The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Subjective shortening with filled and unfilled auditory and visual intervals in humans?

J. H. Wearden [a]; Greg Goodson [a]; Karin Foran [a]
[a] Keele University, Keele, UK

First Published on: 30 January 2007
To cite this Article: Wearden, J. H., Goodson, Greg and Foran, Karin (2007) 'Subjective shortening with filled and unfilled auditory and visual intervals in humans?', The Quarterly Journal of Experimental Psychology, 60:12, 1616 - 1628
To link to this article: DOI: 10.1080/17470210601121916
URL: http://dx.doi.org/10.1080/17470210601121916

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2007, 60 (12), 1616–1628

Psychology Press
Taylor & Francis Group

Subjective shortening with filled and unfilled auditory and visual intervals in humans?

J. H. Wearden, Greg Goodson, and Karin Foran
Keele University, Keele, UK

Two experiments tested humans on a memory for duration task based on the method of Wearden and Ferrara (1993), which had previously provided evidence for subjective shortening in memory for stimulus duration. Auditory stimuli were tones (filled) or click-defined intervals (unfilled). Filled visual stimuli were either squares or lines, with the unfilled interval being the time between two line presentations. In Experiment 1, good evidence for subjective shortening was found when filled and unfilled visual stimuli, or filled auditory stimuli, were used, but evidence for subjective shortening with unfilled auditory stimuli was more ambiguous. Experiment 2 used a simplified variant of the Wearden and Ferrara task, and evidence for subjective shortening was obtained from all four stimulus types.

A by-product of interest in time perception in humans and nonhuman animals over the last two and a half decades has been the phenomenon of *subjective shortening*, the apparent effect that when the duration of some event is retained in working memory, the duration becomes subjectively shorter as retention interval increases. Although prefigured in some earlier work with animals (e.g., Church, 1980) the initial impetus for interest in subjective shortening came from Spetch and Wilkie (1983), who used pigeons in a delayed-matching-to-sample task. To simplify slightly, pigeons were initially presented with visual samples either 2 s or 8 s long, displayed on the centre key of an operant chamber, and were then required to respond on one of two side keys, with the correct choice being based on the duration of the previous sample. In initial training, offset of the sample was accompanied by onset of the side keys, but tests were then introduced where the side keys were illuminated only after a delay had passed since sample offset. When the previous sample had been the shorter of the two, performance accuracy held up well with increasing delay, but deteriorated when the previous sample had been the longer one, as the animals tended to “choose short” with increasing delay.

One interpretation of this “choose short” effect is that the memory of the sample was becoming subjectively shorter as the delay increased, and some evidence for this came from situations in which the training delay between offset of the sample and onset of the response keys was greater than zero (Spetch, 1987). Tests with shorter delays than in training produced a “choose long” effect, as the subjective shortening hypothesis predicts.

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

Karin Foran conducted some of the experimental work reported here while a student in the then-titled Department of Psychology at Manchester University.

http://www.psypress.com/qjep
DOI:10.1080/17470210601121916

Wearden and Ferrara (1993, see also Wearden, Parry, & Stamp, 2002) investigated subjective shortening with humans. Simple variants of the method used by Spetch and Wilkie cannot be employed with humans, even if short durations are used so as to prevent chronometric counting, as participants would just verbally label the samples as "short" and "long" and retain the verbal label throughout the trial. Wearden and Ferrara (1993) introduced a method used in two different forms in the present article. Two stimuli (e.g., tones) are presented on each trial, separated by a gap: The first we call the "sample" (s), the second the "comparison" (c). Offset of c is followed by a prompt asking for a decision as to whether the second stimulus was longer than, shorter than, or of the same duration as the first, and all possibilities are equally likely. The sample duration varies from trial to trial at random, and the comparison has the same duration, or is 100 ms shorter or longer. Sample–comparison delays up to 16 s were used by Wearden and Ferrara (1993), and the general principle of the procedure is that the "fresh" memory of c is compared with an increasingly "subjectively shortened" memory of s as the $s{-}c$ delay increases.

The method, although simple to describe in outline, generates a complex set of results, mainly because the putative subjective shortening will not have the same effect on trials of different types. We define the three different types of trials in terms of the correct response: EQUAL ($s = c$), SHORT ($s > c$), and LONG ($s < c$). On EQUAL trials, subjective shortening with increasing delay will presumably cause increasing rates of error, but the errors should generally change in a predicted direction: People should tend to make the "long" judgement increasingly as $s{-}c$ delay increases. On SHORT trials, s really is longer than c, so performance accuracy may change little with increasing $s{-}c$ delay. On LONG trials, on the other hand, s really is shorter than c, and any subjective shortening should increase the subjective difference between s and c, resulting in an *increase* in performance accuracy on these trials as the $s{-}c$ delay increases. In general, increasing performance accuracy with increasing $s{-}c$ delay on LONG trials can be regarded as one of the "signatures" of subjective shortening, and this result was found with tones (Wearden & Ferrara, 1993) and when the durations of bar-like visual stimuli were judged, but not when their physical lengths were the basis of the decision (Wearden et al., 2002).

The other "signature" of subjective shortening is the pattern of errors on EQUAL trials. In this case "long" errors (erroneously judging that $c > s$) should increase in frequency with increasing $s{-}c$ delay, whereas "short" errors (judging that $c < s$) should decrease, and such error patterns have also been found (Wearden & Ferrara, 1993; Wearden et al., 2002). The subjective shortening hypothesis also makes predictions about error patterns on the other trial types, but these are hard to evaluate as, for SHORT and LONG trials, "equal" errors are much more common than "long" and "short" errors, respectively. Indeed, many participants in the present study never made any of the minority errors at all. We only present errors from EQUAL trials in both experiments reported here.

