided that his or her memory of A was so degraded as to be useless. How could performance when the A comparisons were presented then be carried out? The participant might have had little confidence in his or her memory of A but might have still retained a useful fact about it—namely, that A was a lot shorter than B (far group), a bit shorter than B (close group), or about equal to B (same group)—and the memory of B, even after a delay, might have been usable. This hypothesis suggests that the participant abandoned any attempt to use the putatively degraded memory of A but instead “constructed” a standard for use when the A comparisons were represented. However, the participant could not construct this new standard very effectively, and in the far condition of Experiment 2 participants constructed

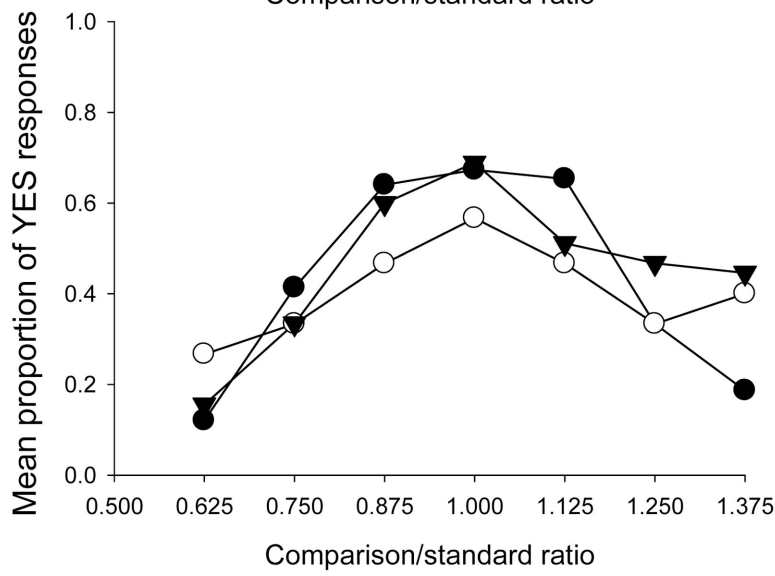
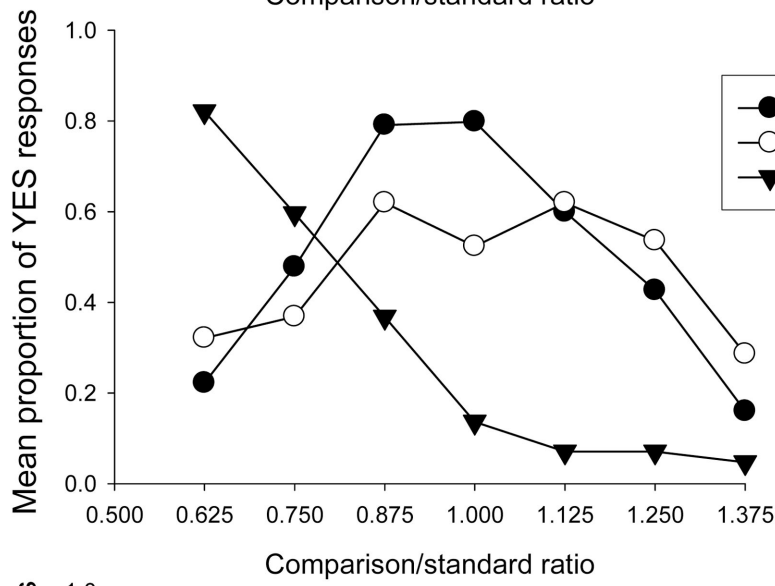
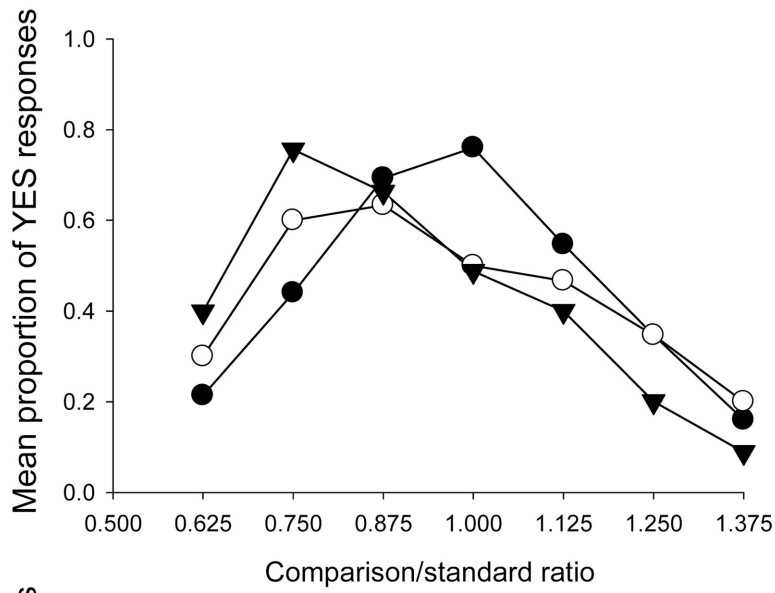


Figure 2. Temporal generalization gradients from Experiment 2. Each panel shows data from the immediate, deferred, and delayed comparisons. Upper panel: *close* comparisons; center panel: *far* comparisons; lower panel: *same* comparisons.

one that was too short (i.e., A really was shorter than B, but not as short as the much shorter constructed standard). In the close condition, the same construction process could have occurred, but the distortion was not as severe (although still marked). In the case of the same group, there was deterioration in performance in the deferred and delayed testing compared with the immediate testing condition but no systematic shift in the temporal generalization gradient, as A and B were the same.

Our proposed effects are reminiscent of those studied under the rubric of adaptation level (AL) theory (Helson, 1964), in which stimulus judgments are shown to be systematically changed by the presence of "anchor" stimuli with which the stimuli judged are implicitly or explicitly compared. In our Experiment 2, for example, the close, far, and same conditions could all be supposed to change the AL with respect to which the delayed A comparisons were judged. This idea is explored further in the General Discussion.

### Experiment 3

How can we test the conjecture that people constructed a standard for the delayed A comparisons? The obvious way is to vary the relation between the B and A standards so that B is sometimes shorter than A rather than always longer, as in Experiments 1 and 2. In Experiment 3, three groups differed in the relation between B and A. For one group (the *longer* group) B was 400 ms longer than A (essentially a near replication of the far group of Experiment 2), for another (the *shorter* group) B was 400 ms shorter than A, and for the third group (the *same* group) B and A were equal in duration.

The critical predictions concerned performance in the delayed testing conditions for the three groups. For the longer group, B was longer than A, so any constructed standard used could be shorter than the real value of A, and a temporal generalization gradient with a very marked leftward shift was expected. In contrast, for the shorter group, the constructed standard might deviate from the real A in the other direction, and the generalization gradient might be expected to be markedly shifted to the right. In the same group, conversely, a constructed standard would produce results similar to those expected with A itself (as  $B = A$ ). In the deferred testing condition, we proposed that people would try to use their memory of A rather than a constructed standard, so some interference effects might be evident, but these were expected to be much less marked than in the delayed testing condition.

### Method

*Participants and apparatus.* Twenty-eight University of Manchester 1st-year psychology undergraduates participated for course credit and were arbitrarily allocated to three groups (9 to the longer group, 8 to the shorter group, and 11 to the same group). The apparatus was the same as in Experiments 1 and 2.

*Procedure.* The procedure was essentially identical to that of Experiment 2, but with the inclusion of an extra condition in which B was shorter than A (by 400 ms). As we were including a shorter condition, it was necessary to alter the length of the standards used from 400–600 ms (as in Experiments 1 and 2) to 800–1,200 ms. A delay drawn from a uniform distribution ranging from 1,500 to 2,000 ms was imposed between presentations of the standard. On

half of all blocks the A comparisons were presented immediately after the final B comparison (deferred conditions), and in the other half there was a 30-s delay prior to presentation (delayed conditions). A total of 18 blocks were presented.

### Results

Figure 3 shows data from the immediate, deferred, and delayed testing conditions from the shorter (upper panel), longer (center panel), and same (lower panel) groups, and Figure 4 shows data from the theoretically critical delayed testing condition from the three groups displayed on the same graph. Inspection of Figure 3 shows apparent effects of testing condition for all groups, and inspection of Figure 4 shows that the temporal generalization gradients for the three groups in the delayed condition differed very radically.

An overall ANOVA conducted on all data from the immediate, deferred, and delayed testing conditions found no effect of group (shorter, longer, or same),  $F(2, 25) = 0.37, p = .696$ , and no effect of testing condition (immediate, deferred, or delayed),  $F(2, 50) = 2.76, p = .073$ . There was a significant effect of comparison/standard ratio,  $F(6, 150) = 21.37, p < .001$ . Two-way interactions between Testing Condition and Comparison/Standard Ratio,  $F(12, 300) = 5.87, p < .001$ , and between Comparison/Standard Ratio and Group,  $F(12, 150) = 1.92, p < .05$ , were both significant, but the interaction between Testing Condition and Group was not,  $F(4, 50) = 0.77, p = .548$ . The three-way interaction between Comparison/Standard Ratio, Testing Condition, and Group was also significant,  $F(24, 300) = 6.05, p < .001$ .

