THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2002, *55B* (1), 1–25

Is subjective shortening in human memory unique to time representations?

J.H. Wearden, A. Parry, and L. Stamp
University of Manchester, Manchester, UK

Three experiments compared forgetting of the duration of a bar-like visual stimulus with forgetting of its length. The main aim of the experiments was to investigate whether subjective shortening (a decrease in the subjective magnitude of a stimulus as its retention interval increased) was observable in length judgements as well as in time judgements, where subjective shortening has been often observed previously. On all trials of the three experiments, humans received two briefly presented coloured bars, separated by a delay ranging from 1 to 10 s, and the bars could differ in length, duration of presentation, or both. In Experiment 1 two groups of subjects made either length or duration judgements, and subjective shortening-type forgetting functions were observed only for duration. Experiments 2 and 3 used the same general procedure, but the stimuli judged could differ both in length and duration within a trial, and different subject groups (Experiment 2) or the same subjects in two conditions (Experiment 3) made either length or duration judgements of stimuli, which were on average physically identical. Subjective shortening was only found with duration, and never with length, supporting the view that subjective shortening may be unique to time judgements.

The last 20 or so years have seen an immense growth of interest in timing processes in humans and animals, including emphasis on the way in which data and theory derived from timing experiments initially carried out with animals might bear on human timing. Wearden and McShane (1988) provide an early example of this work, and more recently Wearden, Edwards, Fakhri, and Percival (1998) have shown how ideas compatible with a currently popular theory of animal timing, the *scalar timing theory* of Gibbon, Church, and Meck (1984) might be used to understand a classic problem in human time psychology, that of why auditory stimuli appear to have greater subjective duration than visual stimuli. A complete review of animal/human comparisons in the domain of timing is inappropriate here, but Allan (1998), Wearden (1994), and Wearden, Denovan, Fakhri, and Haworth (1997) provide many relevant references.

One of the oddest results to emerge from the study of animal timing must surely be the *subjective shortening* found in animals' short-term memory for stimulus duration. Spetch and

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

The experiments reported in the present article were conducted by the second and third authors as part of the requirements for Bachelor's degrees in Psychology awarded by the University of Manchester.

http://www.tandf.co.uk/journals/pp/02724995.html DOI:10.1080/02724990143000108

Wilkie (1983) popularized this result, although evidence for subjective shortening can be found in earlier work (e.g., Church, 1980). Subjective shortening was initially manifested in delayed-matching-to-sample procedures with pigeons, using a method which, in outline, is as follows. On each trial, pigeons were initially presented with one of two stimulus durations (e.g., 2 and 8 s long). After the offset of the stimulus, the pigeons could respond on one of two response keys (e.g., left and right); a response on one of these was correct after the shorter duration, and a response on the other was correct after the longer duration. During training, there was a zero delay between stimulus offset and opportunity to respond, and the discrimination was mastered within a few sessions. Following this training, an above-zero delay was introduced between the stimulus offset and the response opportunity. Spetch and Wilkie (1983) found that introducing this delay produced an unusual pattern of errors; pigeons showed a "choose-short" effect, an increasing tendency to make the response appropriate to the shorter of the two durations presented as the delay increased.

Spetch and Wilkie (1983) accounted for this behaviour by proposing that forgetting of the stimulus duration followed the principle of subjective shortening: The older the memory, the shorter the duration it contained seemed to be. This hypothesis predicts the "choose-short" effect in a simple way: A "short" response and a "long" response are associated with the different stimulus durations experienced with the zero-delay training. When the delay is increased, both the 2- and the 8-s stimuli show subjective shortening: The initially "long" stimulus thus comes to increasingly resemble the memory of the "short" standard, and the "short" standard seems even shorter and consequently even less similar to the "long" standard that it did with zero delay. When the response is delayed after the "long" stimulus, the subjects come to "choose short", whereas short responses persist after the shorter duration, even though it has been subjectively shortened as well.

Further evidence for subjective shortening in pigeons' memory for event duration comes from Spetch (1987), who trained animals to discriminate 2- and 8-s food presentations (which acted as the stimuli in this experiment) with, in one condition, a 20-s delay between stimulus offset and opportunity to respond. Pigeons were then tested with shorter or longer stimulus–response delays, and they exhibited a "choose-long" effect when the delay was reduced between training and testing, whereas they showed a "choose-short" effect when the test delay was longer than that used in training.

More recent work has complicated the simple picture presented earlier. For example, different methods of examining pigeons' memory for event duration do not always produce evidence of a clear "choose-short" effect (Grant & Spetch, 1991). In addition, Santi, Stanford, and Coyle (1998) reported that pigeons failed to show a "choose-short" effect when they remembered the duration of auditory signals, although the effect was present when visual stimuli were used. Another issue that has attracted attention in studies with animals is the effect of the inter-trial interval on duration memory (e.g., Spetch & Rusak, 1992; see also Kelly & Spetch, 2000), and possible interactions between inter-trial interval and the modality of the remembered signal (Santi, Coyle, Coppa, & Ross, 1998). These issues will not be discussed here, as they may not be directly pertinent to humans' memory for stimulus duration, but Grant, Spetch, and Kelly (1997) provide a review.

If subjective shortening very often occurs in short-term memory for duration in pigeons, can similar effects be demonstrated in humans? Wearden and Ferrara (1993) showed that they could (for other demonstrations, see Wearden & Culpin, 1995, 1998). Because humans can label the stimulus durations verbally (e.g., as "short" or "long") and then retain these verbal labels, the method introduced by Spetch and Wilkie (1983) cannot be employed, and Wearden and Ferrara (1993) instead developed a variant of the "roving-standard" method of time psychophysics (Allan, 1979).

On each trial, subjects received two tones (first one called the *sample*, *s*, the second the *comparison*, *c*) separated by a delay, which was usually 1, 2, 5, or 10 s. The sample differed from one trial to another and was drawn randomly from a distribution running from 300–500 ms; the comparison either had the same duration as the sample, or was 100 ms longer or shorter. When the comparison had been presented, subjects had to judge whether it was equal in duration to the standard, longer, or shorter, but no feedback was given. The logic of Wearden and Ferrara's (1993) procedure can be easily grasped by considering the effects of *s–c* delay. As this increases from 1 to 10 s, the "fresh" memory of *c* is compared with an increasingly degraded memory of *s*, and if *s* is exhibiting subjective shortening, then the decision made after *c* should change systematically with *s–c* delay. Although Wearden and Ferrara's experiments were procedurally simple, they generated a complex pattern of results. Given that experiments on memory for duration will be unfamiliar to most readers, and because past experience has shown that the results obtained using the Wearden and Ferrara paradigm can appear complicated and confusing, we will give a fairly detailed account of results from experiments of this type to provide a background for the present work.

Figure 1 shows data from four very similar experiments. One data set comes from Wearden and Ferrara (1993) and shows *s–c* delays up to 16 s; another comes from Wearden and Culpin (1998) and uses *s–c* delays from 1 to 10 s. In this experiment, tone stimuli were used as *s* and *c*, but subjects additionally had to remember a one-syllable word. The third and fourth data sets come from an unpublished study. In one condition, tones were used, and the *s–c* delay varied from 1 to 10 s (essentially a replication of some conditions from Wearden and Ferrara, 1993); in the other condition, the stimuli were visual (squares presented on a computer screen), with the *s–c* delay again varying between 1 and 10 s.

Data are shown separately for the three different trial types of the experiment, where trial type is defined by the *s/c* duration relation or, equivalently, by the correct response. On EQUAL trials $s = c$ (so "equal" is correct), on SHORT trials $s > c$ (so "short", i.e., second shorter than first, is correct), and on LONG trials $s < c$ (so "long" is correct). Inspection of the three panels of Figure 1 immediately suggests that increasing *s–c* delay does not have identical effects on all trial types. The subjective shortening hypothesis is, in fact, consistent with these differences.

Consider first EQUAL trials (top panel of Figure 1). Here, as the *s–c* delay lengthens performance should deteriorate, as *s* becomes progressively subjectively shorter than *c*, and this decline is shown in all the data sets included in Figure 1 (as well as in most others from Wearden & Ferrara, 1993; and Wearden & Culpin, 1995, 1998). On SHORT trials (centre panel of Figure 1), on the other hand, *s* really has a longer duration than *c*, so even if some subjective shortening of *s* occurs, there may be little change in performance accuracy, particularly at short *s–c* delays, where although *s* may have shortened somewhat, it may still be discriminably longer than *c*. Thus performance shows little change over short *s–c* delays on SHORT trials, but may show a decline at longer *s–c* delays where the subjective shortening effect overcomes the fact that $s > c$. The data sets shown in Figure 1, and others in Wearden and Farrara (1993), generally show this effect when auditory stimuli are used. Good