Subjective shortening (or "choose short") effects thus appear fairly robust in humans and have also been found by Lievling, Lane, Cherek, and Tcheremissine (2006), using a method different from that of Wearden and Ferrara (1993). However, a series of experiments mostly by Santi and colleagues, using pigeons, has complicated the initial interpretation of the phenomenon advanced by Spetch and Wilkie (1983). For example, Santi, Stanford, and Coyle (1998b) reported that pigeons failed to show a "choose short" effect with filled auditory stimuli, although one was present with visual stimuli. More recently, Santi, Hornyak, and Miki (2003) found evidence that subjective shortening is absent in pigeons when visual intervals are "unfilled" (i.e., when they start and end with brief stimuli) whereas the effect is present with filled (i.e., continuous) intervals. For example, their Figure 1 (p. 287) shows that with a 9-s delay, performance accuracy was better after short samples when these were filled but after long samples when these were unfilled. In general, responses after unfilled intervals seemed better described by a "choose long" effect

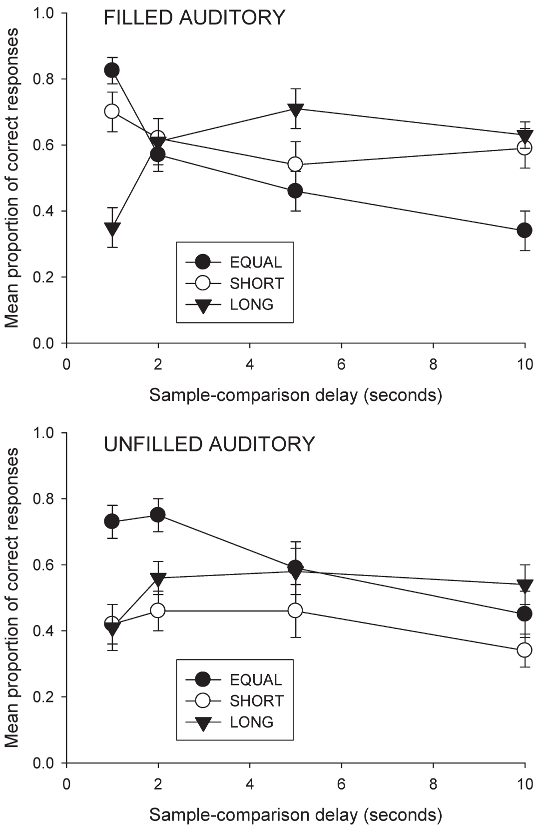

Figure 1. *Mean proportion of correct responses plotted against* $s-c$ *delay. Data are shown separately for the three trial types (EQUAL, SHORT, and LONG). Upper panel: results with filled auditory stimuli. Lower panel: results with unfilled auditory stimuli. Vertical lines indicate the standard error of the mean.*

than by a "choose short" one. Other apparent differences in memory for filled and unfilled intervals in pigeons were obtained by Grant and Talarico (2004).

The observation of the absence of apparent subjective shortening with unfilled intervals in Santi et al. (2003) is the starting point for the present article, which has the simple aim of investigating potential subjective shortening in humans when the basis of the judgements are filled and unfilled auditory, or visual, signals. Previous work (Wearden & Ferrara, 1993; Wearden et al., 2002) has used only filled stimuli (tones, squares, or lines of colour on a computer screen).

EXPERIMENT 1

The methods used in the present work are all small variants on the technique introduced by Wearden and Ferrara (1993). In Experiment 1, one participant group received auditory *s* and *c* stimuli, which, in different conditions, were both continuous tones (filled) or both intervals started and ended with brief clicks (unfilled). Another group received bar-like visual stimuli, and for the filled condition a single bar was presented as *s* and as *c*, and its duration of presentation was the basis of the judgement. For the unfilled condition, two brief presentations of the bar defined an unfilled interval for both *s* and *c*.

Method

Participants

A total of 41 undergraduate psychology students (some from Manchester University, others from Keele University) were arbitrarily allocated to two groups. One group (20 participants) received auditory stimuli, the other (21 participants) visual stimuli.

Apparatus

IBM-compatible computers controlled all experimental events. The experiment trials were programmed in the MEL (Micro Experimental Laboratory) language. The computer keyboard served as the response manipulandum.

Procedure

The basic experimental procedure involved the presentation of two stimuli on each trial, the first the sample (*s*), the second the comparison (*c*). The $s-c$ delays (offset to onset) took 4 values: 1, 2, 5, and 10 s. There were three trial types, defined by the correct answer: EQUAL trials ($s = c$), SHORT trials ($s > c$), and LONG trials ($s < c$). In all conditions, the participant pressed the spacebar to produce the trial events and then after offset of *c* received a display, which, for the filled auditory intervals, was "Was the duration of the SECOND tone longer than (L), shorter than (S), or equal to (E) the duration of the FIRST one?". The display obviously varied for the different stimulus types used.

Auditory conditions. The filled auditory intervals were 500-Hz tones produced by the computer speaker. To produce the unfilled auditory intervals, two 1,000-Hz 20-ms tones separated by a gap defined *s* and *c*. The 20-ms durations were not included in the values of *s* and *c*. The duration of *s* and *c* depended on the trial type. On EQUAL trials, *s* was randomly sampled from a uniform distribution running from 350 to 650 ms, then repeated as *c*. On SHORT trials, *s* was randomly sampled from 500 to 650 ms, and *c* was 150 ms shorter. On LONG trials, *s* was randomly sampled from 350 to 500 ms, and *c* was 150 ms longer. These values prevented participants from identifying the correct response solely on the basis of one of either *s* or *c* alone, and the duration values used had previously resulted in subjective shortening with filled intervals (Wearden & Ferrara, 1993). The three trial types and four $s-c$ delays were combined into a block of 12 stimuli, and five blocks (60 trials) were presented in all, with the trial order within and between blocks being randomized. Half the participants received the filled condition first then after a short break the unfilled one; for the other half the order was reversed.