We then conducted ANOVAs for each group separately, following the arrangement shown in the different panels of Figure 3. For the shorter group (upper panel) there was no significant effect of testing condition,  $F(2, 16) = 0.21, p = .814$ ; but there was a significant effect of comparison/standard ratio,  $F(6, 48) = 6.23, p < .001$ ; and a significant Comparison/Standard Ratio  $\times$  Testing Condition interaction,  $F(12, 96) = 3.83, p < .001$ . For the longer group (center panel) again there was no significant effect of testing condition,  $F(2, 14) = 3.12, p = .076$ ; but there was a significant effect of comparison/standard ratio,  $F(6, 42) = 9.74, p < .001$ ; and the Comparison/Standard Ratio  $\times$  Testing Condition interaction was also significant,  $F(12, 84) = 11.44, p < .001$ . For the same group there was no significant effect of testing condition,  $F(2, 20) = 0.81, p = .461$ ; but there was a significant effect of comparison/standard ratio,  $F(6, 60) = 11.97, p < .001$ ; and the interaction between Comparison/Standard Ratio and Testing Condition was also significant,  $F(12, 120) = 2.04, p < .05$ . If we consider the data in the form shown in Figure 4, an ANOVA of data from the three groups in the delayed condition found a significant overall effect of comparison/standard duration,  $F(6, 150) = 2.52, p < .05$ ; and no effect of group,  $F(2, 25) = 0.68, p = .517$ ; but a significant Comparison/Standard Ratio  $\times$  Group interaction,  $F(12, 150) = 8.45, p < .001$ .

### Discussion

The principal result of Experiment 3, shown in Figure 4, is represented in the temporal generalization gradients from the three groups in the delayed testing condition. Obviously, the relation between B and A had an overwhelming influence on performance



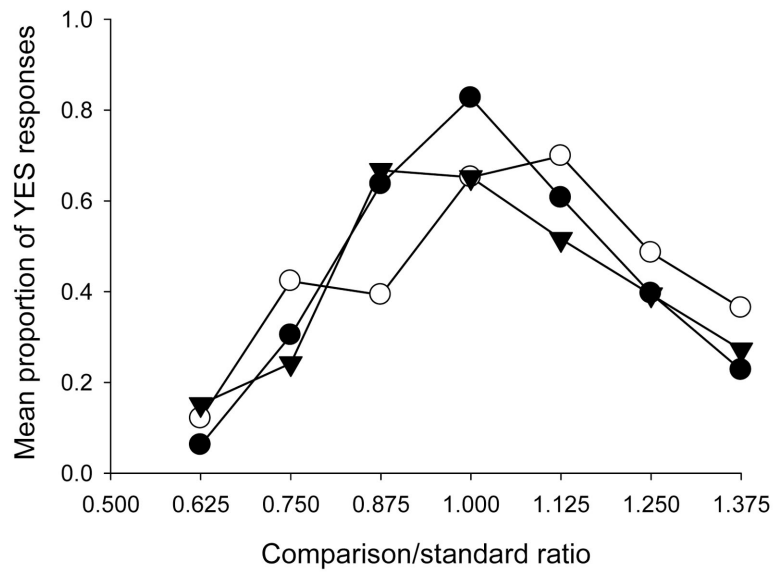
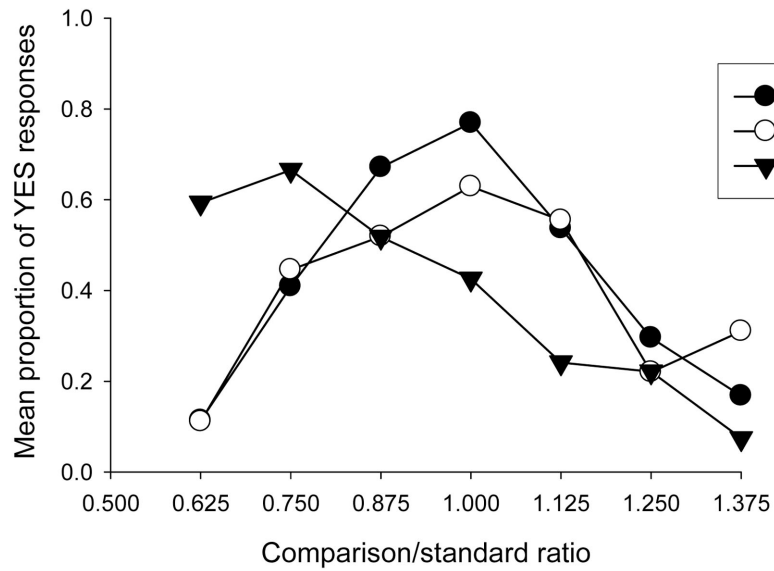
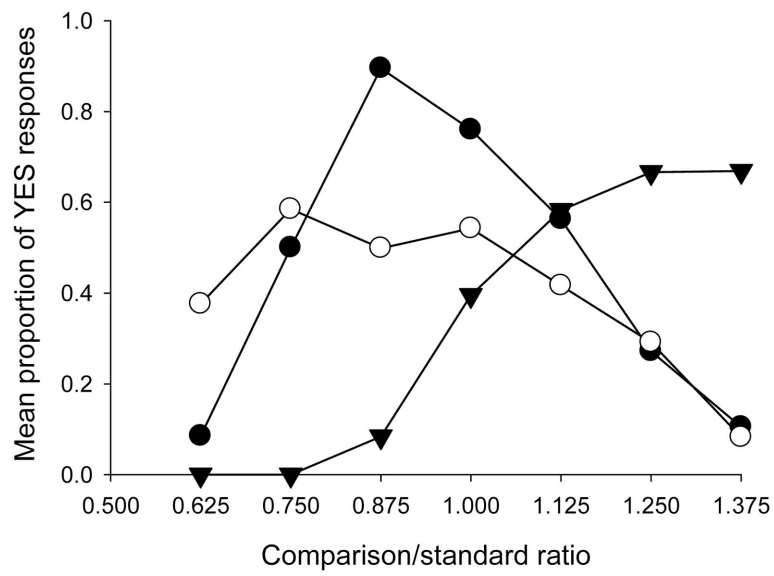


Figure 3. Temporal generalization gradients from Experiment 3. Each panel shows data from the immediate, deferred, and delayed comparisons. Upper panel: *shorter* group; center panel: *longer* group; lower panel: *same* group.

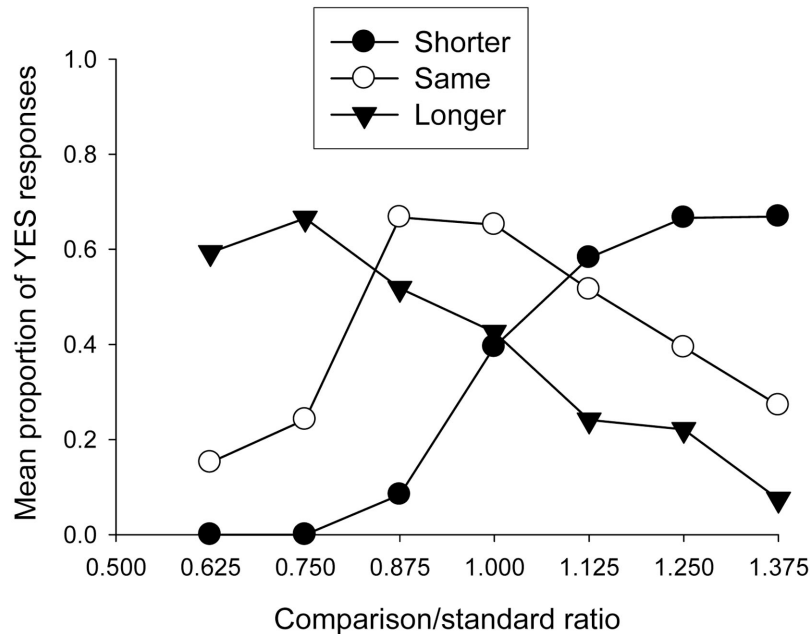


Figure 4. Temporal generalization gradients from critical comparisons in Experiment 3. Temporal generalization gradients are shown from the delayed comparisons, from the *shorter*, *longer*, and *same* groups.

in this condition, and the effect was exactly as predicted. In the longer group ( $B > A$ ) the generalization gradient was shifted very markedly to the left (replicating the effect found in Experiment 2), in the shorter group ( $B < A$ ) the gradient was shifted markedly to the right, and in the same group ( $B = A$ ) the gradient was peaked at comparison/standard ratios close to 1.0. Note that the results from the longer and shorter groups are the opposite of what would be expected if B was assimilated in some way with A or if the A and B standards were mixed together to produce some intermediate value: In the longer group this would have made the "standard" longer, and in the shorter group it would have made it shorter, resulting in generalization gradients shifted in the opposite direction to those actually observed. In general, the results strongly support the suggestion that the delayed testing condition had effectively rendered the memory of A unusable and that performance with the A comparisons in these conditions was based on a constructed standard founded on the remembered relation between B and A.

#### Experiment 4

The results of Experiments 2 and 3, taken together, suggest that imposing a delay between testing of B comparisons and retesting of the A comparisons resulted in participants constructing a new standard based on the relations between B and A and suggest that this construction process was inaccurate. In Experiments 2 and 3, the delay after the B comparisons had been presented was 45 s, which raises the obvious question of what the data would look like if a shorter delay had been imposed. Experiment 4 addresses this question by systematic changes in delay. The experiment consisted of two independent subexperiments, which are reported together as they were procedurally almost identical. In both Experiments, B standards were used that were 400 ms longer than A standards, so

the conditions resembled those in the far group of Experiment 2 and the longer group of Experiment 3, apart from differences in the delay. In Experiment 4A, delays after the presentation of the B comparisons were 0 (deferred), 5, 10, 20, and 30 s. Experiment 4B, which was conducted second, after the results of Experiment 4A were known, employed delays of 0 (deferred), 2, 3, 4, and 5 s.