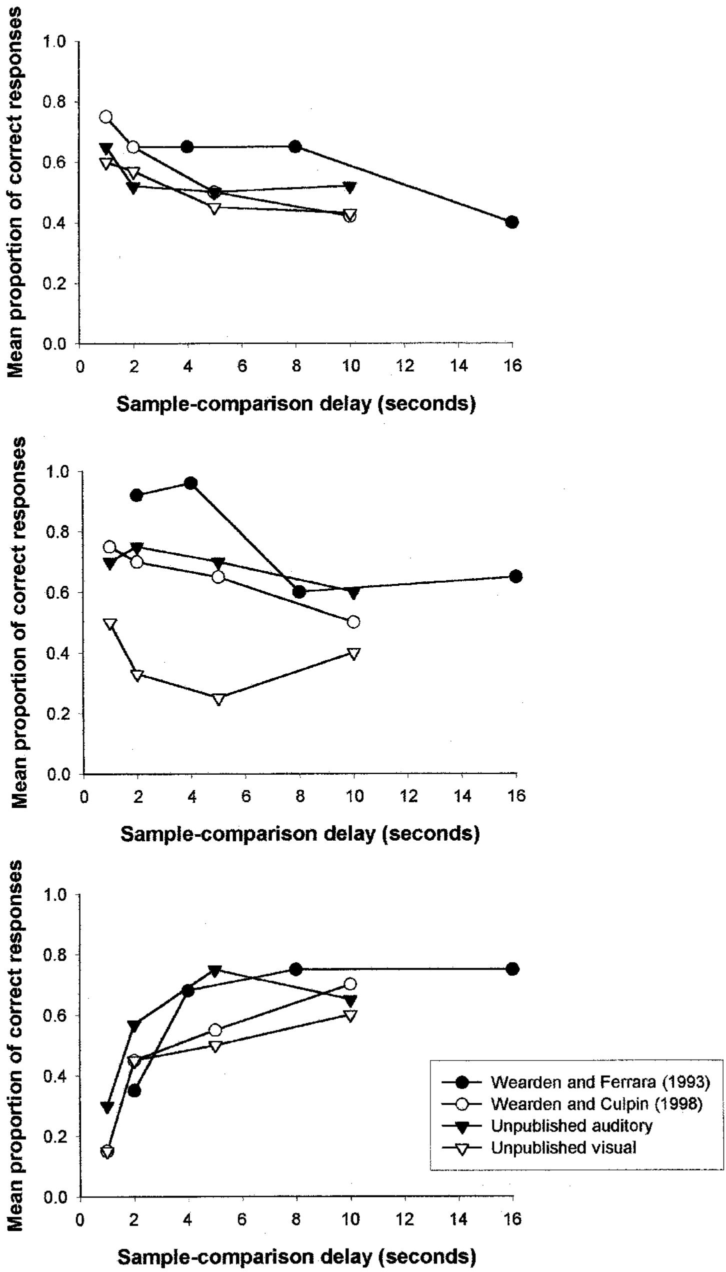

Figure 1. Mean proportion of correct responses from four memory-for-duration studies, plotted against sample–comparison delay. Upper panel shows data from EQUAL trials, centre panel data from SHORT trials, and the lowest panel data from LONG trials. Data sets used are Wearden and Ferrara (1993, Experiment 3), Wearden and Culpin (1998), and two unpublished studies, one using auditory stimuli, the other visual stimuli.

performance on SHORT trials with auditory stimuli may also be aided by the apparent presence of a positive *time–order error* (a tendency to judge that the first stimulus presented lasts longer than the second, regardless of their actual relation) when this stimulus type is used (Wearden & Ferrara, 1993). The stimuli from the visual stimulus condition shown in Figure 1 (and, to anticipate data presented later, the visual stimuli used in the experiments in this article) show much poorer performance on SHORT trials, with the proportion of correct responses rarely exceeding .5, but even for these stimuli, increasing *s–c* delay never produces a monotonic decrease in correct responses, whereas such decreases are usually found on EQUAL trials with visual stimuli.

However, perhaps the most interesting case comes from LONG trials (bottom panel of Figure 1), where $s < c$. In all the data sets shown (as well as in others in the articles cited earlier) performance accuracy actually *improves* with increasing *s–c* delay. At first sight, improvement on a memory task with increasing retention interval (which is effectively what the *s–c* delay is) appears extraordinary, but the subjective shortening hypothesis accounts well for this result. On LONG trials $s < c$, so if *s* appears to shorten progressively with increasing *s–c* delay, it becomes easier for the subject to discriminate the *s/c* difference. Consequently, performance improves as *s–c* delay increases.

Overall, then, the subjective shortening hypothesis deals reasonably well with the pattern of results obtained on the different trial types of the Wearden and Ferrara (1993) procedure. We regard the improving performance with increasing *s–c* delay on LONG trials as being one of the principal "signatures" of subjective shortening in this type of experiment (the other being the pattern of errors committed, particularly on EQUAL trials, as will be discussed later), so in the studies to be reported later, where the question of whether subjective shortening has occurred or not is the main focus of interest, we will be particularly concerned with performance on trials of this type.

The present article uses a method nearly identical to that of Wearden and Ferrara (1993) to address what seems to us one of the most interesting questions about the phenomenon, and this is whether the subjective shortening effect occurs only for duration judgements, or whether it can be observed when non-duration judgements of some types are made. There are three previously published studies of this issue using pigeons and methods similar to that of Spetch and Wilkie (1983). In the first (Honig & Spetch, 1988), pigeons initially received presentations of alternating red and green key colours, which alternated either quickly (1 cycle/s) or more slowly (1 cycle/4 s), then had to make subsequent choice responses based on alternation rate. If the delay between stimulus presentation and response opportunity was lengthened above that used in training, the pigeons showed a "choose-fast" effect, preferentially reporting that the previous rate of alternation had been the faster one. In the second (Fetterman & MacEwen, 1989), pigeons were required to make either a small or larger number of responses on a key, then later were required to choose one of two response alternatives on the basis of their previous behaviour. As the time delay between the initial response period and the choice phase increased, pigeons became increasingly likely to report that they had previously made the fewer number of responses (a "choose-small" effect). In both these instances, however, the effects obtained may be secondary to subjective shortening of duration. In the Honig and Spetch (1988) case, if pigeons remembered the number of alternations correctly, but the duration of presentation of the stimuli was subjectively shortened, the "choose-fast" effect might naturally emerge. Likewise, in the Fetterman and MacEwen study, the number of responses made in the initial phase of the trial was highly correlated with the time this behaviour took, so perhaps their experiment was in fact a disguised temporal discrimination, producing a typical subjective shortening effect. In fact, explanations in terms of subjective shortening of duration are those offered by the two sets of authors themselves.

A final study by Roberts, Macuda, and Brodbeck (1995) avoided the confound between duration and other event dimensions present in the work reviewed earlier. This involved two groups of pigeons performing matching-to-sample tasks with sample stimuli consisting of a number of light flashes. For one group, number of flashes varied (although the duration of the sequence did not) and for another the duration varied (but number did not). Results were complicated by the fact that the pigeons appeared to base their judgements on the number of flashes in both groups, but evidence for a subjective shortening-like "choose-small" effect for number of flashes was found, apparently the only clear evidence for subjective shortening outside the duration domain.

In its simplest form, the assertion that subjective shortening in humans might occur only for duration judgements is impossible to answer positively for obvious logical reasons. If subjective shortening is not observed when the basis of judgement is some stimulus dimension X, another untested dimension Y may show the effect, and so on, so no amount of experimentation could show that *no* stimulus dimension other than duration shows subjective shortening. Of course, a single observation of subjective shortening with some dimension other than duration would disprove the assertion. However, the present studies attempt something more modest; we selected a stimulus dimension (physical length of a bar-like stimulus, described later, and to subjects in the experiments, as a "line") which we argue might show a subjective shortening effect, then compared the results obtained in a Wearden and Ferrara (1993) paradigm using both length and duration judgements. In Experiments 2 and 3, in addition, the length and duration judgements were made on physically identical stimuli. If subjective shortening effects are obtained only with duration judgements, then the conjecture that only duration memory shows this effect is supported; on the other hand, obtaining subjective shortening effects with length would disprove the idea that this phenomenon is unique to time judgements.

Why might physical length (which we will from now on refer to as just *length*, as contrasted with *duration*) be a potential candidate for subjective shortening effects? We conjectured that subjective shortening might be obtained from humans when (1) the stimuli in the experiment are not verbally labelled by subjects, and (2) the stimuli vary along a single dimension. Making the stimulus variable from trial to trial, as well as varying it slightly but noticeably between trials, might ensure condition (1), and physical length of a line seemed to us an appropriate unidimensional stimulus property. The length of a presented line thus seemed to have both the required properties.

As the lines are visual stimuli, an essential prerequisite for the present work was confidence that subjective shortening effects could be obtained from humans when visual stimuli were used, as Wearden and Ferrara (1993) and Wearden and Culpin (1995, 1998) used only auditory events. A pilot study was conducted using squares whose duration had to be compared in a paradigm similar to that of Wearden and Ferrara and, as Figure 1 indicates, the improvement in performance with increasing *s–c* delay on LONG trials, one of our signatures of subjective shortening, was obtained. This result, incidentally, is the first demonstration of subjective shortening in humans with visual stimuli.

EXPERIMENT 1

In Experiment 1 the stimuli used for length and duration judgements were two thin coloured lines (the sample and comparison stimuli, respectively) presented on the computer screen. For the LENGTH group, subjects were required to make a judgement of the relative length of the lines, which either had equal length, or were such that the second one was longer or shorter than the first. Presentation of the lines was separated by a delay which was 1, 2, 5, or 10 s, and from one trial to another the length of the sample line differed randomly. Within each trial, the duration of presentation of the bars was always the same, but presentation duration varied randomly between trials, and subjects were told that duration was irrelevant. For the DURATION group, the basis of the judgement was the relative duration of presentation of the sample and comparison lines, and the sample duration varied randomly from trial to trial, with the comparison having the same presentation duration or being 100 ms longer or shorter. The length of the sample and comparison lines was always the same within the trial, but varied randomly between trials, and subjects were instructed that length was irrelevant.