Visual conditions. The stimuli used were yellow lines approximately 4 cm long and 3 mm wide presented in the centre of the computer screen. For the filled intervals *s* and *c* comprised a single line, and the duration of presentation of this line was the basis of the judgement. For unfilled intervals, the lines were presented twice, each time for 300 ms, to define unfilled intervals for *s* and *c*. The 300-ms "marker" durations were not included in the values of *s* and *c*. Participants received detailed experimental instructions including diagrams making clear what the appropriate events to be judged in the filled and unfilled cases were.

Duration values were selected that had previously produced subjective shortening with filled intervals (Wearden et al., 2002). On EQUAL trials, *s* was randomly sampled from a range from 250–550 ms and was repeated as *c*. On SHORT trials, *s* was randomly selected from 400 to 550 ms, and *c* was 150 ms shorter. On LONG trials, *s* was randomly selected from 250–400 ms, and *c* was 150 ms longer. The three trial types and four $s-c$ delays were combined into a block of 12 stimuli, and six blocks (72 trials) were presented in all, with the trial order within and between blocks being randomized. A total of 10 of the participants received the filled condition first then after a short break the unfilled one; for the other 11 the order was reversed.

Results

Figure 1 shows the mean proportion of correct responses for the auditory group, plotted against $s-c$ delay. Data are shown separately for the EQUAL, SHORT, and LONG trials. Data from the filled intervals are in the top panel, those from the unfilled intervals in the lower panel.

Inspection of data in the filled interval case suggests that the proportion of correct responses decreased with increasing $s-c$ delays on EQUAL trials, showed a slight decline on SHORT trials, and increased on LONG trials. An overall analysis of variance (ANOVA) found no overall effect of trial type (EQUAL, SHORT, or LONG), $F(2, 38) = 0.546$, nor $s-c$ delay, $F(3, 57) = 2.56$, $p = .06$, although the effect approached significance, but a significant Trial Type × $s-c$ Delay interaction, $F(6, 114) = 11.89$, $p < .001$, the latter result confirming that the effect of $s-c$ delay was different for the different trial types. ANOVAs using each trial type separately found that the mean proportion correct decreased with increasing $s-c$ delay on EQUAL trials, $F(3, 57) = 16.56$, $p < .001$, showed no change on SHORT trials, $F(3, 57) = 1.73$, and increased on LONG trials, $F(3, 57) = 8.27$, $p < .001$.

The lower panel shows analogous data from judgements of unfilled auditory durations. Inspection of the data suggests a similar pattern of results to that obtained with filled intervals, although the increase in the proportion of correct responses on LONG trials appears less marked. An overall ANOVA found significant effects of trial type, $F(2, 38) = 6.05$, $p < .05$, $s-c$ delay, $F(3, 57) = 3.95$, $p < .05$, and a trial type by delay interaction, $F(6, 114) = 2.66$, $p < .05$. The proportion of correct responses significantly decreased on EQUAL trials, $F(3, 57) = 5.96$, $p < .01$, but showed no change on either SHORT, $F(3, 57) = 1.18$, or LONG trials, $F(3, 37) = 1.91$.

Figure 2 shows the mean number of errors made on EQUAL trials, plotted against $s-c$ delay for the filled (upper panel) and unfilled (lower panel) intervals. Inspection suggests that the mean number of "short" errors (i.e., erroneous judgements that $s > c$) showed little change with increasing $s-c$ delay, but that "long" errors (judging that $s < c$) increased with $s-c$ delay in both cases. In fact, the change in "short" errors was not significant in either condition: filled, $F(3, 57) = 2.48$, $p = .07$; unfilled, $F(3, 57) = 1.96$, $p = .13$, whereas the change in "long" errors was significant in both conditions: filled, $F(3, 57) = 8.805$, $p < .001$; unfilled, $F(3, 57) = 8.79$, $p < .001$.

Figure 3 shows mean proportion of correct responses plotted against $s-c$ delay for judgements of visual stimuli (filled: upper panel; unfilled: lower panel). Inspection of the data in both panels suggested that the mean proportion of correct

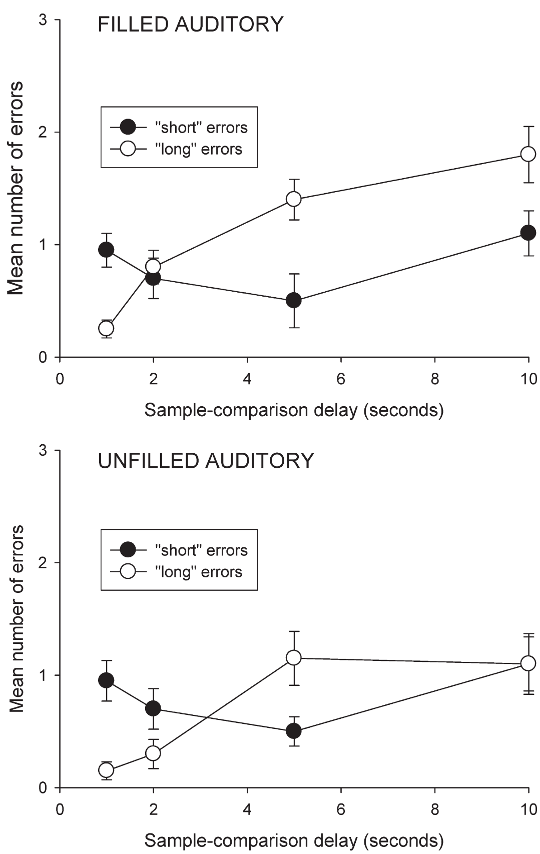

Figure 2. *Mean number of "long" and "short" errors on EQUAL trials, plotted against* s−c *delay. Upper panel: results with filled auditory stimuli. Lower panel: results with unfilled auditory stimuli. Vertical lines indicate the standard error of the mean.*