As delays are increased, what outcome is expected? There are a number of distinct possibilities: One is that systematically increasing the delay from zero (the deferred condition) produces some gradual shift in the participants' behavior, from attempting to use their putatively degraded memory of A when the delay is short to constructing a new standard when the delay is long. Therefore, we might expect a shift in generalization gradient shape with increasing delay, from a shape peaked approximately at the standard (i.e., comparison/standard ratio = 1.0) at short delays to one markedly shifted to the left (because  $B > A$ ) at longer delays, with some generalization gradients from intermediate delays showing mixtures of the two sorts of shapes. Another possibility is that the change in gradient shape is abrupt, with gradients peaked approximately at a comparison/standard ratio of 1.0 up to some delay, then being markedly skewed at longer delays. Another possibility is that the imposition of all delays, even very short ones, markedly skews the shape of the generalization gradient.

#### Method

*Participants and apparatus.* Nineteen University of Manchester 1st-year psychology undergraduates (Experiment 4A) and seventeen undergraduates (Experiment 4B) participated for course credit. The apparatus was the same as in Experiments 1, 2, and 3.

*Procedure.* The procedure in Experiment 4A was essentially identical to that of Experiment 3 but used only the longer (i.e.,  $B > A$ ) condition. The within-subject variable was the length of delay

between the end of the presentation of the B comparisons and the start of the retesting of the A comparisons. The delay conditions were deferred (no delay), 5-s, 10-s, 20-s, and 30-s delays. The procedure in Experiment 4B was essentially identical to that of Experiment 3, but the delay conditions were now deferred (no delay), 2-s, 3-s, 4-s, and 5-s delays. On half of all blocks, A was the same length as B; on the other half, B was 400 ms longer than A. There were 30 blocks in total.

*Results*

*Experiment 4A.* The upper panel of Figure 5 shows data from the immediate, deferred, and delayed (5, 10, 20, and 30 s) testing conditions. Inspection of the data suggests that after as little as a 5-s delay prior to retesting of the A comparisons, the peak of the temporal generalization gradient shifted markedly to the left, from peaking at the standard (comparison/standard ratio = 1.0) to

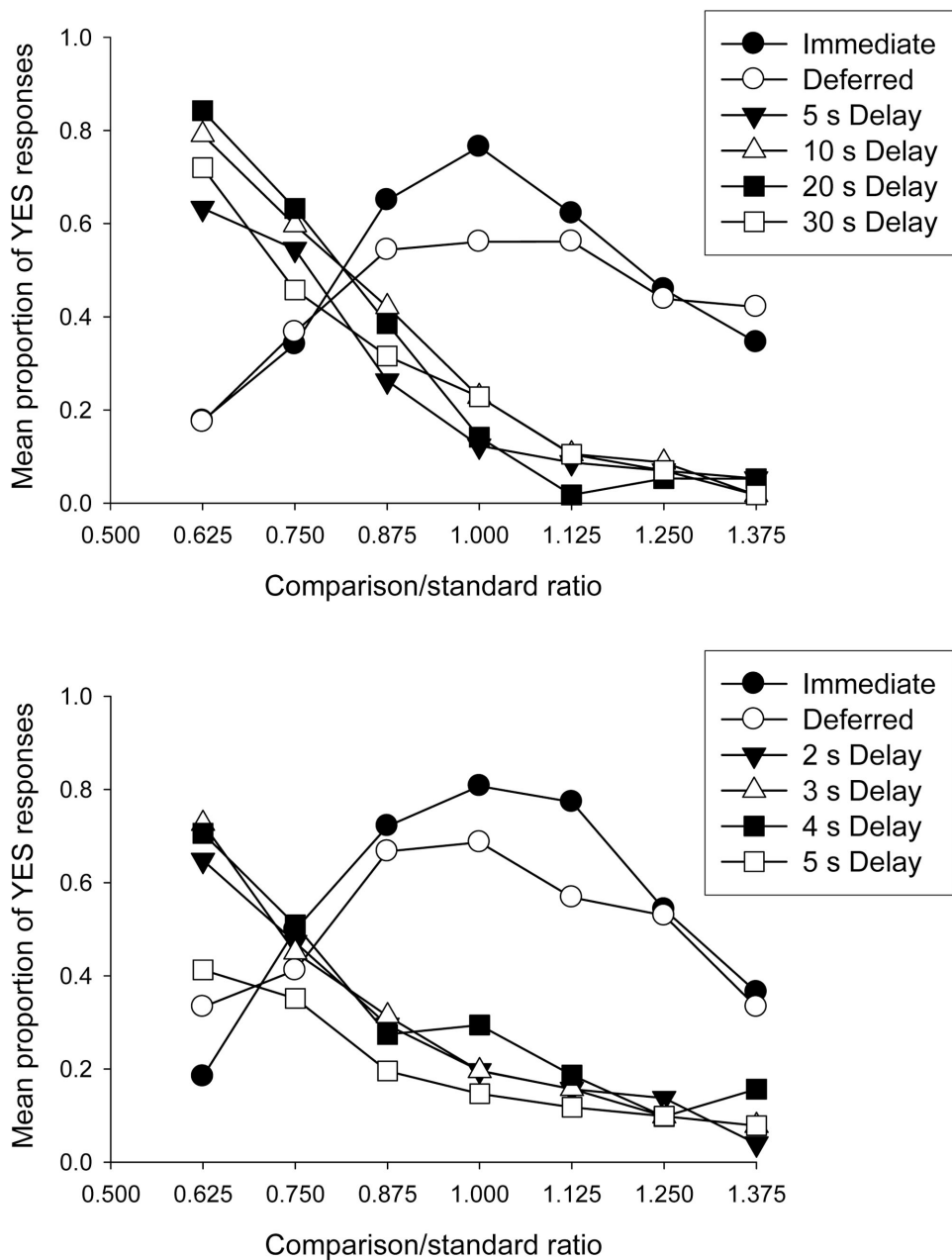


Figure 5. Temporal generalization gradients from Experiment 4A (upper panel) and Experiment 4B (lower panel). In each panel, temporal generalization gradients from the immediate and deferred comparisons are shown, along with those from the delayed comparisons with different delays (e.g., 5, 10, 20, and 30 s in the upper panel).

peaking at the shortest of the comparisons (comparison/standard ratio = 0.625). The leftward skew of the generalization gradient appeared to be about the same for all the delays from 5 to 30 s. All these suggestions were supported by statistical analysis.

An overall ANOVA found a significant effect of delay condition (immediate, deferred [0 s], 5 s, 10 s, 20 s, or 30 s),  $F(5, 90) = 23.07, p < .001$ , and a significant effect of comparison/standard ratio,  $F(6, 108) = 52.73, p < .001$ . There was also a significant Delay  $\times$  Comparison/Standard Ratio interaction,  $F(30, 540) = 18.75, p < .001$ . When an ANOVA was conducted on the delay conditions alone (5, 10, 20, or 30 s), the effect of delay just failed to reach conventional significance,  $F(3, 54) = 2.65, p = .058$ . There was, however, a significant effect of comparison/standard ratio,  $F(6, 108) = 80.97, p < .001$ , but no significant Delay  $\times$  Comparison/Standard Ratio interaction,  $F(18, 324) = 1.55, p = .07$ .

*Experiment 4B.* Results from Experiment 4B are shown in the lower panel of Figure 5. Inspection of the data suggested that there was a marked generalization gradient shift to the left at all delays, compared with the immediate and deferred (0 s) conditions. An overall ANOVA found a significant effect of delay condition (immediate, deferred, 2 s, 3 s, 4 s, or 5 s),  $F(5, 80) = 18.55, p < .001$ , and a significant effect of comparison/standard ratio,  $F(6, 96) = 20.97, p < .001$ . There was also a significant Delay  $\times$  Comparison/Standard Ratio interaction,  $F(30, 480) = 11.11, p < .001$ .

An ANOVA conducted on just the delay conditions (2, 3, 4, or 5 s) found no significant effect of delay,  $F(3, 48) = 2.48, p = .077$ , but there was a significant effect of comparison/standard ratio,  $F(6, 96) = 29.73, p < .001$ . There was no significant Delay  $\times$  Comparison/Standard Ratio interaction,  $F(18, 288) = 1.41, p = .126$ .

## Discussion

The results of Experiments 4A and 4B show that interposing even a short delay of 2 s between the end of presentation of B comparisons and the start of retesting with the A comparisons was sufficient to produce a marked leftward shift of the temporal generalization gradient, compared with either immediate testing or a 0-s delay (deferred testing). Furthermore, further increases in the delay led to no additional change in the shape of the generalization gradients, which suggests that the effect of interposed delay was immediate and dramatic. These results were obtained even though the same participants experienced all the different delay conditions in their particular experiment. The effects of even very short delays extend and confirm results from Experiments 2 and 3 and suggest that participants used a constructed A standard for their judgments of the delayed A comparisons, even after short interposed delays.