Method

Subjects

A total of 30 Manchester University psychology undergraduates participated for course credit, which was not contingent on performance. They were arbitrarily allocated to the equal-sized LENGTH and DURATION groups.

Apparatus

The experiment took place in a small cubicle, which provided isolation from external noises. An Opus 16X (IBM-compatible) computer with a colour monitor controlled all experimental events, and the computer keyboard was used to register responses. The experimental program was written in the MEL language (Micro-Experimental Laboratory: Psychology Software Tools, Inc.), which ensured millisecond accuracy for the timing of stimuli and responses.

Procedure

All subjects received a single experimental session. The visual stimuli used were yellow-coloured lines, 5 pixels wide (about 0.35 cm, i.e., there were about 18 pixels/cm), presented on a black background. The different groups were initially given appropriate instructions as to the stimulus dimension that was the basis of the judgement, which was either length or duration of presentation. The general arrangement for a trial was as follows. After the subject pressed in response to a "Press spacebar for next trial" prompt, the sample stimulus was presented after a 1-s delay. Termination of the sample was followed by the s–c delay, which was 1, 2, 5, or 10 s (offset to onset), then by the presentation of the comparison stimulus. When this terminated, the subject was asked "Was the length [duration] of the second line longer than (L), shorter than (S), or equal to (E) the first one? Press the appropriate key". The key press was followed by a return to the "Press spacebar for next trial" prompt, and no performance related feedback was given. For the LENGTH group, the sample stimulus length differed on each trial. On EQUAL trials($s = c$), the sample length was randomly selected from a uniform distribution running from 170–230 pixels (about 9.4–12.8 cm) and then was repeated after the s–c delay as the comparison. On SHORT trials ($s > c$), the sample was randomly selected from the range 190–230 pixels, and the comparison was 20 pixels shorter. On LONG trails ($s < c$) the sample was randomly selected from the range 170–210 pixels, and the comparison was 20 pixels longer. The duration of presentation of *s* and *c* varied between trials and was randomly selected from a uniform distribution running from 400 to 600 ms, but was constant within trials. Events for the DURATION group were in general similar. On EQUAL trials, the sample line was presented for a random duration between 400 and 600 ms, and this random value was repeated for the comparison duration. On SHORT trials the sample length was between 500 and 600 ms, and the comparison was 100 ms shorter. On LONG trials, the sample presentation duration was between 400 and 500 ms, and the comparison was 100 ms longer. The length of the presented line on all trials varied randomly between trials from 170 and 230 pixels, and the sample and the comparison stimulus always had the same length within the trial.

The three trials types (EQUAL, SHORT, and LONG) were combined with the four *s–c* delays to yield a block of 12 different trials. Trials were randomly ordered within this block and were presented successively until all had been delivered (i.e., each of the 12 trials was presented once within the block). Six blocks were given in the experimental session.

Results

Figure 2 shows the mean proportion of correct judgements from both the LENGTH and DURATION groups, with data coming from EQUAL, SHORT, and LONG trials (upper, centre, and lower panels of Figure 2, respectively).

Inspection of the data in Figure 2 suggests that all three of the experimental variables group (LENGTH vs. DURATION), judgement type (EQUAL, SHORT, and LONG), and *s–c* delay (1, 2, 5, and 10 s) played some role in determining judgements. An overall analysis of variance (ANOVA) found significant main effects of group, $F(1, 28) = 9.82$, $p < .01$, type, $F(2, 56) = 5.28$, $p < .01$, and *s–c* delay, $F(3, 84) = 3.42$, $p < .05$. The group by *s–c* delay interaction was also significant, $F(3, 84) = 4.32$, $p < .01$, but the group by trial type interaction was not, $F(2, 56) = 1.11$, nor was the type by *s–c* interaction, $F(6, 168) = 1.65$. However, the three-way group by type by *s–c* delay interaction was significant, $F(6, 168) = 3.08$, $p < .01$.

The results are probably easier to understand when simple ANOVAs are conducted, for example, analyses that follow the form of Figure 2. That is, we consider trial types separately, and analyse effects of group (LENGTH vs. DURATION) and *s–c* delay.

On EQUAL trials, there was no effect of group, $F(1, 28) = 0.7$, but there was a significant effect of *s–c* delay, $F(3, 84) = 4.64$, $p < .01$, with performance deteriorating with increasing delay in both groups. The group by delay interaction was not significant, $F(3, 84) = 0.71$. On SHORT trials, there was an effect of group, $F(1, 28) = 5.35$, $p < .05$, but neither *s–c* delay nor the group by delay interaction was significant, $F(3, 84) = 1.13$, and 0.83, respectively. Evidently, in this case, the length judgement was easier than the duration judgement. On LONG trials, there was a significant effect of group, $F(1, 28) = 7.02$, $p < .05$, but not delay, $F(3, 84) = 0.75$, although the group by delay interaction was highly significant, $F(3, 84) = 8.93$, $p < .001$. It is this last result that is of the greatest interest for the subjective shortening effect. Statistically analysing the simple effect of *s–c* delay in these cases poses some problems with the design we have used, but we will follow the suggestion of Howell (1997, p. 468), and use repeated measures ANOVAs for data from the DURATION and LENGTH groups separately. Such analyses showed that increasing *s–c* delay significantly *increased* performance accuracy in the DURATION group, $F(3, 42) = 5.53$, $p < .01$, but significantly *decreased* it in the LENGTH group, $F(3, 42) = 3.87$, $p < .05$.

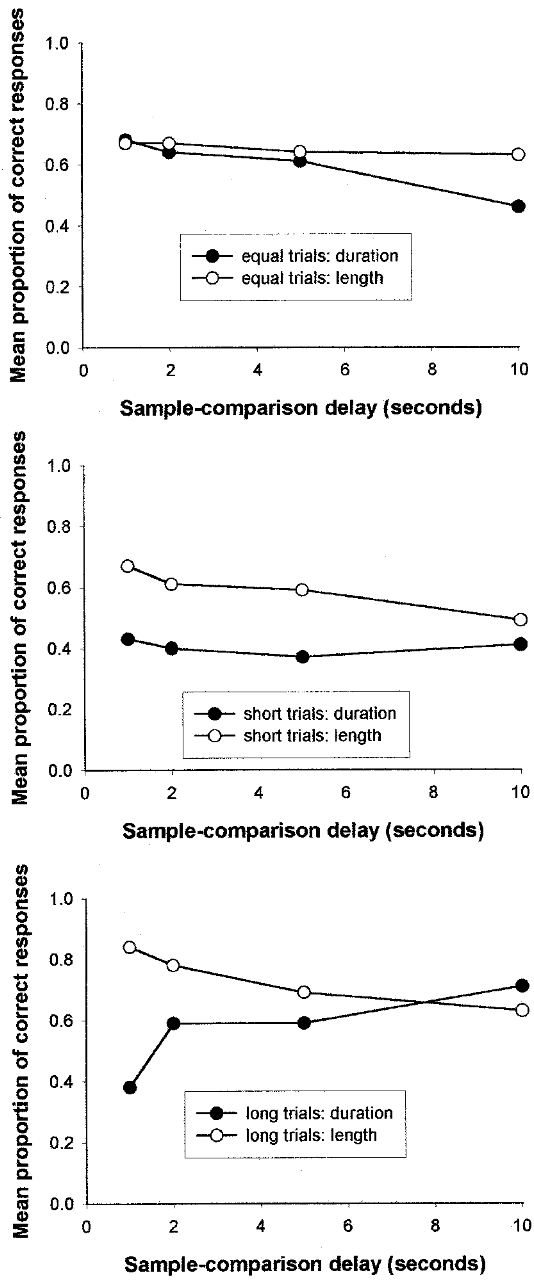

Figure 2. Mean proportion of correct responses plotted against sample–comparison delay from Experiment 1. Upper panel: EQUAL trials; centre panel: SHORT trials; lowest panel: LONG trials. Within each panel, data are shown separately for the DURATION (filled circles) and LENGTH (open circles) groups.

Discussion

The results from Experiment 1 show that one of the characteristic "signatures" of subjective shortening in the Wearden and Ferrara (1993) paradigm—an increase in performance accuracy with increasing *s–c* delay on LONG trials (but not on other trial types)—was found in duration judgements, but not when length was the basis of task performance. This result was obtained in spite of the fact that, according to the subjects' reports, neither the duration nor the length judgements were based on verbal labelling of the stimuli. In addition, both the duration and length judgements involved a unidimensional quantity. This finding supports the contention that subjective shortening effects in humans might be restricted to duration judgements.

However, a potential problem with Experiment 1 is that the stimuli judged in the two groups were not physically identical; in the DURATION group, duration varied within the trial but length was constant, whereas for the LENGTH group, length differed but duration was constant within trials. In addition, the length judgement was overall significantly easier than the duration judgement. Would the apparent difference in forgetting of length and duration judgements remain if the stimuli were physically identical? That is, suppose that within each trial both length and duration of the sample and comparison could vary, and the only difference between the groups was the instructions given. A DURATION group would be instructed to use presentation duration as the criterion for judgements and to ignore length, whereas for the LENGTH group, length should be attended to and duration ignored. Experiment 2 is a study of exactly this type, where different subject groups made either length or duration judgements (using a paradigm similar to that of Wearden & Ferrara, 1993, and nearly identical to that of Experiment 1) of stimuli that were, on average, physically identical.