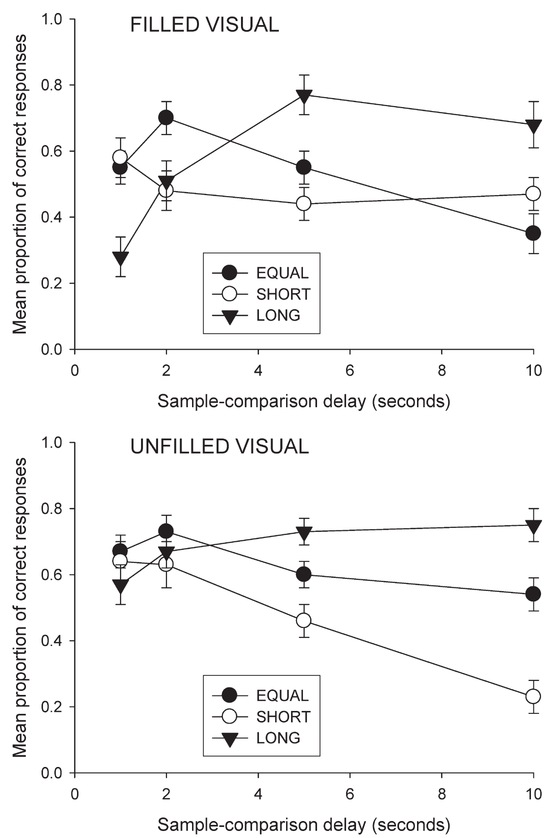

Figure 3. *Mean proportion of correct responses plotted against* s−c *delay. Data are shown separately for the three trial types (EQUAL, SHORT, and LONG). Upper panel: results with filled visual stimuli. Lower panel: results with unfilled visual stimuli. Vertical lines indicate the standard error of the mean.*

responses decreased with $s-c$ delay on EQUAL trials, decreased or stayed roughly constant on SHORT trials, and increased on LONG trials. An overall ANOVA on data from judgements of filled durations found no effect of trial type, $F(2, 40) = 0.71$, but significant effects of $s-c$ delay, $F(3, 60) = 5.27$, $p < .01$, and Type × Delay interaction, $F(6, 120) = 11.09$, $p < .001$. Simpler analysis found that the proportion of correct responses declined with $s-c$ delay on EQUAL trials, $F(3, 60) = 9.10$, $p < .001$, and increased on LONG trials, $F(3, 60) = 20.77$, $p < .001$, but showed no change on SHORT trials, $F(3, 60) = 1.247$.

The same analyses of data from judgements of unfilled intervals found significant effects of trial type, $F(2, 40) = 5.92$, $p < .01$, $s-c$ delay, $F(3, 60) = 7.10$, $p < .001$, and Type × Delay interaction, $F(6, 120) = 9.06$, $p < .001$. Mean proportion of correct responses decreased with increasing $s-c$ delay on both EQUAL, $F(3, 60) = 3.37$, $p < .05$, and SHORT trials, $F(3, 60) = 18.54$, $p < .001$, but increased on LONG trials, $F(3, 60) = 3.11$, $p < .05$.

Figure 4 shows the mean number of "short" and "long" errors from EQUAL trials for judgements of filled intervals (upper panel) and unfilled intervals (lower panel). Inspection of the data suggests that the mean number of "short" errors tended to decline with increasing $s-c$ delay, whereas the number of "long" errors tended to increase. In fact, the decline in "short" errors was significant only for the unfilled condition, $F(3, 60) = 6.18$, $p < .001$, although the effect for the filled condition was very close to significance, $F(3, 60) = 2.64$, $p = .058$. The increase in "long" errors was significant for both filled, $F(3, 60) = 6.58$, $p < .01$, and unfilled, $F(3, 60) = 12.38$, $p < .001$, intervals.

Discussion

To summarize the data in the present study, we can use the "signatures" of subjective shortening. One is an increase in performance accuracy on LONG trials with increasing $s-c$ delay. This was found with filled auditory intervals and both filled and unfilled visual intervals. Although there was a suggestion of such an increase with unfilled auditory intervals, the effect was not significant. The other "signature" has two parts: an increase in the number of "long" errors and a decrease in the number of "short" errors on EQUAL trials. The increase in the number of "long" errors was found in all conditions. The decrease in the number of "short" errors was found only for the unfilled visual condition, although there was no significant change in other cases (i.e., the number of "short" errors never significantly increased with increasing $s-c$ delay). Overall, therefore, the condition that was associated with fewest "signatures" of subjective

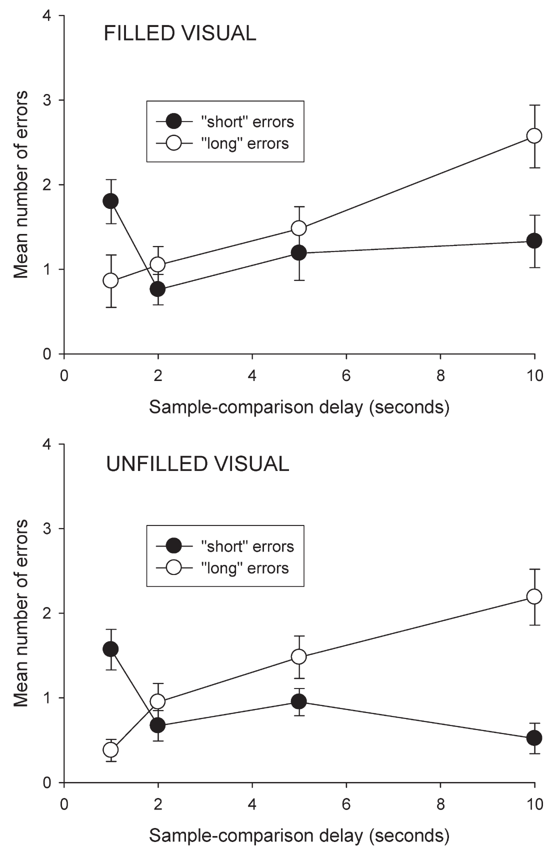

Figure 4. *Mean number of "long" and "short" errors on EQUAL trials, plotted against* s−c *delay. Upper panel: results with filled visual stimuli. Lower panel: results with unfilled visual stimuli. Vertical lines indicate the standard error of the mean.*

shortening (one out of the three) was unfilled auditory, whereas the one associated with all three was unfilled visual. Other conditions had two of the three "signatures".