Although the results of Experiments 2, 3, and 4 provide a consistent picture of the effects of interposed delays on temporal generalization gradients and their data are consistent with our construction hypothesis, they do not eliminate all alternative possibilities as to why marked gradient shifts occurred. One is that the effect of delay was based solely on the delay from the A standards to the retesting of A comparisons and that the presence of the B standards and comparisons was not necessary to obtain gradient shifts. Experiments 1 to 4 all involved A and B standard/comparison sets, so they did not provide an appropriate control

condition to examine the effects of "pure delay" since A (i.e., without the presence of B), but Experiment 5 provides data on this issue. Another question that arises is whether the gradient shifts would be obtained no matter where the delays were positioned. In Experiments 2, 3, and 4, the delays were instituted after B comparisons, but suppose that the delays had been interposed after the initial (immediate) testing with the A comparisons: Would the gradient shift effect be obtained? Experiment 6 provides this control condition also.

## Experiment 5

Experiment 5 was designed to test whether the results from Experiments 2 to 4B occurred primarily because of interference effects (i.e., the effect of B on A) or simple mnemonic delay since presentation of the A standards. Experiment 5 used only one standard to investigate the effects of delay on temporal generalization performance, with performance on judgments of comparison stimuli that took place immediately after presentations of a standard (immediate) being contrasted with performance after a 15-s (short) or 45-s (long) delay.

## Method

*Participants and apparatus.* Fifteen University of Manchester students participated for course credit. Apparatus was as in Experiments 1 to 4.

*Procedure.* Participants were presented with the standard three times. This was a 500-Hz tone with a duration randomly sampled from a uniform distribution running from 400 to 600 ms. Between presentations of the standard there was a delay drawn from a uniform distribution ranging from 1,500 to 2,000 ms. There was then either no delay or a delay of 15 s or 45 s before the start of the comparisons. Then the participants were presented with seven comparison stimuli, which were the standard multiplied by 0.625, 0.750, 0.875, 1.000, 1.125, 1.250, or 1.375, with comparisons presented in a random order. Fifteen such blocks were given. On five of these (immediate) comparison testing began immediately after the standard presentation, for another five (short) a delay of 15 s was interposed between the end of the standard presentations and the start of comparison testing, and for another five (long) this delay was 45 s. The different blocks were presented in a random order. Other details were as in Experiments 1 to 4.

## Results and Discussion

Figure 6 shows the temporal generalization gradients obtained from Experiment 5. Inspection of Figure 6 indicates that the mean proportion of *yes* responses peaked at the standard with immediate testing. However, for testing after a short delay (15 s) and a long delay (45 s), the mean proportion of *yes* responses peaked at a duration shorter than the standard. The gradients for the conditions with delays did not appear flatter than the gradient for the responses after immediate testing, which suggests that responding did not become more variable as testing delay increased.

A repeated-measures ANOVA with within-subject factors of comparison/standard ratio and delay before presentation of comparisons found a significant effect of comparison/standard ratio,  $F(6, 78) = 22.24, p < .001$ ; but no significant effect of delay,  $F(2,$

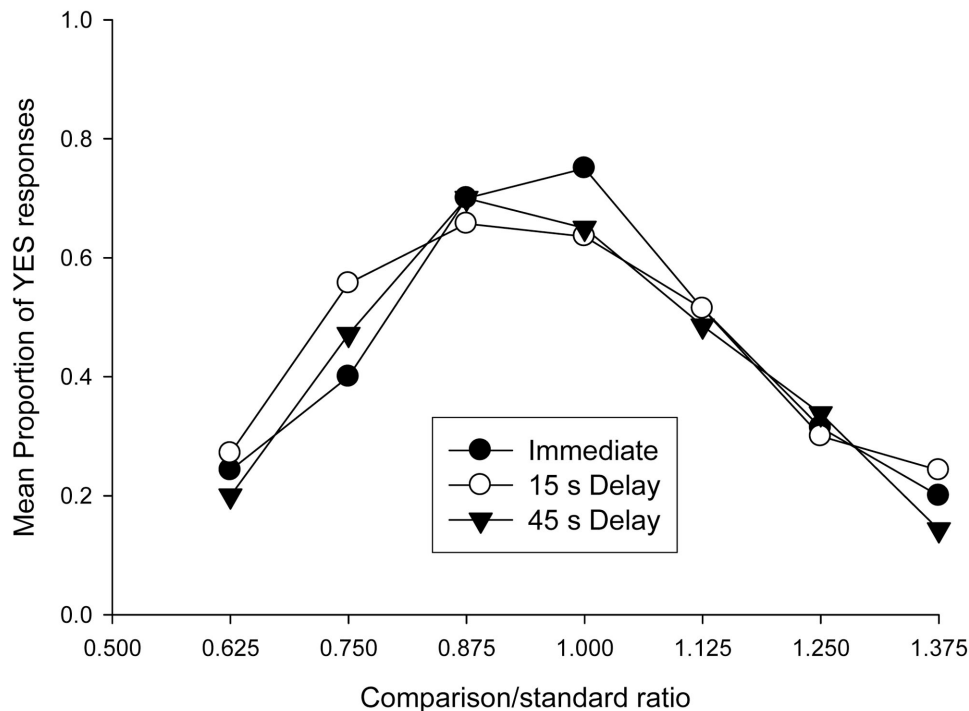


Figure 6. Temporal generalization gradients from Experiment 5. Data are shown separately for the immediate comparisons and those delayed by 15 and 45 s.

26) = 0.89,  $p = .424$ ; nor Comparison/Standard Ratio  $\times$  Delay interaction,  $F(12, 156) = 0.82, p = .626$ . These results suggest that imposing a delay before the presentation of the comparisons did not have a significant effect on the participants' ability to identify a comparison as the standard, in the absence of any B standard/comparison presentations.

### Experiment 6

The difference in performance between the deferred and delayed conditions observed throughout the series of experiments reported here suggests that a delay (however short) prior to the retesting of the A comparisons was required for interference to occur. Experiment 6 examined whether the location of the delay affected responding during retesting of the memory of standard A. In the delayed responding conditions of Experiments 2 to 4, not only were A comparisons tested after a delay since the A standard/comparison set (during which the B standard/comparison set was interposed), but there was also a delay after the end of the B comparisons, and generalization gradients were markedly skewed even when this last delay was short. Testing with A comparisons immediately after B comparisons finished (our deferred conditions), in contrast, produced generalization gradients that approximately peaked at the standard. The delayed conditions we used thus involved a delay since the B standard/comparison set as well as since the A standard/comparison set, whereas the deferred conditions involved only the former. Perhaps one of the essential ingredients in producing the highly skewed generalization gradients is the delay since B, which suggests that the gradients might

only have been skewed as a result of some sort of change in the memory of B as well as A.

If the delay since B plays a critical role, then interposing delays elsewhere in the block sequence should change the results. For example, suppose that delays ranging from 0 to 30 s were imposed after the initial A standard/comparison set but before the B standard/comparison set, with the A comparison retesting taking place immediately after the last B comparison. This condition would have the same delay since the initial A standard/comparison set as some of our delayed conditions but the same B standard/comparison set prior to retesting of A comparisons as our deferred conditions. What would behavior look like in this case? If generalization gradients were skewed markedly to the left, the delay since the B standard/comparison set apparently would not be critical, whereas if the gradients were peaked near the standards, the latter delay would be implicated as a critical element in producing skewed generalization gradients.

Experiment 6 tested this condition, using delays ranging from 0 to 30 s (the same values as used in Experiment 4A), but this time interposed after the initial testing of the A comparisons and before the B standard/comparison set. The retesting of A comparisons occurred immediately after the last B comparison, as in the deferred conditions of our earlier experiments.

### Method

*Participants and apparatus.* Eighteen University of Manchester students participated for course credit. Apparatus was an in Experiments 1 to 5.

*Procedure.* The procedure was identical to that used in Experiment 4A except that in different blocks a delay of 0, 5, 10, 20, or 30 s was interposed prior to the presentation of standard B. Retesting with the A comparisons commenced immediately after the last B comparison presentation, as in the deferred conditions of other experiments. There was a total of 30 blocks.

### Results and Discussion

Figure 7 shows data from the immediate, deferred (0-s delay), and delayed (5-s, 10-s, 20-s, and 30-s delay) testing conditions. Inspection of the data suggests that there was no systematic deterioration in performance from immediate testing to any of the delayed retests and that peak responding remained at or around the standard in all testing conditions. These suggestions were supported by statistical analysis.

An overall ANOVA found no effect of testing condition (immediate, 0 s, 5 s, 10 s, 20 s, and 30 s),  $F(5, 85) = 2.27, p = .55$ . There was a significant effect of comparison/standard ratio,  $F(6, 102) = 18.41, p < .001$ , but no significant Testing Condition  $\times$  Comparison/Standard Ratio interaction,  $F(30, 510) = 1.19, p = .224$ .