An additional feature of Experiment 2 was that errors were used to investigate the hypothesis of subjective shortening as well as correct responses. As Wearden and Ferrara (1993) point out, the process of subjective shortening, if it occurs, should not only affect the pattern of correct responses observed, but also the errors made. The simplest case to consider is EQUAL trials, where the sample and comparison stimuli are actually the same. If subjective shortening is occurring, the sample will appear to be increasingly shorter than the comparison as the *s–c* delay increases. It might, therefore, be expected that LONG errors (i.e., making the LONG response on EQUAL trials) should become more common as *s–c* delay increases, and that SHORT errors should decrease. Wearden and Ferrara (1993) found this pattern in data from their Experiment 3. Experiment 2 provides evidence about the type of errors made in the DURATION and LENGTH groups on all trial types.

EXPERIMENT 2

Method

Subjects and apparatus

A total of 31 Manchester University psychology undergraduates participated for course credit. Of these, 14 were arbitrarily allocated to DURATION group and 17 to the LENGTH group. The apparatus was the same as that for Experiment 1.

Procedure

General details of the stimuli and procedure were as those for Experiment 1. Both groups received the same stimuli and differed only according to the dimension of the task that was relevant. For the LENGTH group this was physical length (and subjects were instructed to ignore duration of presentation differences, if any), and for the DURATION group the dimension was duration of presentation (and subjects were told to ignore any length differences that they observed). The *s–c* delays between presentation of the stimuli on the trial were 1, 2, 5, and 10 s, and after the comparison stimulus had been presented subjects were asked the same question as that in Experiment 1. On each trial, both the length and duration of *s* and *c* could be different, but only one of the dimensions (either length or duration) was relevant, depending on group. Thus, the length of the comparison could be equal to, or shorter or longer than, that of the sample, and the duration could also be equal, shorter, or longer. All length and duration differences were used at each *s–c* delay: So, for example, the comparison could be physically longer than the sample (a LONG trial for the LENGTH group), but have shorter duration of presentation (thus making a SHORT trial for the DURATION group). The nine combinations of length and duration, coupled with the four different *s–c* delays, made a basic block of 36 trials, and within each block the trials were randomly ordered. The experimental session consisted of two 36-trial blocks.

When *s–c* had equal lengths, *s* was randomly sampled from a uniform distribution running from 150–250 pixels, and the sampled value was repeated as *c*. When *s* was longer than *c*, *s* was sampled from a distribution running from 170–250 pixels, and *c* was 20 pixels shorter. When *s* was shorter than *c*, *s* was sampled randomly from between 150 and 230 pixels, and *c* was 20 pixels longer. When *s* and *c* had equal durations, *s* was randomly sampled from a uniform distribution running from 250–550 ms, and this value was repeated as *c*. When *s* had a longer duration than *c*, *s* was sampled from 400–550 ms, and *c* was 150 ms shorter. When *s* had a shorter duration than *c*, *s* was sampled from 250–400 ms, and *c* was 150 ms longer.

Results

Figure 3 shows the proportion of correct responses from the LENGTH and DURATION groups, plotted against *s–c* delay, with different panels once again showing data from the different trial types separately.

An overall ANOVA of all three trial types combined found a just-significant main effect of group, $F(1, 29) = 4.55$, $p = .042$, but no main effects either of trial type, $F(2, 58) = 2.13$, or *s–c* delay, $F(3, 87) = 1.4$. However, two-way interactions between group and trial type, $F(2, 58) = 4.08$, $p < .05$, and trial type by delay, $F(6, 174) = 3.22$, $p < .01$, were both significant, and the group by delay interaction approached significance, $F(3, 87) = 2.67$, $p = .052$. The three-way interaction between group, trial type, and delay was also significant, $F(6, 174) = 6.99$, $p < .001$.

Once again, the data may be easier to understand when simpler analyses are performed. For example, consider possible between-group effects from the different trial types separately. On EQUAL trials, there was no effect of group, $F(1, 29) = 1.09$, nor any significant group by delay interaction, $F(3, 87) = 1.92$, but there was a significant effect of delay, $F(3, 87) = 3.55$, $p < .05$. On SHORT trials, there was an effect of group, $F(1, 29) = 13.30$, $p < .01$, which obviously favoured higher performance accuracy in the LENGTH group, and there were also significant effects of delay, $F(3, 87) = 2.97$, $p < .05$, and group by delay interaction, $F(3, 87) = 13.3$, $p < .001$. Analysis of the effect of *s–c* delay on data from the groups separately showed that the proportion of correct responses decreased with *s–c* delay in the DURATION group, $F(3, 39) = 7.14$, $p < .01$, whereas the apparent increase in performance accuracy in the LENGTH group just failed to reach significance, $F(3, 48) = 2.62$, $p = .06$.

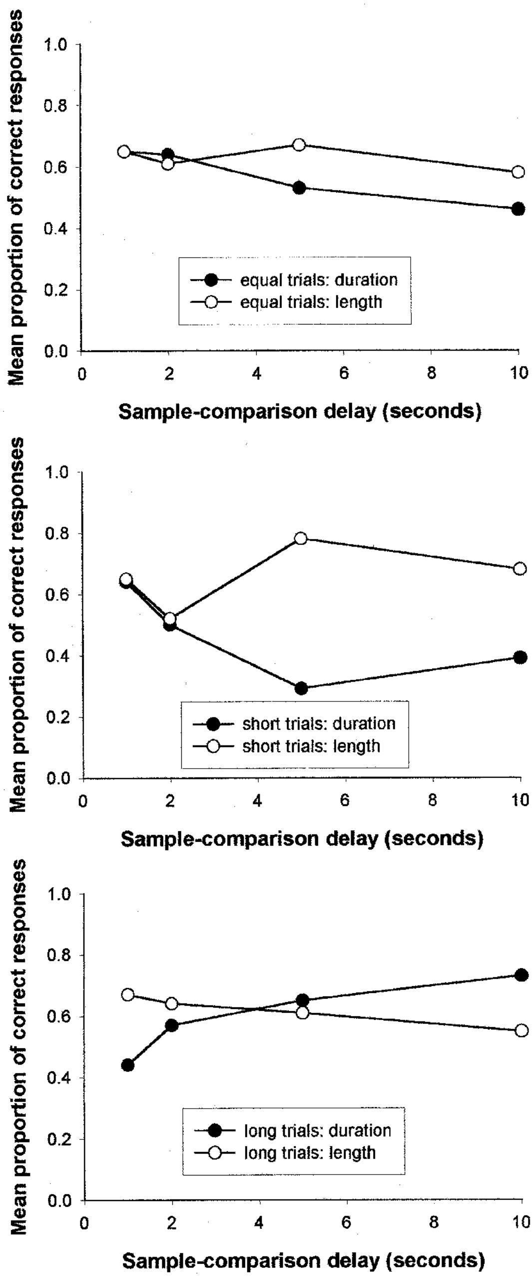

Figure 3. Mean proportion of correct responses plotted against sample–comparison delay from Experiment 2. Other details as Figure 2.

On LONG trials, there was no effect of group, $F(1, 29) = 0.09$, or delay, $F(3, 87) = 1.35$, but there was a significant group by delay interaction, $F(3, 87) = 6.35$, $p < .001$. Comparison of the effects of delay on LONG trials for the two groups found no significant effect of delay in the LENGTH group, $F(3, 48) = 2.62$, $p = .062$, although the decline in performance with increasing delay approached significance, whereas in the DURATION group, performance accuracy *increased* significantly, $F(3, 39) = 4.86$, $p < .01$, with increasing *s–c* delay. As in Experiment 1, this signature of subjective shortening was observed in the DURATION group but not the LENGTH group.

Figure 4 shows the mean number of errors made by subjects in the DURATION and LENGTH groups, on the different trial types (EQUAL, SHORT, and LONG; top, middle, and bottom panels, respectively).

Inspection of the upper panel of Figure 4 suggests that the pattern of errors committed on EQUAL trials was different for the DURATION and LENGTH groups. Consistent with the idea of subjective shortening, the mean number of LONG errors (i.e., making the response "long" on EQUAL trials) increased with increasing *s–c* delay in the DURATION group, whereas the mean number of SHORT appeared to decrease slightly with increasing *s–c* delay. The pattern appeared different in the LENGTH group with SHORT errors increasing and LONG errors remaining roughly constant with *s–c* delay.

An overall ANOVA of errors on EQUAL trials included error type (SHORT or LONG errors), group (DURATION or LENGTH), and *s–c* delay as factors. There were significant effects of error type, $F(1, 29) = 9.57$, $p < .01$, *s–c* delay, $F(3, 87) = 5.74$, $p < .01$, and a significant effect of group, $F(1, 29) = 9.85$, $p < .01$. None of the two-way interactions (error type by group, delay by group, and error type by delay) was significant, but the three-way interaction between these variables was significant, $F(3, 87) = 3.07$, $p < .05$.