Evidence of subjective shortening when filled auditory and visual stimuli are used replicates results in previous articles: Subjective shortening has been obtained when tones were used as stimuli (Wearden & Ferrara, 1993) and when squares or lines on a computer screen were employed (Wearden et al., 2002). When stimuli are unfilled intervals the position is apparently more complicated. In the present study, unfilled visual intervals exhibited all three "signatures" of subjective shortening, in contrast to the work of Santi et al. (2003), where such stimuli generated a reliable "choose long" effect. On the other hand, the ambiguity of subjective shortening when unfilled auditory intervals were used is reminiscent of results from Santi, Ross, Coppa, and Coyle (1999) with pigeons, where unfilled auditory intervals resulted in a "choose long" effect, at least at the longest delays used. Although in the present study, unfilled auditory intervals provided some evidence for subjective shortening (increase in proportion of "long" errors on EQUAL trials), the effect seems more fragile than that with other sorts of stimuli and did not reach conventional significance levels with the same number of participants as produced significant effects with other stimulus types.

Although Experiment 1 produced clear results for most of the stimulus types, it had a number of potential problems. One of these was that the structure of the equivalent auditory or visual conditions (filled or unfilled) was not identical in terms of number of trials, or organization of trials into blocks (although trial order was randomized in all cases, so the block structure was unlikely to be evident to participants). A second potential problem lies with the stimuli used: In the auditory case, the stimuli (a tone or a pair of clicks) were more distinctive than those in the visual case, where a single line or two lines were used, and participants may find the duration of a thin line more difficult to perceive than the interval between two successive presentations. Thirdly, the durations of the auditory and visual stimuli used in Experiment 1 were not exactly the same. Fourthly, the pattern of results produced by the Wearden and Ferrara (1993) method used in Experiment 1 is both highly complex and not always easy to interpret, particularly insofar as errors are concerned. We discussed the three potential "signatures" of subjective shortening, but all three may not be equally indicative or strong. In particular, the pattern of change of errors on EQUAL trials may be hard to interpret: For example, if some participants make fewer errors than others on such trials, they will necessarily have fewer opportunities to show changes in error patterns with changing $s-c$ delay, so leading to problems of statistical evaluation of potential effects. For example, when auditory filled intervals were used in Experiment 1, around a quarter of participants made no "short" or "long" errors, even with a 10-s $s-c$ delay.

To overcome some of these problems, we modified the procedure used above for Experiment 2. In Experiment 2, the only responses permitted were "short" or "long", but the majority of trials were EQUAL trials, so (a) all the responses produced on these trials were necessarily errors, and (b) we fixed in advance how many such errors would occur in all participants, as their errors were forced. Although SHORT and LONG trials were used in Experiment 2, they were infrequent and were present only to disguise the fact that the majority of trials were EQUAL trials. The prediction from the subjective shortening hypothesis is that the number of "long" responses on EQUAL trials would increase as $s-c$ delay increased. Another difference from Experiment 1 was that the difference between the filled and unfilled intervals was intended to be more equivalent in the auditory and visual cases. In the auditory case, the filled stimuli were tones, and the unfilled stimuli were click-defined intervals (as in Experiment 1), and in the visual case the filled interval was a square continuously presented on the screen, whereas the unfilled visual intervals were defined by the line presentations used in Experiment 1. A final change was that the durations of the auditory and visual stimuli used in Experiment 2 were exactly equated.

EXPERIMENT 2

Method

Participants

A total of 32 Keele University undergraduates were arbitrarily allocated to two equal-sized groups.

Apparatus

The apparatus was the same as that in Experiment 1.

Procedure

A total of 16 participants received both the filled and unfilled auditory stimuli (auditory group), and 16 received the filled and unfilled visual stimuli (visual group). In each group 8 participants received the filled condition first and 8 the unfilled condition first. The filled/unfilled experimental sessions were separated by a break of a few minutes. For the auditory group the stimuli were identical to those used in Experiment 1: 500-Hz tones (filled); 1,000-Hz tones 20 ms long (unfilled). For the visual group, the filled stimulus was a 10 × 10-cm light-blue square presented in the centre of the computer screen, and the unfilled interval was identical to that used in Experiment 1 (the time between two yellow lines).

The sample–comparison delays (offset to onset) were 1, 2, 5, and 10 s, with all being equally likely. The trial types were not, however, equally likely, with EQUAL trials being three times as frequent as other trial types. Each block consisted of 3 EQUAL trials, 1 LONG, and 1 SHORT trial at each delay (making 20 trials per block), and three blocks were given in all, making a total of 60 trials in each experimental condition (filled or unfilled). On EQUAL trials (from which the data in Experiment 2 were taken), the sample duration was a random value drawn from a uniform distribution running from 350 to 650 ms, and this was repeated as the comparison. On SHORT trials, the sample was drawn from a uniform distribution running from 500 to 650 ms, and the comparison was 150 ms shorter; on LONG trials, the sample was drawn from a uniform distribution running from 350 to 500 ms, and the comparison was 150 ms longer. That is, the durations for all stimulus types were the same as those for the auditory stimuli in Experiment 1.