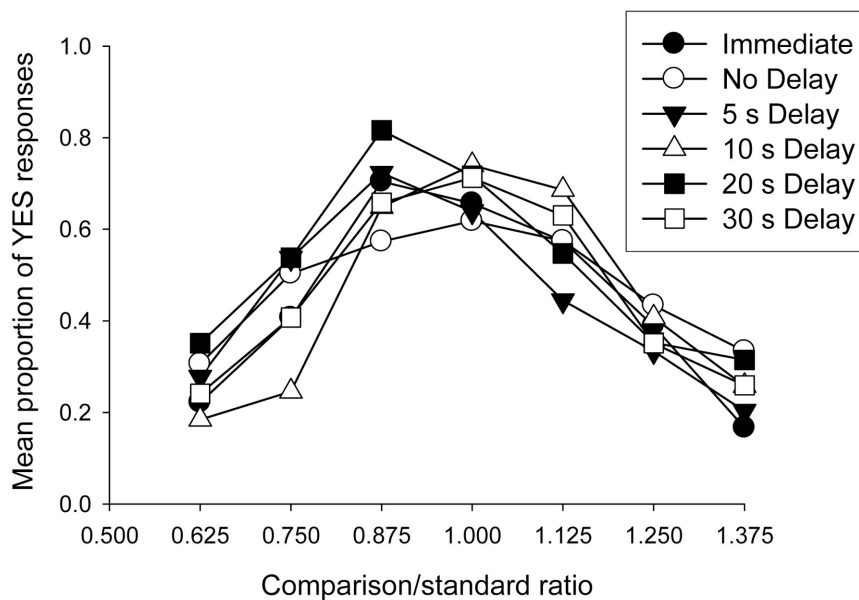
It is clear from comparing data in Figure 7 with those in Figures 3, 4, 5, and 6 that relocating the delay from after the B comparisons (in Experiments 2, 3, and 4) to prior to the presentation of the B standard (Experiment 6) changed the shape of the temporal generalization gradients. When the delay was located after B comparisons, even short delays produced marked skewing of temporal generalization gradients, with peaks of responding at or close to the most remote stimuli presented, whereas when the delay was located after the A comparisons, gradients were approximately centered on the standard value (i.e., comparison/standard ratio = 1.0). These results suggest, obviously, that a delay after the end of

the B comparisons is an essential ingredient in producing the markedly skewed generalization gradients. One possibility, therefore, is that the skew was a result of some sort of complex interaction between the interference from the B standard/comparison set and some sort of degradation of the memory of B itself.

### General Discussion

Perhaps the most striking feature of the results presented above is the contrast between the interfering effects obtained with different combinations of potential delay since initial learning of a standard and subsequent testing. If a substantial delay was imposed between presentations of a standard and its subsequent testing, without any potentially interfering second standard being presented, little or no effect on temporal generalization gradients was obtained (Experiment 5), which suggests that pure delay had a minimal effect on the integrity of memory. Likewise, interposing a putatively interfering condition (B standards and comparisons) between a target standard (A, in this case) and its subsequent retesting produced a small effect of flattening the temporal generalization gradient, but no marked change in gradient shape, if the retesting of A occurred without any delay after the B comparisons had been presented (Experiment 1 and the deferred conditions of Experiments 2, 3, and 4). These results, suggesting that some manipulations have small effects on temporal memory, are consistent with other work by Jones and Wearden (2004) and Grondin (2005), who likewise found small, albeit usually significant, effects, and together suggest that temporal reference memory is robust across manipulations of delay since the memory was established and across some sorts of interference.

Contrasting sharply with this picture is the effect obtained in our Experiments 2, 3, and 4, which showed that interposing a delay



*Figure 7.* Temporal generalization gradients from Experiment 6, in which the delay was interposed after the A comparisons. Data are shown separately for the immediate, no delay, and various delay conditions (see text for details).

between the end of the B comparisons and retesting of A produced very marked shifts in the shape of the temporal generalization gradient, in all cases skewing the gradient so much that the peak appeared at the shortest or longest of all the comparison durations presented. Experiment 4 also showed that even a short interposed delay would produce this effect, with no further change in gradient shape at longer delays, and Experiment 6 showed that the interposed delay had a critical location after the B comparisons.

Taken together, the results of all the studies reported in this article suggest that the important "ingredients" for producing marked shifts in temporal generalization gradients during delayed testing of A compared with immediate or deferred testing were (a) the presence of a delay since the B comparisons and (b) the presence of B and its comparisons, the putatively interfering stimuli. When one of these factors was present but not the other one, generalization gradients differed little between the testing conditions.

How can we interpret the marked shift in temporal generalization gradients sometimes found? The potential explanation that we advanced is in terms of construction: The conditions rendered the memory of A effectively unusable for the participants, who, although they might have had no clear recollection of A, knew something about it, namely its relation to B (slightly or a lot shorter or longer or more or less the same) and so used their memory of B to construct a standard for A, which they then used for the delayed A comparisons. This construction process was flawed and resulted in the standard employed being markedly shorter or longer than the real A, therefore shifting the peak of the temporal generalization gradient toward the longest or shortest comparison durations presented, at least with the range of durations that we used. This interpretation is supported most clearly by the results of Experiment 3 (see Figure 4, e.g.), in which the direction of the shift of the gradient was determined by the duration relations between B and A.

But what is the basis for this construction process? As mentioned earlier, the effects of interposing delays between testing of B comparisons and retesting of A comparisons have some similarities with results often interpreted in terms of AL theory (Helson, 1964). The basic idea of AL theory is that judgments of the magnitude or other characteristics of stimuli are made with respect to an AL that is determined by characteristics of the set of stimuli judged, of which the to-be-judged stimulus is a member (e.g., their range); potential "anchors" or "references" with which the stimuli judged might be compared; or context values derived from background stimuli. Fitting AL theory to data in practice is complex, and the equations used are commonly tailored specially for the case modeled (see examples in Helson, 1964, Chapter 4), but the general idea is that anchors set the AL with respect to which subsequent stimuli are judged.

AL effects are found in duration judgments, as in the study of Behar and Bevan (1961), who presented people with visual stimuli ranging in duration from 1 to 5 s, where the task was to classify each duration presented on an 11-point scale ranging from *very very short* to *very very very long*. Three conditions were used: In a no-anchor condition, no reference stimulus was provided, but in the other two conditions, every fourth stimulus was an anchor, which, for different conditions, was either short (0.2 s) or long (9 s). The judgment categories used for the 1- to 5-s stimuli were systematically shorter in the long anchor condition than in the

short one, with the no-anchor condition producing data somewhere in between.

The obvious interpretation of Behar and Bevan's (1961) result is that the anchor stimuli set the AL, and this shifted the judgments accordingly. At first sight, such an effect appears reminiscent of the ones we obtained in the delayed conditions of Experiments 2 to 4B. In most of the cases reported, the B standards and B comparisons were longer than the A standard and comparisons, so the longer B values might have shifted the AL. Thus, by comparison to the immediate condition, in which the AL was presumably determined just by the A standards and comparisons, the delayed A comparisons seemed shorter than they did during immediate testing, resulting in a leftward shift of the generalization gradient. This interpretation appears to be supported by two additional results. In Experiment 2, the close B standard/comparison values produced a less marked shift in generalization gradient than the far comparisons. This is consistent with the close B condition shifting the AL less than the far one. In addition, when the B standard/comparison values were systematically shorter than the A ones, as in our Experiment 3, the generalization gradient was shifted to the right, and this is compatible with the idea that the shorter B values reduced the AL, thus making the delayed A comparisons seem longer.

However, although this use of ideas derived from AL theory fits some of our data well at first glance, there are some aspects of our results that remain unexplained. If shifts in AL resulting from B standard/comparison sets produced our effects, why was there no shift of generalization gradients in the deferred conditions or when a delay was interposed between the initial testing of A comparisons and B standards (as in our Experiment 6)? In both these cases, B standard/comparison sets would be expected to shift AL, but effects depend on a delay between the B standard/comparison block and retesting of A comparisons and do not occur otherwise. At the very least, this suggests that the shifts in AL depend on an additional change based on delay.

In addition, the similarity between effects reported in studies such as that of Behar and Bevan (1961) and our own may be more apparent than real, as a result of differences in the responses required from participants. In Behar and Bevan's study, durations were classified on an 11-point scale, whereas in our studies, we did not ask people to judge how long the delayed A comparisons seemed to last on some categorical scale but asked whether they were or were not identical to the previously presented A standard. Even if all our delayed A comparisons seemed shorter or seemed longer than they did during immediate testing, no effect on generalization gradient might be expected, because the memory of the A standard would also, surely, seem shorter or longer because of the AL shift. However, it may be that the construction process that we propose uses the AL established by the B standard/comparison set as the basis for the constructed standard, which is then used to judge delayed A comparisons.

Although the marked shifts in temporal generalization gradients and our explanation of them may seem very novel, in fact, some similar phenomena have been reported before, both in the study of timing and outside it. The idea that participants construct a temporal standard that is used for subsequent time judgments is, in fact, a fairly commonplace notion in the study of temporal bisection. Temporal bisection can be investigated with different procedures, but a common one (called classification by Wearden &

Ferrara, 1995) proceeds as follows. People initially receive examples of a standard short and a standard long duration (e.g., tones 200- and 800-ms long, respectively), then are required to classify a range of comparison values (e.g., from 200 to 800 ms in 100-ms steps) in terms of their similarity to the short and long standards previously presented. In spite of the fact that the obvious thing to do is to compare each comparison duration value with the remembered short and long standards and respond according to judged similarity, and in spite of the fact that theoretical models that are slightly elaborated versions of this idea fit data well (e.g., Wearden, 1991a, 1991b), there is evidence that people do not afford the standards any special status. In fact, it appears that people construct some criterion value based on the durations of all the stimuli received. Wearden and Ferrara (1995) proposed that this criterion is the arithmetic mean of all the stimuli presented, so if a comparison is judged longer than the mean the participant responds *long*, and if it shorter he or she responds *short*. This is not the only possibility, but the general point here is that there may be a number of situations in which people do not simply remember the standards given to them in timing tasks but instead construct some other basis for responding, using the standards as well as other duration values (Allan, 2002a, 2002b; Allan & Gerhardt, 2001).