Simpler analyses may give a clearer picture of the results obtained. For the DURATION group, there was an effect of error type, $F(1, 13) = 9.33$, $p < .01$, showing that LONG errors were more frequent than SHORT ones, and of *s–c* delay, $F(3, 39) = 4.23$, $p < .05$. In addition, the error type by delay interaction approached significance, $F(3, 39) = 2.34$, $p = .09$. The mean number of LONG errors increased significantly with increasing *s–c* delay, $F(3, 39) = 4.78$, $p < .01$, whereas the mean number of SHORT errors did not change significantly. For the LENGTH group, the only significant effect was a just-significant effect of *s–c* delay, $F(3, 48) = 3.00$, $p = .04$.

Overall, therefore, the increase in the mean number of LONG errors with increasing *s–c* delay predicted by the subjective shortening hypothesis was obtained in the DURATION group but not in the LENGTH group, where *s–c* delay had no significant effect on any error type. SHORT errors in the DURATION group decreased with increasing *s–c* delay, as the subjective shortening hypothesis predicted, although the decrease was not statistically significant. However, SHORT errors were less common than LONG errors (again, a result consistent with the subjective shortening hypothesis), and some individual subjects only made one or two errors of this type in the experiment as a whole, which may have contributed to the lack of statistical significance.

Errors committed on EQUAL trials are far easier to interpret than those committed on the other two types of trials, for three reasons. One of these is that on SHORT and LONG trials, one type of error seems much more likely than another. For example, on SHORT trials, the comparison is actually shorter than the standard, so if stimulus representations contain

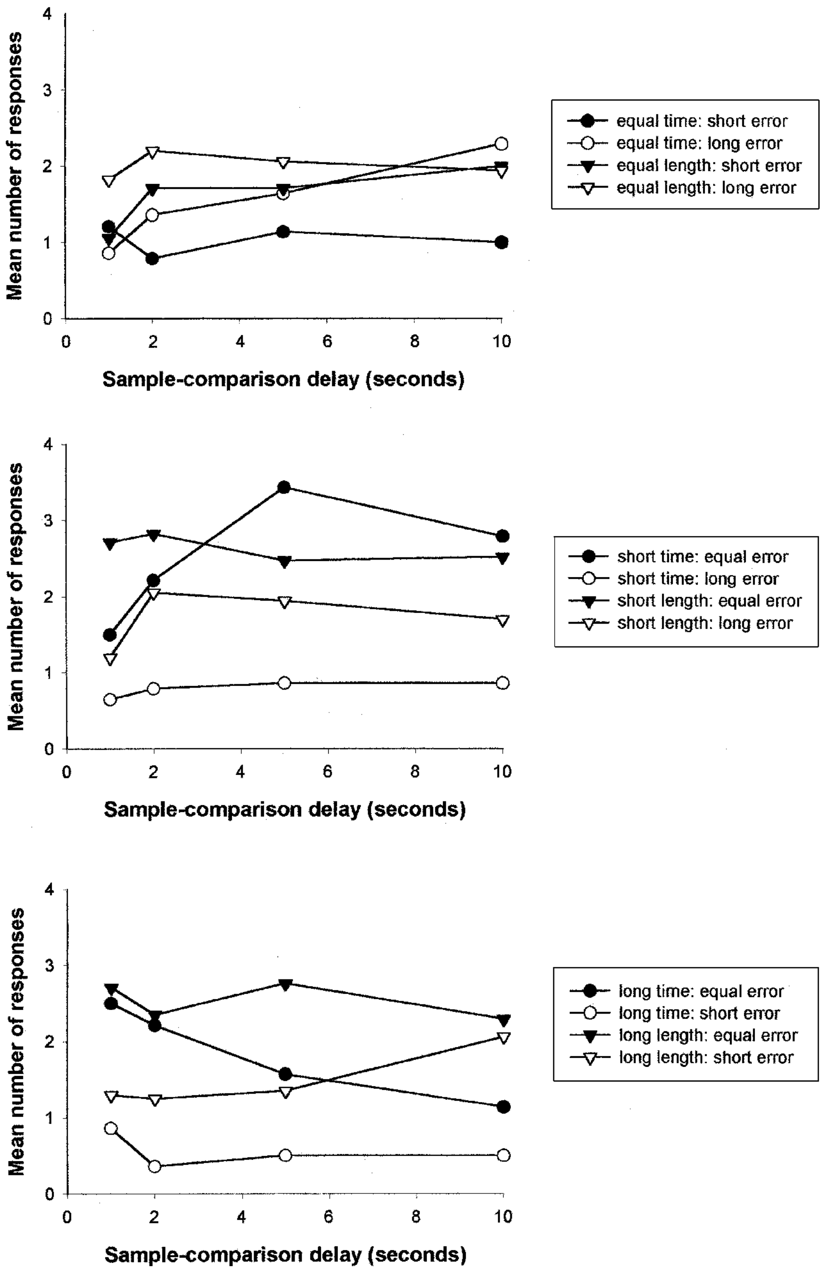

Figure 4. Mean number of errors from the DURATION and LENGTH groups of Experiment 2 plotted against sample–comparison delay. Upper panel shows data from EQUAL trials, centre panel data from SHORT trials, bottom panel data from LONG trials. Within each panel different judgement types (time or length) and the different error types are distinguished in the associated key.

variance, as a result of variability of perceptual encoding or storage in memory, the resulting confusion seems more likely to result in judging that the durations are equal (committing an "EQUAL error") than judging that the second is longer than the first (a "LONG error"). Conversely, for LONG trials, an EQUAL error seems more likely than a SHORT error. In these two cases, making EQUAL errors may just be a consequence of the actual physical values of the stimuli presented (i.e., both SHORT and LONG trials, on average, involve stimuli that are physically closer to those presented on EQUAL trials than those on LONG trials are to those on SHORT trials, and vice versa), and does not necessarily imply any "bias" towards making EQUAL responses.

A second problem, which follows on from the first, is that some subjects may make very few, or even no, SHORT errors on LONG trials, and vice versa, so statistically confirming statements about these types of error can be difficult. Finally, the error patterns on SHORT and LONG trials need to be looked at alongside the patterns of correct responses, as some types of error may change with *s–c* delay, not because another type of error is changing in the opposite direction, but because the pattern of correct responses is changing.

In spite of these caveats, examination of error patterns on SHORT and LONG trials may help to clarify whether processes like subjective shortening are occurring with changes in *s–c* delay, and whether such changes are found only for duration judgements.

The centre panel of Figure 4 shows errors from SHORT trials. Inspection of the data suggests that for both duration and length judgements, EQUAL errors were more common than LONG errors (as we conjectured earlier that they would be) but that some other differences between duration and length judgements can be observed. On SHORT trials, the comparison stimulus is actually shorter than the sample, so subjective shortening of the sample with increasing *s–c* delay should make the task more difficult, probably in the direction of increasing numbers of EQUAL errors. An increase in EQUAL errors can be observed for the duration judgements, whereas for length judgements, the number of EQUAL errors seems likely to slightly decrease.

An overall ANOVA of errors on SHORT trials included error type (EQUAL or LONG errors), group (DURATION or LENGTH), and *s–c* delay as factors. There was an overall effect of group, $F(1, 19) = 18.72, p < .001$, error type, $F(1, 29) = 31.06, p < .001$, and *s–c* delay, $F(3, 87) = 4.60, p < .01$. There were also significant interactions between error type and group, $F(1, 29) = 4.65, p = .04$, and between group and *s–c* delay, $F(3, 87) = 7.47, p < .001$, but no other significant effects. Some of these results confirm the suggestions made earlier: length and duration judgements generated different numbers of errors overall on SHORT trials, EQUAL errors were more common than LONG errors for both duration and length judgements, and *s–c* delay played a significant role in at least one comparison.

Despite the absence of a significant three-way interaction, further analyses were conducted, as interpretation is easier when the DURATION and LENGTH groups are considered separately. Consider first the DURATION group. ANOVA found the following: significant effects of error type, $F(1, 13) = 80.58, p < .001$, showing that EQUAL errors were more common than LONG errors; *s–c* delay, $F(3, 39) = 7.14, p < .01$; and a just significant error type by delay interaction, $F(3, 39) = 2.91, p < .05$. This latter result suggested that the effect of *s–c* delay differed for the two different types of error, and individual ANOVA for each error type confirmed this. EQUAL errors increased with increasing *s–c* delay, $F(3, 39) = 6.71$, $p < .01$, whereas LONG errors showed no change, $F(3, 39) = 0.21$.

Consider next the LENGTH group. Here, there was a significant effect of error type, $F(1, 16) = 6.37, p < .05$, confirming that EQUAL errors were more common that LONG errors, but no other significant effects.

Overall, therefore, the pattern of errors on SHORT trials is consistent with that predicted by the subjective shortening hypothesis for the DURATION group, but the LENGTH group showed no change in error patterns with increasing *s–c* delay.

The bottom panel of Figure 4 shows errors from LONG trials. Inspection suggests that EQUAL errors are more common than SHORT errors, as predicted earlier, and suggests some effects of *s–c* delay. On LONG trials, the comparison is actually longer than the sample, so increasing subjective shortening with increasing *s–c* delay would predict that errors in general would decrease—in particular, EQUAL errors would decline with increasing *s–c* delay.