The only responses permitted were "short" or "long": For example, for the filled auditory case, a display read "Was the duration of the SECOND tone longer than (L) or shorter than (S) the duration of the FIRST tone?", with equivalent displays being presented for the other stimulus types. General instructions and other details were the same as those in Experiment 1.

Results

Data were taken from EQUAL trials, and Figure 5 shows the mean number of "long" responses plotted against $s-c$ delay for auditory stimuli (upper panel), and visual stimuli (lower panel), with data from the filled and unfilled conditions shown in each panel. The maximum possible number of responses at each delay was 9.0, so the mean number of "short" responses (which is not shown) was just 9.0 minus the number of "long" responses.

Inspection of Figure 5 suggests that the mean number of "long" responses increased with increasing $s-c$ delay for all stimulus types, with no obvious difference between either the auditory or visual conditions, or the filled/unfilled conditions. This was confirmed by an overall ANOVA, which involved filled/unfilled (type), and $s-c$ delay (delay) as within-subject factors and auditory/visual (group) as the between-subject factor. There was no significant effect of type of stimulus, $F(1, 30) = 0.03$, but there was a highly significant effect of delay, $F(3, 90) = 37.18$, $p < .001$. The group effect (auditory versus visual) was not significant, $F(1, 30) = 0.57$, nor was any interaction significant or approaching significance: Type × Group, $F(1, 30) = 1.78$; Delay × Group, $F(3, 90) = 1.48$; Type × Delay, $F(3, 90) = 1.53$; Type × Delay × Group, $F(3, 90) = 1.26$.

When data from the auditory and visual groups were analysed separately, ANOVAs found the only significant effect to be $s-c$ delay: auditory, $F(3, 45) = 21.43$; visual, $F(3, 45) = 18.32$, both $p < .001$. No other main effect or interaction approached significance in either group. When

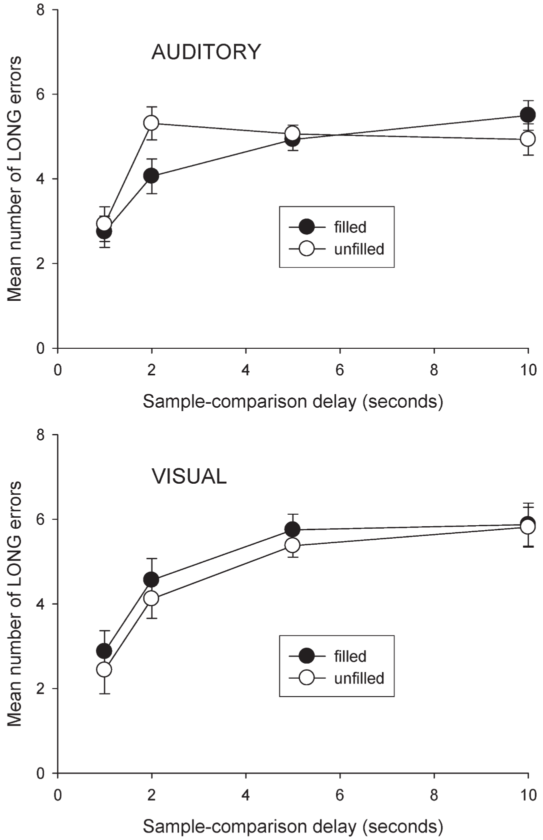

Figure 5. *Mean number of "long" responses on EQUAL trials from Experiment 2, plotted against* s−c *delay. Upper panel: auditory intervals. Lower panel: visual intervals. In each panel, data from filled intervals are indicated by filled circles, data from unfilled intervals by unfilled circles. Vertical lines indicate the standard error of the mean.*

data from each stimulus type were analysed separately, there was always a significant effect ($p < .001$) of *s−c* delay: *F* values, all $F(3, 45)$, were: filled auditory, 10.81; unfilled auditory, 10.45; filled visual, 8.37; unfilled visual, 13.72.

Discussion

In the data collected in Experiment 2, which came solely from the errors made on the predominant EQUAL trials, the subjective shortening hypothesis predicts increasing numbers of "long" responses with increasing *s−c* delay, and such an effect was found with all stimulus types. There were no significant differences (or trends towards significance) when comparing auditory and visual stimuli, or filled and unfilled stimuli. The procedure of Experiment 2 was not only simpler than that of Experiment 1, but also perhaps more sensitive in detecting any effects of *s−c* delay. The number of errors on EQUAL trials was equated across participants (as all responses on these trials were errors), so differences between participants in accuracy on these trials were eliminated. In addition, the filled and unfilled stimuli in the auditory and visual cases were distinct (unlike in Experiment 1), and the number of trials, and organization of trials into blocks, was identical across conditions. The number of EQUAL trials (36) was also greater than that in Experiment 1 (where it was 20 and 24, respectively, for the auditory and visual conditions), so probably resulted in more reliable data. Obviously, our data suggest a clear subjective shortening effect for both filled and unfilled auditory and visual stimuli, although another possible interpretation is discussed below.

GENERAL DISCUSSION

Overall, the data from our two experiments support the view that subjective shortening in humans' memory for stimulus duration is found with all the four stimulus types that we employed, unlike reported data from pigeons, where effects apparently depend on whether stimuli are auditory or visual, or filled or unfilled.

Although our data are consistent with the general notion of subjective shortening, another possibility should be discussed. In the original Wearden and Ferrara (1993) method, and variants used since (Wearden & Culpin, 1998; Wearden et al., 2002; the present article), two stimuli are present on each trial, raising the question of the involvement of the *time order error* (TOE). TOEs (see Hellstrom, 1985, for a review), are a persistent, if rather mysterious, phenomenon affecting many sorts of judgements of sequentially presented stimuli. For example, if A and B are two stimuli about which some relative judgement is made (e.g., which is longer, heavier, more pleasant), TOEs are said to occur if the judgements made when the stimuli are presented in the order AB are not the same as when the order is BA. In general, a *positive* TOE is one in which A is "weighted" more than B, a *negative* TOE the reverse. In the case of erroneous performance on the EQUAL trials of our experiments, for example, a positive TOE would lead to "short" responses, and a negative one to "long" responses. One interpretation of some of our data is that, rather than the results being due to a subjective shortening process, they are "simply" due to a change of TOE type with $s-c$ delay, from positive at short delays to negative at long delays.