Our proposition that people constructed standards for the delayed judgments of the A comparisons suggests that people were, in fact, constructing a sort of “false memory” of A—that is, a time value that was not actually presented in the experiment. Making this analogy between performance on our task and more common sorts of memory errors immediately suggests that the effects we obtained may be frequently found in other sorts of memory. For example, if people receive a categorized word list (e.g., containing examples of words for fruits, animals, articles of furniture), their memory of the actual words presented may be weak after some manipulation, yet they may remember the word categories and, in fact, construct new items using this knowledge, so they say “dog” when the word was never presented. Such effects have been found by Roediger and McDermott (1995), who looked at false recall and recognition using lists of related words. Participants were presented with lists of 12 words, all of which were strongly related to a nonpresented word, the critical lure. Their task was to listen to the lists and then recall each word after each list had been presented. Finally, after all lists had been recalled, a recognition test was completed, including 12 studied words, the critical lures, unrelated words, and weakly related words. Participants also rated their confidence that a word had appeared in the list by stating whether it was probably old, surely old, possibly new, or surely new. The critical lure was consistently recalled, with about the same probability of recall as a word presented in the middle of the list. In the recognition test, critical lures were rated old almost as often as presented words.

While it is generally accepted that recalling the critical lure in the Roediger and McDermott (1995) study constituted the creation of a false memory, Miller and Wolford (1999) suggested that the false recall of the critical lure might have been the result of a criterion shift. They proposed that participants developed meta-knowledge of the relationships between the stimuli in the lists during the task. This metaknowledge enabled participants to produce words that were not presented in the task but were related to the categories of words presented, often the critical lure. Likewise, a criterion shift is better able to account for the shifts in temporal

generalization gradients found in this study than just the creation of an actual false memory as such. The bidirectional effect of B observed in Experiment 3 suggests that it is unlikely that the shifts in peak responding were the result of some sort of decay process affecting participants’ memory of A, resulting in it becoming much shorter or longer than it actually was. Instead, as suggested before, it seems that participants abandoned their use of their memory of A altogether and instead used a new construction-based strategy using the relationship between A and B to create a new standard for comparison.

The principal effects of our manipulations of delayed testing appeared to be on shifts in the peak of the temporal generalization gradient, but it may be that other aspects of the gradient were also affected, such as its spread. In general, the question arises of how “normal” the gradients were from the delayed and deferred conditions, compared with those obtained in more standard procedures, such as that of Wearden (1992), and in what ways, if any, the psychological processes involved in generating the abnormal gradients differed from those used in generating normal ones. One way of addressing this question, perhaps the only way, is to use computer modeling, employing variants of a standard model of temporal generalization.

We conducted extensive computer modeling of most of the results obtained in the experiments reported above using a variant of the modified Church and Gibbon (1982; MCG) model, developed by Wearden (1992) as an account of the performance of humans on temporal generalization tasks. As its name suggests, the MCG model is a variant of an earlier model proposed by Church and Gibbon (1982) as an explanation of temporal generalization performance in rats. Wearden (2004) discussed in some detail the use of ideas derived from SET for modeling of time judgments in humans. As noted above, SET proposes that time judgments are based on an interaction of clock, memory, and decision processes, and the principal difference between one model and another lies in the decision processes proposed. For example, in temporal generalization, the basis of the decision is some sort of comparison between the memory of the standard and the representation of the just-presented comparison duration. In other tasks, such as bisection (Wearden, 1991b; Wearden & Ferrara, 1995), the judgment of some comparison duration may depend on its relation to more than one standard or to some sort of average of the durations presented. As Wearden (2004) noted, the different models have many features in common, often using the principle of thresholded normalized difference; that is, the difference between two (or more) duration representations is expressed as a fraction of something else (normalized), then compared with a threshold. Modeling can be said to be compatible with SET in a number of ways. For example, the modeling of performance on one task can use performance rules similar to those used for another task previously treated by SET. In addition, some of the parameters of the models can be related to principles of SET and thus provide tests of features, such as the scalar property of variance (Lejeune & Wearden, 2006).

The MCG model proposes that a person responds *yes* (i.e., judges that a comparison duration is the same as the standard) when

$$\text{abs}(s^* - t)/t < b^* \quad (1)$$



In Equation 1,  $s^*$  is a sample drawn from the memory of the standard duration,  $s$ ;  $t$  is the duration of the comparison stimulus;  $b^*$  is a threshold, and  $\text{abs}$  indicates absolute value. The memory of the standard contains trial-to-trial variability, whereas  $t$  is supposed to be timed without variance. To produce  $s^*$ , a Gaussian distribution is constructed with a mean of  $s$  and a coefficient of variation,  $c$ , which is one of the parameters of the model, and on each trial a sample ( $s^*$ ) is randomly drawn from this distribution. The mean of the distribution is  $s$ , so, on average, the standard is remembered accurately, and increasing or decreasing  $c$  increases or decreases trial-by-trial variance in the memory of  $s$ , making this memory more or less “fuzzy” over trials. The threshold also varies from trial to trial; the sample  $b^*$  is drawn from a Gaussian distribution with a mean of  $b$  and some coefficient of variation, which in most simulations using the MCG model in recent articles, including those used here, has a constant value of 0.5 (i.e., a standard deviation half the mean of  $b$ ).

Equation 1 produces generalization gradients that (a) are peaked at the standard,  $s$ ; (b) are skewed to the right, so that comparison durations longer than the standard are generally more confused with it than those that are shorter by the same amount; and (c) exhibit superposition when plotted on the same relative scale. All these properties are found in data from normal temporal generalization—that is, the experimental situation in which  $s$  is constant for the whole experimental session (see Wearden, 1992, and Wearden, Denovan, Fakhri, & Haworth, 1997, for examples).

The experiments in the present article all used the “changing standard” temporal generalization method introduced by Jones and Wearden (2003), whereby the standard is in force only for one set or one block of comparisons before being replaced by another standard/comparison set. Generalization gradients from the changing standard method sometimes differ from those obtained with normal temporal generalization in two main ways. First, gradient shape can be different from normal (with peak location displaced away from  $s$ , a leftward skew, or both), and, second, superposition of generalization gradients, when these are plotted on the same relative scale, is often violated (Wearden & Lejeune, in press). For the present purposes, only the first of these concerns us. An additional problem with modeling the present data is that it is obvious that in some cases the peak of the distribution of *yes* responses was very markedly displaced from  $s$  (e.g., in Figure 4). A potential solution to modeling data from situations in which generalization gradient peaks are displaced from  $s$  is to use Equation 2:

$$\text{abs}(ks^* - t)t < b^*. \quad (2)$$

Here, all terms are as in Equation 1, except for the addition of a value  $k$ , which multiplies  $s$  and allows the internal representation of the standard,  $ks$ , to differ from its real value,  $s$ . If  $k = 1.0$ , then  $s$  is represented accurately, and Equation 2 converges on the simple MCG model, Equation 1. If  $k < 1.0$ , the standard  $s$  is represented as shorter than it really is, and if  $k > 1.0$ , it is represented as longer. Equation 2 has been used successfully in articles modeling generalization performance in children (e.g., Droit-Volet, Clément, & Wearden, 2001), in which gradient peaks are sometimes displaced from  $s$ . Using Equation 2 to model the present data enables us to estimate the actual value of the standard

used in the putatively constructed conditions obtained when testing of  $A$  comparisons is delayed.

We simulated all the generalization gradients shown in Figures 1 to 7. The model described by Equation 2 was instantiated as a Visual Basic 6.0 program. Each fit yielded three theoretical parameters:  $c$ , the coefficient of variation of the memory distribution of  $s$ ;  $b$ , the threshold mean (the coefficient of variation of the threshold was always 0.5); and  $k$ , the multiplier for the memory of  $s$ . For each fit, 10,000 trials were run, and the three parameters were varied by small increments over a wide range until the smallest mean absolute deviation (MAD) between predictions of the model and data was found. The MAD was the sum of the absolute deviations between the predictions of the model and the data points, divided by the number of data points, which was usually seven.

Table 1 shows parameter values from fits of the model and MAD values for 45 cases representing all the generalization gradients shown in Figures 1 to 7. MAD values less than 0.05 have been used as an index of reasonable fit in previous work (e.g., Wearden, 1992; Wearden et al., 1997). One gradient of the total of 45 could not be fitted, although the reason for its irregular shape is unclear; 3 more could be fitted if one data point was ignored; and the other 41 produced fits with MAD values less than or equal to 0.05, using all data points. Figure 8 shows some representative cases of fits of Equation 2 (lines in each panel) to data from selected data conditions. These were chosen to illustrate fits to immediate, deferred, and delayed conditions (unconnected symbols), including those in which the generalization gradients were markedly skewed to the left or right.