An overall ANOVA found significant effects of group, $F(1, 29) = 18.11, p < .001$, error type, $F(1, 29) = 37.54, p < .001$, *s–c* delay, $F(3, 87) = 2.74, p < .05$, and significant interactions between delay and group, $F(3, 87) = 6.23, p < .01$, and between error type and delay, $F(3, 87) = 3.93, p < .05$.

Again, despite the absence of a significant three-way interaction, ANOVA of data from the two groups separately clarify the effects. In the DURATION group, there was a significant effect of error type, $F(1, 13) = 27.27, p < .001$, showing that EQUAL errors were more common than SHORT errors, and a significant effect of *s–c* delay, $F(1, 13) = 11.25, p <.01$, whereas the error type by delay interaction just failed to reach significance, $F(1, 13) = 3.60, p = .08$. Both EQUAL errors, $F(3, 39) = 3.46, p < .05$, and SHORT errors, $F(3, 39) = 3.63, p = .02$, decreased with increasing *s–c* delay. In the LENGTH group, on the other hand, only the effect of error type was significant, $F(1, 16) = 37.07, p < .01$.

Overall, therefore, errors committed on LONG trials tend to support the notion of subjective shortening in the DURATION group. With increasing *s–c* delay, not only does the proportion of correct responses increase (Figure 3), but both types of error decline significantly, with the more common EQUAL errors showing a marked effect. In the LENGTH group, on the other hand, increasing *s–c* delay had no significant effect, although EQUAL errors were more common than SHORT ones.

Discussion

Experiment 2, in which the stimuli judged in the two groups were on average physically identical, confirmed the essential results of Experiment 1. In particular, the subjective shortening signature of improving performance with increases in *s–c* delay on LONG trials was found only when subjects were judging duration and not when they were judging length. This occurred in spite of the fact that there was no overall effect of group on performance accuracy on LONG trials, indicating that over all *s–c* delays, the length judgement and the duration judgement were not significantly different in difficulty, just different in the way that they were affected by the *s–c* delay. Analysis of errors committed also confirmed that duration and length judgements showed different patterns of change with increasing *s–c* delay.

The confirmation of subjective shortening in the DURATION group was obtained even though, a priori, the method used seemed to have some potential pitfalls. For example, although over all the trials the length and duration of the presented stimuli were completely independent for both groups, trials could occur where, for example, the comparison had a shorter duration than the sample but had greater length. If subjects tend to judge that stimuli that were physically longer also lasted a longer time, judgement in these conditions could have been particularly difficult. Conversely, any tendency to judge that stimuli that have shorter presentation durations also have shorter length would have affected judgements in the LENGTH group. Overall, however, the subjective shortening effect in the DURATION group was evidently strong enough to overcome these problems, if they arose.

Although Experiment 2 replicated and extended the main findings of Experiment 1, one problem remains. In both Experiments 1 and 2, overall performance accuracy was higher in the LENGTH group than in the DURATION group, although the difference only just reached significance in Experiment 2 and, as mentioned earlier, was not shown on LONG trials in Experiment 2. Perhaps part of the apparent difference between the way lengths and durations are remembered comes from this differential difficulty. Ideally, the length and duration judgements would be made in conditions of identical difficulty, but arranging exactly equal task difficulty poses some problems in the paradigm we have used. For example, do the length and duration tasks have to be equally difficult overall, or equally difficult on all three trial types (SHORT, LONG, and EQUAL trials)? It may be nearly impossible to find a stimulus set that meets these requirements but that also allows performance to change with *s–c* delay. To address the task difficulty issue, Experiment 3 replicated Experiment 2 with two variations. First, the stimuli were changed so that the duration judgement produced higher overall accuracy than the length judgement. Second, to reduce between-subject variance, each subject performed both the length and the duration tasks.

EXPERIMENT 3

Method

Subjects and apparatus

A total of 21 Manchester University psychology undergraduates participated for course credit. The apparatus was the same as that for Experiment 1.

Procedure

Each subject served in a LENGTH condition and a DURATION condition, carried out on two daily sessions, with 10 subjects receiving the LENGTH condition first and 11 the DURATION condition first. All procedural details for the different conditions were identical to those used for the LENGTH and DURATION groups of Experiment 2, except that the length judgement was made more difficult by reducing the difference between *s* and *c* on those trials where they were different. When *s* and *c* had equal length, *s* was randomly sampled between 150–250 pixels and repeated as *c*. When *s* was longer than *c*, *s* was sampled from 160–250 pixels, and *c* was 10 pixels shorter. When *s* was shorter than *c*, *s* was sampled from 150–240 pixels, and *c* was 10 pixels longer.

Results and discussion

Figure 5 shows the proportion of correct responses from the different conditions (LENGTH and DURATION) plotted against *s–c* delay, for the three trial types separately.

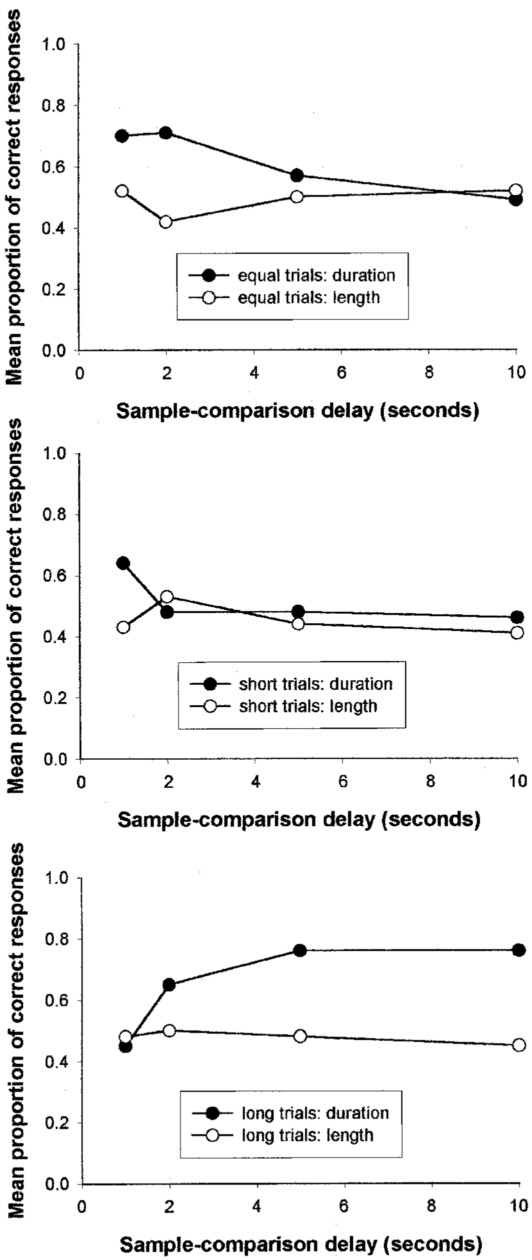

Figure 5. Mean proportion of correct responses for the DURATION and LENGTH conditions of Experiment 3. Other details as Figure 2.

Inspection of the results in the different panels immediately suggests that changing the stimulus differences within trials between Experiments 2 and 3 had the intended effect as, in most cases, more correct judgements occurred for the DURATION condition than for LENGTH. Consistent with this, an overall ANOVA found a significant effect of condition, $F(1, 20) = 19.03$, $p < .001$. There was also a significant effect of trial type, $F(2, 40) = 6.34$, $p < .01$, but no main effect of *s–c* delay, $F(3, 60) = 0.73$. The interaction between condition and trial type was significant, $F(2, 40) = 4.03$, $p < .05$, but that between condition and delay was not, $F(3, 60) = 0.12$. The remaining two-way interaction, between trial type and delay, was significant, $F(6, 120) = 3.99$, $p < .01$, as was the three-way interaction between condition, trial type, and delay, $F(6, 120) = 6.42$, $p < .001$.

Next, we analysed the different trial types separately. On EQUAL trials, there was a significant effect of condition (LENGTH or DURATION), $F(1, 20) = 14.86$, $p < .01$, but no effect of delay, $F(3, 60) = 1.94$, although the condition by delay interaction was significant, $F(3, 60) = 4.48$, $p < .01$. On SHORT trials, neither the effect of condition, $F(1, 20) = 3.02$, nor the effect of delay, $F(3, 60) = 2.57$, was significant, although the latter effect approached significance ($p = .06$). The condition by delay interaction was, however, significant, $F(3, 60) = 3.44$, $p < .05$. On LONG trials, there was a significant effect of condition, $F(1, 20) = 19.40$, $p < .001$, and delay, $F(3, 60) = 4.16$, $p < .05$, and, most importantly from the point of view of demonstrating subjective shortening, a significant condition by delay interaction, $F(3, 60) = 7.22$, $p < .001$. An analysis of data from LONG trials for the two conditions showed a significant increase in the proportion of correct responses with increasing *s–c* delay in the DURATION condition, $F(3, 60) = 10.75$, $p < .001$, but no effect of delay in the LENGTH condition, $F(3, 60) = 0.24$.

Figure 6 shows errors made in Experiment 3 for both the DURATION and LENGTH conditions. Data from EQUAL, SHORT, and LONG trials are shown in the upper, centre, and bottom panels of Figure 6, respectively.