The word "simply" is highlighted above because such an explanation would in reality be far from simple. The discussion of the relation between TOEs and subjective shortening effects is somewhat complicated, but we attempt it here because the issue is often raised by reviewers and other readers and deserves some treatment.

The first point to make is that subjective shortening effects and TOEs are not mutually exclusive: Wearden and Ferrara (1993) in fact interpreted their results overall as being due to a positive TOE superimposed on a subjective shortening effect, and this possibility is discussed with respect to the present data later.

A second point concerns the theoretical or logical status of the TOE as an explanation. A TOE of some sort could be regarded as an unexplained "bias" (e.g., "people tend to judge the first stimulus of two as bigger, but we don't know why"), or it could be derived from some deeper theoretical considerations. Unexplained "biases" may be theoretically unsatisfactory, but their usage is not entirely vacuous, as the treatment by Baum (1974) shows. Baum was considering data from animal subjects performing under concurrent schedules. To simplify the procedure, pigeons were confronted with two response keys, one green, one red, and responding on each key was reinforced according to an independent schedule of reinforcement. In the experimental procedure, the schedules on the two keys were changed from one condition to another, and the animal's choice (i.e., the distribution of its responses between them) was measured. The focus of interest in general in such experiments is how the pigeon's choice varies as a function of the relative rates of reinforcement that the two keys provide (e.g., see Herrnstein, 1961, for the classic result). Baum (1974), however, argued that a proper description of choice would take into account both the relative rates of reinforcement delivered in the situation and idiosyncratic "biases" that might be present or absent, or differ from one pigeon to another. For example, in the case cited above, a particular pigeon might "prefer" pecking red to green, another might have the reverse tendency, and yet a third might be indifferent. These "biases" were regarded as being consistent across conditions within a particular animal, thus independent of the independent variable manipulation, the change of relative rate of reinforcement. Baum (1974) showed how the bias could be measured independent of the effects of relative reinforcement rate, so a model using such bias would fit data much better than one without, even though the source of bias was unexplained. The important point here is that the bias is independent of the variable that is experimentally manipulated across conditions.

A TOE might be used in the same way as Baum's "bias"—for example, people *always* tend to judge that the first of two presented stimuli is longer than the second—that is, *independent of anything else*. The bias might or might not be explainable, but it is always present. In our study, such a usage might explain the numbers of responses at some particular $s-c$ delay, but cannot account for effects of changing $s-c$ delay as, by definition, the TOE effect is independent of the $s-c$ delay manipulation.

A more pernicious use of the TOE concept would involve changing the TOE type as some experimental variable, such as $s-c$ delay, is manipulated, so the TOE changes from positive to negative as $s-c$ delay increases. This seems to us a pseudoexplanation, and an abuse of the term TOE, as no reason is given for why the form of the TOE should change, and the "explanation" gives an entirely false sense of theoretical closure. Any data can be explained by "biases" if these are allowed to change with manipulations of independent variables of interest. To give an absurd example, "there is no such thing as forgetting, only a bias to give the wrong response, or no response, which increases with time since an item was stored". Here, obviously, an explanation in terms of an internal process of forgetting is replaced by a pattern of bias that changes conveniently to mimic the obtained results.

It seems to us, therefore, theoretically incoherent to account for the totality of our results with a TOE, but that does not mean that one is not present in our data, or that a TOE cannot contribute to a partial explanation of our findings. The simplest data to discuss in this context are the "long" responses from Experiment 2. As Figure 5 shows, the mean number of such responses goes from around 2.5 at an $s-c$ delay of 1 s to nearly 6 at 10 s, in the different conditions. In Figure 5, all responses are errors, and the maximum number possible at any given $s-c$ delay is 9.0. If there were no TOE, and a "perfect" subjective shortening process, then the number of "long" responses would be expected to be close to the chance level (4.5) at 1 s (when subjective shortening is not expected to be very pronounced) to approaching 9 at 10 s, but the numbers were always systematically smaller than these. The obvious suggestion is that a positive TOE (a tendency to judge that the first stimulus is longer than the second, which reduces the number of "long" responses) is present throughout, but that a subjective shortening process, which increases in strength with $s-c$ delay, is superimposed on top of it. Can this suggestion (similar to that of Wearden & Ferrara, 1993) be tested? If the results are solely the result of a positive TOE at the shortest $s-c$ delays, which "wears off" as $s-c$ delay increases, then the number of "long" responses should be below chance at 1 s, but not significantly different from chance at 10 s. We tested this by comparing the number of "long" responses at 1 s and 10 s (aggregating data from the auditory and visual conditions together, a step justified by the lack of significant difference between them) with the chance value of 4.5. At an $s-c$ delay of 1 s, the number of "long" responses was significantly below chance (i.e., assuming that "long" and "short" responses were equally likely) for both filled, $t(31) = -5.57$, $p < .001$, and unfilled, $t(31) = -5.42$, $p < .001$, conditions, whereas at an $s-c$ delay of 10 s, the number of "long" responses was significantly above chance for both filled, $t(31) = 3.88$, $p < .01$, and unfilled, $t(31) = 2.88$, $p < .01$, intervals.