Inspection of the results shown in Table 1 suggests that some generalizations can be drawn. First, although threshold values ( $b$ ) varied between experiments, virtually all values were between 0.2 and 0.3 and appeared to show little systematic effect of condition, in that values derived from immediate, deferred, and various delayed conditions appeared to overlap. Second, although most coefficients of variation of memory representations ( $c$ ) were found between .15 and .28, some systematic variation seemed to be present. For example, if we compare  $c$  values from the deferred and the corresponding immediate condition in experiments in which both  $A$  and  $B$  standards were used (11 cases), values from fits to the deferred condition were greater than those from the immediate one in 10 of 11 cases, suggesting an inflation of variability in deferred compared with immediate conditions. Comparison of values from delayed conditions with immediate ones (18 cases) showed a similar trend.  $c$  values were higher in 15 of 18 cases, equal in 2, and lower in 1. However, comparison of  $c$  values derived from delayed and deferred conditions showed a less clear picture. In 5 (of 17) cases the delayed condition produced the higher  $c$  value, in 1 case it was the same, and in the other 11 cases the delayed condition produced a lower  $c$  value. Overall, therefore, it seems that (a) interposing a delay between standard presentation and comparison testing, whether this was a deferred or delayed condition, produced an increase in memory variance with respect to immediate testing and (b) values from delayed and deferred conditions were more intermixed. However, modeling of data from the pure delay conditions of Experiment 5 shows that this manipulation had little or no effect on any of the three parameters of the model ( $c$ ,  $b$ , or  $k$ ), confirming the conclusions to be drawn from the statistical analysis of the results.

Table 1  
Parameter Values From Fits of Equation 2 to Data Shown in  
Figures 1 to 7

Experiment	Condition	Delay	<i>c</i>	<i>b</i>	<i>k</i>	MAD
1	Close	Imm	.25	.29	0.97	.05
1	Close	Def	.25	.30	1.00	.03
1	Far	Imm	.16	.30	0.92	.04
1	Far	Def	.23	.30	0.92	.05
1	Same	Imm	.25	.29	0.95	.04
1	Same	Def	.30	.30	0.96	.05
2	Close	Imm	.22	.25	0.92	.05
2	Close	Def	.27	.24	0.86	.04
2	Close	Del	.24	.26	0.78	.03
2	Far	Imm	.20	.26	0.93	.04
2 <sup>a</sup>	Far	Def	.25	.24	0.97	.05
2	Far	Del	.22	.27	0.60	.03
2	Same	Imm	.20	.21	0.96	.04
2 <sup>a</sup>	Same	Def	.27	.19	0.96	.03
2	Same	Del	.28	.25	0.98	.05
3	Shorter	Imm	.15	.22	0.90	.04
3	Shorter	Def	.29	.24	0.85	.05
3	Shorter	Del	.18	.21	1.33	.05
3	Longer	Imm	.18	.22	0.95	.05
3	Longer	Def				
3	Longer	Del	.26	.27	0.70	.04
3	Same	Imm	.15	.22	1.00	.05
3 <sup>a</sup>	Same	Def	.20	.22	1.00	.05
3	Same	Del	.23	.22	0.97	.04
4A	Longer	Imm	.19	.24	1.00	.03
4A	Longer	Def	.25	.23	1.00	.03
4A	Longer	5 s	.23	.22	0.65	.04
4A	Longer	10 s	.20	.25	0.65	.03
4A	Longer	20 s	.20	.25	0.65	.05
4A	Longer	30 s	.19	.22	0.65	.05
4B	Longer	Imm	.17	.25	1.00	.04
4B	Longer	Def	.28	.26	0.95	.04
4B	Longer	2 s	.28	.25	0.63	.05
4B	Longer	3 s	.42	.29	0.58	.05
4B	Longer	4 s	.45	.22	0.58	.04
4B	Longer	5 s	.30	.16	0.65	.05
5		Imm	.20	.23	0.92	.04
5		15 s	.20	.22	0.90	.04
5		45 s	.21	.22	0.90	.01
6	Longer	Imm	.21	.22	0.90	.04
6	Longer	Def	.26	.23	0.96	.04
6	Longer	5 s	.22	.22	0.90	.04
6	Longer	10 s	.20	.23	0.97	.03
6	Longer	20 s	.22	.27	0.89	.04
6	Longer	30 s	.22	.22	0.95	.04

Note. Experimental condition items pertain to B (e.g., *longer* indicates that B was longer; 5 s indicates that B came after a 5-s delay). Parameter values: *c* = coefficient of variation of memory representation of the standard; *b* = mean threshold; *k* = multiplier of standard. The coefficient of variation of the threshold distribution was always 0.5. MAD = mean absolute deviation; Imm = immediate; Def = deferred; Del = delayed.  
<sup>a</sup> Fit was based on six rather than seven data points.

The parameter that was clearly the most systematically affected by the condition was *k*. When a delay was interposed after testing of B comparisons and further testing of A comparisons (as in Experiments 2 to 4B), the effect on *k* was very marked and depended on the relation between B and A. Having B longer than A reduced *k* values to a range from 0.58 to 0.78 (the average of 11 conditions was 0.65), whereas when B was shorter than A a value of 1.33 was obtained. This suggests that the delayed conditions

caused participants to remember the standard as around 33%–35% shorter or longer than it really was or to construct a standard 33%–35% shorter or longer. In contrast, when  $A = B$ , the *k* value was 0.97. In general, however, in immediate and deferred conditions, *k* values were always 1.00 or less, suggesting that the changing standard method produced representations of *s* that were shorter than its real time value. We have no explanation as to why this should be, although research on this issue is ongoing.

In general, modeling confirmed the suggestions obtained by inspection of the data in Figures 2 to 7 and statistical analysis, namely that in the delayed conditions the memory representation of the mean duration of the standard was very markedly affected. The modeling further shows that this was by far the most major change between conditions, with effects on the *c* and *b* parameters (reflecting memory variability and decision thresholds, respectively) being much more minor, even when some systematic trends were observable. Perhaps the most remarkable result of the modeling is that the variant of the MCG model fitted the data at all, considering that, superficially, the generalization gradients obtained in some of our delayed testing conditions bore little resemblance to those obtained in normal temporal generalization, as in Wearden (1992) and Wearden et al. (1997). Nevertheless, the modeling suggests that the basic underlying psychological processes at work when temporal generalization gradients are taken from delayed conditions are not different from those in the normal case, except for the different standard value used. The striking correspondence between the data points and the model's fit, illustrated in Figure 8, emphasizes their similarity very strongly. Here, gradients could be exactly peaked at the standard, have their peaks slightly displaced from the standard and be slightly flattened, or be very markedly skewed toward the smallest or largest comparison/standard ratios and still be consistent with the processes specified in Equation 2. As mentioned above, the MCG model and its variants were originally developed from work by Church and Gibbon (1982), which modeled temporal generalization gradients in animals. The good fit of the MCG model to previous data (e.g., Wearden, 1992) and to data from most of the conditions used in the present study thus provides theoretical continuity over a range of tasks and emphasizes the usefulness of the clock-memory-decision structure of SET as an explanatory tool, even of data like those shown in some of our figures, which deviate markedly from those normally found in temporal generalization.

In the present article, we have interpreted the effects of our delay manipulations on memory of the standard, *s*. But what kind of memory is really involved, and can working and reference memories for duration be clearly distinguished? Jones and Wearden (2004) provided a lengthy discussion of the development of the concepts of reference and working memory in timing studies, tracing the idea back to early studies of animal timing, for which the distinction between a semipermanent "reference memory" (e.g., of the time of reinforcement) and a more fluid "working memory" (e.g., of the elapsed time in the current trial) seems intuitively very reasonable. They also pointed out that the distinction transfers less well to human studies, in which "references" may be stored only for a few minutes or even a few seconds, without any suggestion of permanence over hours or days.

There are two ways of addressing the question of what sort of memory is involved in the work reported in the present article, although these are not mutually contradictory. One approach is to