Consider first errors on EQUAL trials. Inspection of the data suggests immediately that the error pattern was different for the two types of judgement. In the DURATION condition, SHORT errors decreased with increasing *s–c* delay, whereas LONG errors increased. In the LENGTH condition, on the other hand, neither error type appeared to change markedly with changing *s–c* delay. An overall ANOVA found significant main effects of condition, $F(1, 20) = 17.77$, $p < .001$, error type, $F(1, 20) = 44.30$, $p < .001$, and *s–c* delay, $F(3, 60) = 6.18$, $p < .01$, as well as significant interaction between error type and delay, $F(3, 60) = 6.48$, $p < .01$, and a three-way interaction between condition, error type, and delay, $F(3, 60) = 4.08$, $p < .05$. The last two effects indicate that the mean number of different error types was differentially affected by *s–c* delay and that there was a difference between the DURATION and LENGTH conditions. These effects are in accord with the subjective shortening hypothesis (i.e., LONG errors increase with delay and SHORT errors decrease, but only in the DURATION condition).

The results are easier to understand with simpler analyses involving the conditions separately. In the DURATION condition, there were: a significant effect of error type, $F(1, 20) = 9.14$, $p < .01$, showing that LONG errors were more common than SHORT errors; a significant effect of *s–c* delay, $F(3, 60) = 5.01$, $p < .01$; and a significant interaction between error type and delay, $F(3, 60) = 8.94$, $p < .01$, suggesting that the mean number of LONG and SHORT errors changed in different ways with changing *s–c* delay. This was confirmed by separate

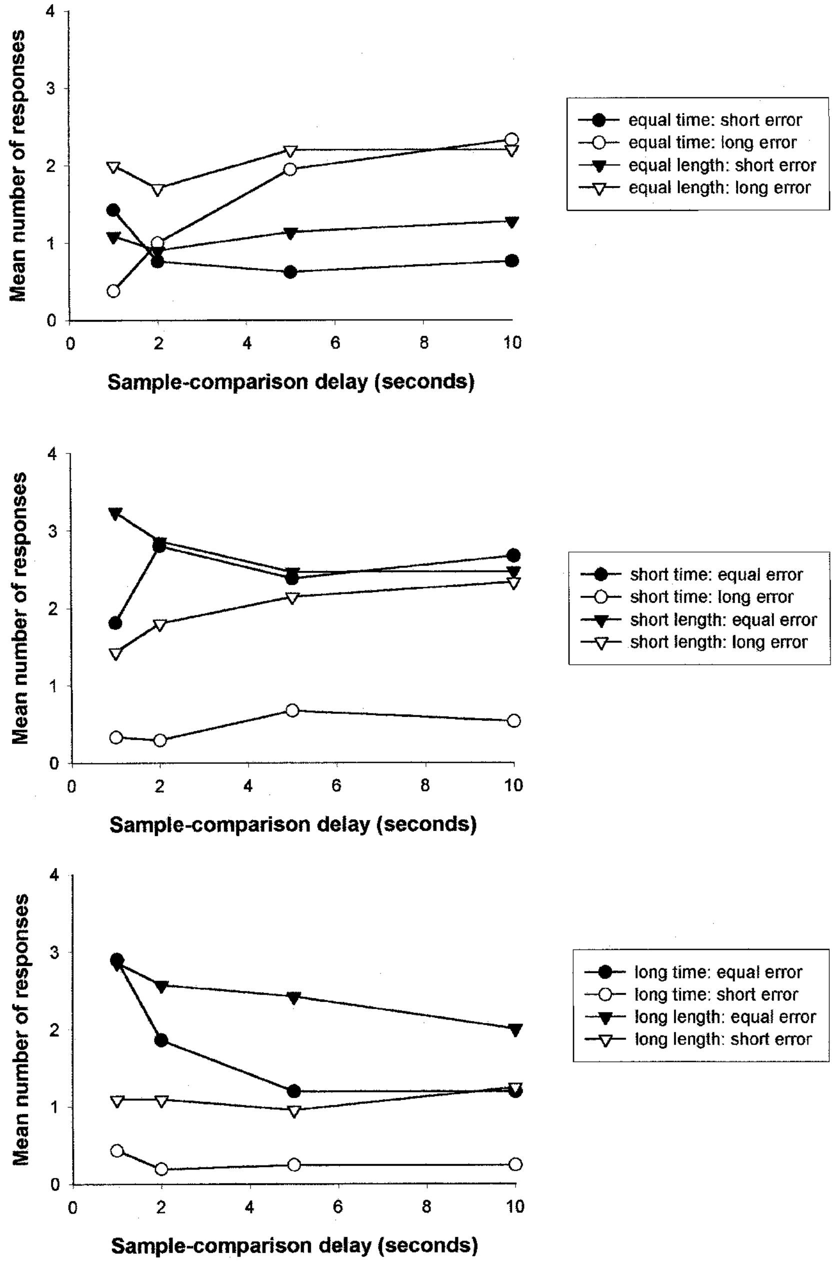

Figure 6. Mean number of errors from the DURATION and LENGTH conditions of Experiment 3 plotted against sample–comparison delay. Upper panel shows data from EQUAL trials, centre panel data from SHORT trials, bottom panel data from LONG trials. Within each panel different judgement types (time or length) and the different error types are distinguished in the associated key.

analyses of the different errors: The mean number of LONG errors increased significantly with increasing *s–c* delay, $F(3, 60) = 11.74$, $p < .001$, whereas the decrease in the mean number of SHORT errors just failed to reach significance, $F(3, 60) = 3.45$, $p = .07$. This is probably because these errors were uncommon in some subjects (e.g., 8 of the 21 subjects had two or fewer of these errors in the whole experiment).

In the LENGTH condition, the only significant effect was of error type, with LONG errors being more common than SHORT errors, $F(1, 30) = 26.30$, $p < .001$. No other effects (e.g., interactions, or effects of *s–c* delay on SHORT and LONG errors separately) approached significance.

Consider next error data from SHORT trials (centre panel of Figure 6). An overall ANOVA revealed significant main effects of condition, that is, DURATION versus LENGTH, $F(1, 20) = 63.62$, $p < .001$, and of error type, $F(1, 20) = 79.43$, $p < .001$. The other significant effects were interactions between trial condition and error type, $F(1, 20) = 15.93$, $p < .01$, and the three-way interaction between condition, error type, and *s–c* delay, $F(3, 60) = 3.18$, $p < .05$. This last effect indicates different effects of delay on the different sorts of errors committed in the DURATION and LENGTH conditions; in other words, error patterns differed for the two different types of judgements.

Once again, interpretation is easier with separate analyses of data from the DURATION and LENGTH conditions. Considering first DURATION judgements, ANOVA revealed significant effects of error type, that is, EQUAL errors were more common than LONG ones, $F(1, 20) = 162.05$, $p < .001$, and of *s–c* delay, $F(3, 60) = 2.49$, $p = .03$, although neither EQUAL errors nor LONG errors, considered individually, changed significantly with increases in *s–c* delay. When LENGTH judgements were made, in contrast, only error type produced a significant result, $F(1, 20) = 10.21$, $p < .01$; that is, EQUAL errors were, not surprisingly, more common than LONG errors.

The bottom panel of Figure 6 shows error data from LONG trials. Overall ANOVA found significant effects of judgement type, $F(1, 20) = 38.96$, $p < .001$, error type, $F(1, 20) = 62.86$, $p < .001$, and *s–c* delay, $F(3, 60) = 17.79$, $p < .001$. There was also a significant interaction between error type and delay, $F(3, 60) = 4.62$, $p < .01$. Although the three-way interaction was not significant, ANOVA of the DURATION judgements separately found significant effects of error type, $F(1, 20) = 96.10$, $p < .001$, *s–c* delay, $F(3, 60) = 10.99$, $p < .001$, and error type by delay interaction, $F(3, 60) = 4.57$, $p < .01$. The first of these effects shows that EQUAL errors were more common than SHORT errors, and the last shows that the different error types were differentially affected by delay. This was confirmed by separate analyses of EQUAL and SHORT errors: The former decreased significantly with *s–c* delay, $F(3, 60) = 13.76$, $p < .001$, whereas the latter showed no change. This pattern of results, taken together with data in Figure 5, is consistent with the pattern expected from subjective shortening: The number of EQUAL errors decreases with *s–c* delay, and these errors are replaced by correct responses.

When LENGTH judgements are considered, on the other hand, only error type was significant, $F(1, 20) = 78.72$, $p < .001$, so overall the assertion that the pattern of effect of *s–c* delay on LONG trials differ for LENGTH and DURATION judgements was supported.

GENERAL DISCUSSION

Data from Experiments 1 to 3, taken together, show that subjective shortening effects in stimulus memory occurred for duration judgements, but not for judgements of the relative length of presented stimuli. This conclusion was supported from data from several different conditions. In Experiment 1, the length judgement was much easier than the duration judgement, although stimuli were not physically identical. In Experiments 2 and 3, on the other hand, the stimuli used for the duration and length judgements were on average physically identical, although the relative difficulty of the length and time judgements varied between Experiments 2 and 3. For the former, the length judgement was slightly easier, for the latter, the duration judgement was easier.