The significant effect at 1 s suggests that a positive TOE operates in our Experiment 2, but the significant result at 10 s shows that a reduction of the TOE with $s-c$ delay cannot be the sole explanation of the results: Subjective shortening appears necessary in addition to a TOE. Why should a TOE be present at all? One possibility is that the positive TOE in the present case results from the rapid sequential processing of two durations, so when these are close together (i.e., short $s-c$ delay) the time taken to produce a representation of the sample interferes with the formation of representation of the comparison. For example, suppose that both stimulus durations are timed by a pacemaker–accumulator clock of the type proposed by scalar expectancy theory (Gibbon, Church, & Meck, 1984), a leading account of timing in animals and humans. Onset of the sample causes the switch of the clock to close, and termination of the sample causes it to open again, with the contents of the accumulator being transferred to working memory. However, this transfer may not be instantaneous so, at short $s-c$ delays, the comparison may be presented before the process is finished, thus possibly resulting in "missed pulses" when the comparison is timed. In general, this would lead to the comparison being judged as shorter than the sample when the $s-c$ delay is short, but when the $s-c$ delay is longer, there is sufficient time for the processing of the sample before the comparison is presented. This proposed mechanism, while speculative, provides an account of why a positive TOE might be manifested at short $s-c$ delays and not at longer delays but, as shown above, this "wearing off" of the TOE is not sufficient to produce the results observed, without a subjective shortening mechanism also operating.

Other data obtained using the Wearden and Ferrara (1993) method also suggest that a positive TOE, as well as subjective shortening, is present in the results. Although, as mentioned above, the number of "short" and "long" errors on EQUAL trials may have problems of statistical evaluation because of between-subject differences in the number of errors committed, in general, a positive TOE would suggest that on EQUAL trials there would be more "short" errors than "long" errors at shortest $s-c$ delays used, and inspection of previous results shows that this is true. For example, in Wearden and Ferrara (1993), Experiment 3, which used filled auditory stimuli, with a 2-s $s-c$ delay (the shortest used), there were many more "short" errors than "long" errors (see their Figure 5, p. 179), although the data shown were the total for the group rather than the mean, and no statistical analysis was carried out. In Wearden et al. (2002), which used filled visual stimuli, the same effect can be noted in Experiment 2 (Figure 4, p. 14) and Experiment 3 (Figure 6, p. 20). In both cases, on EQUAL trials, there were more "short" errors at the 1-s $s-c$ delay than "long" errors, although once again the difference was not evaluated statistically. In the data reported in the present article, there were once again always more "short" errors than "long" errors at the 1-s $s-c$ delay (see Figures 2 and 4), and the difference was in all cases significant: filled auditory, $t(19) = 2.18$; unfilled auditory, $t(19) = 3.76$; filled visual, $t(20) = 2.32$; unfilled visual, $t(20) = 4.34$, all $p < .05$.

In general, then, TOEs may very well exist in data from procedures where two stimulus durations are presented sequentially, but it seems unlikely that any coherent explanation in terms of TOE alone can account for the results: Subjective shortening seems a real phenomenon, which is necessary to account for the data obtained in experiments like those reported here.

Why are the data in the present study different, at least by implication, from those obtained from pigeons by Santi and colleagues? There is no clear answer to this question, but we might note that the experimental procedures used in experiments with pigeons are complicated, and the possibility exists that the results obtained in some cases are due to "instructional ambiguity" (i.e., the pigeons cannot master task requirements and use features of the experimental situation, such as stimuli present in the intertrial interval, which are actually irrelevant as far as the experimenter is concerned; see Santi, Coyle, Coppa, & Ross, 1998a, for discussion, and Zentall, 2005, for a review of "instructional" effects in experiments with animals). Instructional ambiguity is unlikely to play a role in the present work, where verbally competent undergraduate students appeared to understand, and appeared to try to perform in accordance with, the experimental instructions. A possible complicating factor in comparison of data from pigeons and humans with unfilled intervals is the fact that pigeons appear to judge unfilled intervals as longer than filled intervals (Miki & Santi, 2005), the opposite result to that usually obtained with humans (e.g., see Fraisse, 1964, for discussion of the "filled duration illusion"). In our unpublished observations of comparison of intervals defined by continuous tones and those defined by brief clicks, we have found the filled intervals to be judged as up to 40% longer when verbal estimation methods are used. Whether this species difference in the subjective duration of filled and unfilled intervals contributes to the apparent species differences in subjective shortening with different stimulus types remains to be seen. However, in the case of humans, subjective shortening has been found with filled auditory and filled visual stimuli, even though auditory stimuli have longer subjective durations than visual stimuli of the same real-time values (Wearden, Edwards, Fakhri, & Percival, 1998; Wearden, Todd, & Jones, 2006), so absolute or relative differences in subjective durations between different stimulus types do not seem to play an important role in determining whether or not subjective shortening occurs.

Original manuscript received 19 May 2006
Accepted revision received 26 August 2006
First published online 30 January 2007

REFERENCES

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

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

Fraisse, P. (1964). *The psychology of time*. London: Eyre and Spottiswoode.

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

Grant, D. S., & Talarico, D. C. (2004). Processing of empty and filled time intervals in pigeons. *Learning and Behavior*, *32*, 477–490.

Hellstrom, A. (1985). The time-order error and its relatives: Mirrors of cognitive comparing. *Psychological Bulletin*, *97*, 35–61.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.

Lievling, 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*, *58B*, 31–45.

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

Santi, A., Hornyak, S., & Miki, A. (2003). Pigeons' memory for empty and filled time intervals signalled by light. *Learning and Motivation*, *34*, 282–302.

Santi, A., Ross, L., Coppa, R., & Coyle, J. (1999). Pigeons' memory for empty time intervals marked by visual and auditory stimuli. *Animal Learning and Behavior*, *27*, 190–205.

Santi, A., Stanford, L., & Coyle, J. (1998b). 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., & 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., & 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, Germany: Hogrefe and Huber.

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

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

Zentall, T. R. (2005). Timing, memory for intervals, and memory for untimed stimuli: The role of instructional ambiguity. *Behavioural Processes*, *70*, 209–222.