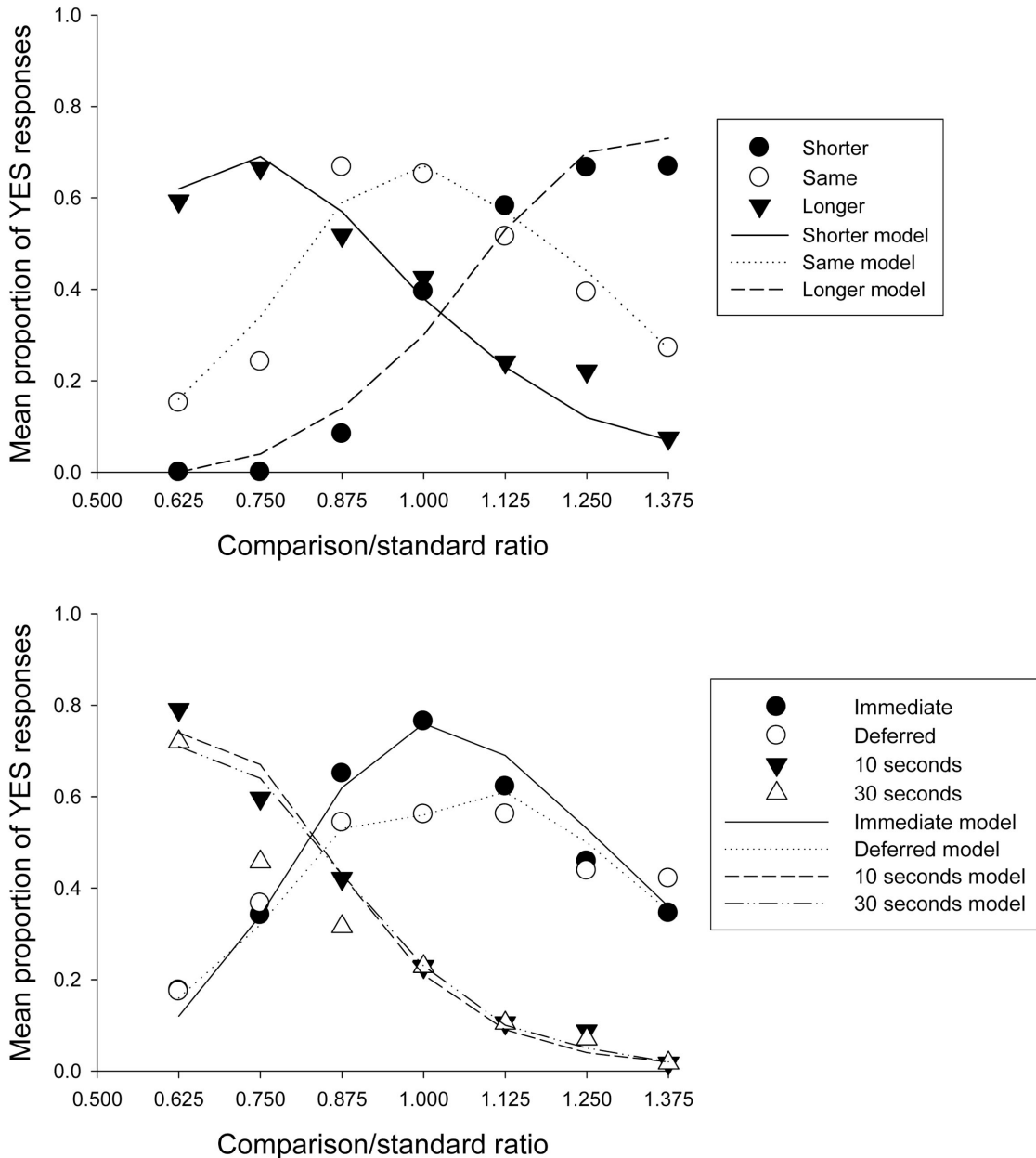


Figure 8. Selected temporal generalization gradients (unconnected points) and fits of the model embodied in Equation 2 (lines). Upper panel: data and modeled values from the *longer*, *shorter*, and *same* delayed conditions of Experiment 3. Lower panel: Data and modeled values from the immediate and deferred conditions of Experiment 4A and two delayed conditions (delays of 10 and 30 s). In all cases, the obtained and modeled proportion of *yes* responses is plotted against the comparison/standard ratio.

say that the memory of the standard in temporal generalization experiments in humans has been considered to be an example of reference memory in a number of previous studies (e.g., from Wearden, 1992, onward), so, by this previous custom and practice, the effect of delay in our studies was a reference memory effect. The second approach is to argue that the terminology of reference and working memories in human timing carries unwanted surplus meaning and, ideally, should be abandoned, although it is perhaps now too established to be replaced in practice. A

better terminology might be between standards that are used for only a single trial (single trial referents, as in episodic temporal generalization; see Wearden & Bray, 2001) and those that are used for multiple trials (blocks of comparisons, as in the changing standard method, or whole sessions, as in normal temporal generalization), which we might call multiple trial referents. This distinction has the virtue of neutrality and has no implications that one or another of the memories is long term or more permanent or stable than the other. It does not necessarily

assume that behavior in single-trial or multiple-trial referent conditions will be the same when the same physical durations are used as standard and comparison (and Filippopoulos & Wearden, 2006, provided evidence that it is not), nor does it assume any structural difference between the two memory types, although one may well exist.

Given this terminology, our effects clearly involve multiple trial referents, and the data suggest that the effects we observed when delays were interposed between the presentation of such referents and their subsequent testing by comparison presentations differ from the effects of mnemonic delays when single-trial referents are used, as in the subjective shortening experiments of Wearden and Ferrara (1993) and others (see Wearden et al., 2007, for review and discussion).

In conclusion, we noted earlier in this article that the study of memory for duration, and the study of reference memory in particular, is relatively neglected by comparison to work on the processes subserving the internal clock proposed by SET (e.g., Penton-Voak, Edwards, Percival, & Wearden, 1996), in spite of the prominence given to memory processes in this theory. We know relatively little about the formation and maintenance of reference memories (Jones & Wearden, 2003) and little about potential interference between different references (Jones & Wearden, 2004). Work in the present article suggests that there are conditions in which reference memory seems little changed by interposed delays (Experiments 5 and 6) and some sorts of potential interference (e.g., the deferred conditions of Experiments 1 to 4), while in other circumstances some combinations of delay and interference produce an apparently severe loss of the reference, which might be the potential basis of performance when people judge comparison stimuli (e.g., the delayed conditions of Experiments 2 to 4). This combination of apparent stability in some cases and fragility in others is puzzling, but the consistency of results between our experiments and the consistency of the process embodied in the model that fits almost all our data suggest that the effects are real and warrant further investigation, which may help us to better understand the process involved in the development, maintenance, and use of reference memory in timing tasks.

## References

- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*, 101–117.
- Allan, L. G. (2002a). Are the referents remembered in temporal bisection? *Learning and Motivation*, *33*, 10–31.
- Allan, L. G. (2002b). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *55*(B), 43–60.
- Allan, L. G., & Gerhardt, K. (2001). Temporal bisection with trial referents. *Perception & Psychophysics*, *63*, 524–540.
- Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *American Journal of Psychology*, *74*, 17–26.
- Brown, S. W., & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica*, *75*, 103–121.
- Church, R. M., & Gibbon, G. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*, 165–186.
- Creelman, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, *34*, 582–593.
- Droit-Volet, S., Clément, A., & Wearden, J. H. (2001). Temporal generalization in 3- to 8-year-old children. *Journal of Experimental Child Psychology*, *80*, 271–288.
- Droit-Volet, S., Wearden, J. H., & Delgado-Yonger, M. (2007). Short-term memory for time in children and adults: A behavioral study and a model. *Journal of Experimental Child Psychology*, *97*, 246–264.
- 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: Comparative and Physiological Psychology*, *50*(B), 217–237.
- Filippopoulos, P. C., & Wearden, J. H. (2006, July). *The role of reference memory in time judgements*. Paper presented at the meeting of the Experimental Psychology Society, Plymouth, England.
- Fortin, C. (1999). Short-term memory in time interval production. *International Journal of Psychology*, *34*, 308–316.
- Fortin, C., & Breton, R. (1995). Temporal interval production and processing in working memory. *Perception & Psychophysics*, *57*, 203–215.
- Fortin, C., & Rousseau, R. (1987). Time estimation as an index of processing demand in memory search. *Perception & Psychophysics*, *42*, 377–382.
- Fortin, C., & Rousseau, R. (1998). Interference from short-term memory processing on encoding and reproducing brief durations. *Psychological Research*, *61*, 269–276.
- Fortin, C., Rousseau, R., Bourque, P., & Kirouac, E. (1993). Time estimation and concurrent non-temporal processing: Specific interference from short-term-memory demands. *Perception & Psychophysics*, *53*, 536–548.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.
- Grondin, S. (2005). Overloading temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 869–879.
- Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology*, *10*, 228–240.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper and Row.
- Jones, L. A., & Wearden, J. H. (2003). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *56*(B), 321–343.
- Jones, L. A., & Wearden, J. H. (2004). Double standards: Memory loading in temporal reference memory. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *57*(B), 55–77.
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, *59*, 1875–1908.
- Lieving, L. M., Lane, S. D., Cherek, D. R., & Tcheremissine, O. V. (2006). Effects of delays on human performance on a temporal discrimination procedure: Evidence of a choose-short effect. *Behavioural Processes*, *71*, 135–143.
- Miki, A., & Santi, A. (2005). The perception of empty and filled time intervals by pigeons. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *58*(B), 34–45.
- Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, *106*, 398–405.
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1770–1787.
- 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.
- Rattat, A. C., & Droit-Volet, S. (2005). The long-term retention of time: A

- developmental study. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 58(B), 163–176.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814.
- Schneider, W. (1988). Micro Experimental Laboratory [computer software]. Pittsburgh, PA: Psychology Software Tools.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” *Psychological Monographs*, 77(Whole No. 576).
- Wearden, J. H. (1991a). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation*, 22, 59–83.
- Wearden, J. H. (1991b). Human performance on an analogue of an interval bisection task. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 43(B), 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: Vol. 2. Information-processing perspectives* (pp. 21–39). Göttingen, Germany: Hogrefe & Huber.
- Wearden, J. H. (2004). Decision processes in models of timing. *Acta Neurobiologiae Experimentalis*, 64, 303–317.
- Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory: Episodic temporal generalization and bisection in humans. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 54(B), 289–310.
- Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 502–511.
- Wearden, J. H., & Ferrara, A. (1993). Subjective shortening in humans' memory for stimulus duration. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 46(B), 163–186.
- Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection in humans. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 48(B), 289–310.
- Wearden, J. H., Foran, K., & Goodson, G. (2007). Subjective shortening with filled and unfilled auditory and visual intervals in humans? *Quarterly Journal of Experimental Psychology*, 60, 1616–1628.
- Wearden, J. H., & Grindrod, R. (2003). Manipulating decision processes in the human scalar timing system. *Behavioural Processes*, 61, 47–56.
- Wearden, J. H., & Lejeune, H. (in press). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*.
- Wearden, J. H., Parry, A., & Stamp, L. (2002). Is subjective shortening in human memory unique to time representations? *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 55(B), 1–25.

Received February 4, 2007

Revision received September 3, 2007

Accepted September 11, 2007 ■

## Low Publication Prices for APA Members and Affiliates

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.