If we consider the critical LONG trials, increasing *s–c* delay significantly improved performance accuracy when duration was the basis of judgement, in all three experiments, but had different effects when length was being judged. This conclusion held over LONG trials in which the length judgement was easier (Experiment 1), where there was no significant difference between the two (Experiment 2), or where the duration judgement was easier (Experiment 3). It was also independent of whether *s–c* delay affected performance on LONG trials when length was the relevant stimulus dimension. In Experiment 1, performance accuracy on LONG trials decreased significantly with increasing delay when length was the relevant stimulus dimension, whereas in Experiments 2 and 3 no significant effect of delay was obtained. The pattern of errors obtained in Experiments 2 and 3 on the different trial types was also consistent with the subjective shortening hypothesis for DURATION judgements, but not for length judgements. For example, on EQUAL trials, the numbers of LONG errors significantly increased with increasing *s–c* delay (and SHORT errors usually decreased, albeit not significantly) with DURATION judgements, but not with judgements of LENGTH. On LONG trials, when DURATION judgements were made, EQUAL errors decreased significantly and were apparently replaced with correct responses rather than SHORT errors, but only when DURATION judgements were made. On SHORT trials, EQUAL errors increased with increasing *s–c* delay (significantly in Experiment 2) with DURATION judgements, but not with judgements of LENGTH.

The pattern of errors, although rather complex and sometimes not easy to grasp immediately, was generally highly consistent with the subjective shortening hypothesis for DURATION judgements, but never for LENGTH ones, supporting the view derived from analysis of correct responses in all three experiments that subjective shortening occurred for DURATION judgements but not for judgements of LENGTH.

The error patterns also argued against some other interpretations of the data. For example, one suggestion is that the subjective shortening-type effects found in analyses of correct responses and EQUAL trial errors for DURATION judgements arise not because of any change in the representations of the sample duration with increasing *s–c* delay (as the subjective shortening hypothesis argues) but because of changing patterns of “bias”. For example, the subjects may just have an increasing tendency to respond LONG as *s–c* delay increases, which would improve accuracy on LONG trials and produce the increasing number of LONG errors with increasing *s–c* delay on EQUAL trials. There are two problems with this interpretation. First, suggesting a “bias” that changes with the value of an independent variable (*s–c* delay in this case) begs the obvious question of why bias changes in this way. In most cases, suggested "biases" are independent of other experimental variables (e.g., see Baum, 1974, for an influential example of the use of the idea of "bias"), so linking them to variables like *s–c* delay makes them difficult to distinguish logically from effects like subjective shortening and gives this "bias" interpretation an ad hoc character. Second, the error data from SHORT trials show no significant tendency for LONG errors to increase with *s–c* delay when DURATION judgements are made, and such an increase would be predicted if the "bias" interpretation is accepted as a potential interpretation of the results. In fact, the subjective shortening hypothesis gives a simpler account of the error data than do arbitrary changes in "bias" with *s–c* delay.

Overall, therefore, it appears as if length and duration are remembered differently, even when the observations on the chosen dimension are extracted from stimuli that are physically the same on average, and even when the overall relative task difficulty of the duration and length judgements is manipulated.

Our findings might lead the study of subjective shortening down two different paths. One of these would be to search for some non-duration-related stimulus dimension that does show subjective shortening. For example, areas of figures, pitches of tones, intensities of visual stimuli, and so on, could all provide unidimensional stimuli that might not be verbally labelled, and that might be used in procedures similar to those of Wearden and Ferrara (1993). Numerosity judgements might also be used (following the results of Roberts et al., 1995, with pigeons) although preventing verbal labelling might be difficult. Although we have no desire to restrict research, our view is that this approach is probably wasteful of experimental effort. A second line of attack seems potentially much more interesting, and this is to try to understand why subjective shortening could occur for *any* stimulus dimension and to ask the more general question of how stimuli are encoded in memory-for-duration tasks.

One starting point is the observation that subjective shortening is compatible with what Wearden (1994) called *quantitative encoding* of time—that is, the idea that *longer* durations are represented by *more* of something. For example, if stimulus duration is encoded in terms of the number of "ticks" arising from the pacemaker of the type of internal clock proposed by Gibbon et al. (1984), then longer durations will be represented by more ticks. Subjective shortening is thus consistent with a memory process resembling physical erasure; as time passes, ticks are erased or lost from the duration representation. These ideas led us to speculate exactly which memory processes are invoked when durations are stored for a few seconds in a task similar to that of Wearden and Ferrara (1993). Wearden and Culpin (1995, 1998) were concerned with memory of the type of auditory stimuli used in Wearden and Ferrara's original task, and they sought evidence that these tones were stored in the phonological loop of the working-memory stems proposed by Baddeley (1987, 1996). Thus, for example, the sample duration would be encoded and phonologically rehearsed during the *s–c* delay. Wearden and Culpin (1995, 1998) found evidence that manipulations designed to interfere with phonological rehearsal (such as articulatory suppression or pre-loading of the phonological loop with competing material) did in fact decrease performance accuracy on a Wearden and Ferrara task. However, the possibility that these effects were produced by non-specific interference and would have resulted whatever competing task was used concurrently with duration memory was not explored.

Wearden and Culpin's (1995, 1998) position is also less consistent with performance when visual stimuli are used in memory-for-duration tasks. If the subjective shortening observed by

Wearden and Ferrara (1993) results in some way from the phonological loop, why then does it also occur when visual stimuli are used, as these will obviously not undergo phonological encoding? One possibility is that stimuli that cannot be phonologically encoded are instead stored in another subsystem of the working memory model, the visuo-spatial sketchpad.

In the experiments presented here, a notable feature of performance is that memory for the length of lines was in most cases hardly affected at all by increases in *s–c* delay. Why is this? One possibility is that line length enters the visuo-spatial sketchpad and is maintained there by a non-verbal rehearsal process, so in fact as only a single item (a single line length) is encoded in the sketchpad, it is well maintained even over long delays (and in unpublished work we have found that even *s–c* delays of up to 16 s produce little or no forgetting of a single line length). Although the lack of effect on *s–c* delay on line length judgements may be intuitively surprising, it may in fact be just like giving the subject the word "cat" to remember during the *s–c* delay. The item would enter the phonological loop and be rehearsed there, and no-one would be surprised if it showed no forgetting over *s–c* delays of up to 10 s.

In conclusion, the work reported in the present article supports the view that short-term duration representations possessed by humans are unusual, and may even exhibit a unique form of forgetting, subjective shortening, which is very different from that exhibited by representations of other types of stimulus, even if the reasons for the apparent peculiarities of duration representations remain at present mysterious.

REFERENCES

Allan, L.G. (1979). The perception of time. *Perception and Psychophysics, 26,* 340–354.

Allan, L.G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes, 44,* 110–117.

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

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A,* 5–28.

Baum, W.M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behaviour, 22,* 231–242.

Church, R.M. (1980). Short-term memory for time intervals. *Learning and Motivation, 11,* 208–219.

Fetterman, J.G., & MacEwen, D. (1989). Short-term memory for responses: The "choose small" effect. *Journal of the Experimental Analysis of Behaviour, 52,* 311–324.

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

Grant, D.S., & Spetch, M.L. (1991). Pigeons' memory for event duration: Differences between choice and successive matching tasks. *Learning and Motivation, 22,* 180–199.

Grant, D.S., Spetch, M.L., & Kelly, R. (1997). Pigeons' coding of event duration in delayed-matching-to-sample. In C.M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: Psychological and neurobiological analyses* (pp. 217–264). Amsterdam: Elsevier.

Honig, W.K., & Spetch, M. (1988). Short-term memory for rate of alternation in the pigeon. *Bulletin of the Psychonomic Society, 26,* 152–154.

Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.

Kelly, R., & Spetch, M.L. (2000). Choice biases in delayed matching-to-sample duration with pigeons: Manipulation of ITI and delay illumination. *Quarterly Journal of Experimental Psychology, 53B,* 309–323.

Roberts, W.A., Macuda, T., & Brodbeck, D.R. (1995). Memory for number of light flashes in the pigeon. *Animal Learning and Behavior, 23,* 182–188.

Santi, A., Coyle, J., Coppa, R., & Ross, L. (1998). The timing of auditory and visual signals by pigeons: Differential sensitivity to intertrial interval duration. *Learning and Motivation, 29,* 245–365.

Santi, A., Stanford, L., & Coyle, J. (1988). Pigeons' memory for event duration: Differences between visual and auditory signals. *Animal Learning and Behavior, 26,* 163–171.

Spetch, M.L. (1987). Systematic errors in pigeons' memory for event duration: Interaction between training and test delay. *Animal Learning and Behavior, 15,* 1–5.

Spetch, M.L., & Rusak, B. (1992). Temporal context effects in pigeons' memory for event duration. *Learning and Motivation, 23,* 117–144.

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

Wearden, J.H. (1994). Prescriptions for models of biopsychological time. In M. Oaksford & G. Brown (Eds.), *Neurodynamics and psychology* (pp. 215–236). London: Academic Press.

Wearden, J.H., & Culpin, V. (1995). Exploring scalar timing theory with human subjects. In M. Richelle, V. DeKeyser, G. d'Ydewalle, & A. Vandierendock (Eds.), *Time and the dynamic control of behavior* (pp. 21–33). Liège: University of Liège Press.

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

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

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

Wearden, J.H., & Ferrara, A. (1993). Subjective shortening in humans' memory for stimulus duration. *Quarterly Journal of Experimental Psychology, 46B,* 163–186.

Wearden, J.H., & McShane, B. (1988). Interval production as an analogue of the peak procedure: Evidence for similarity of human and animal timing processes. *Quarterly Journal of Experimental Psychology, 40B,* 363–375.

Manuscript received 20 October 1998
Accepted revision received 16 January 2001