Filled-duration illusions

J.H. Wearden[1] and Ruth S. Ogden[2]

1. School of Psychology, Keele University, Keele, ST5 5BG, U.K, and Division of Neuroscience and Experimental Psychology, Manchester University, Manchester, M13 9PL, U.K.
2. School of Psychology, Liverpool John Moores University, Liverpool, L3 3AF, UK.

Acknowledgement: The research for the preparation of this article was aided by an Emeritus Fellowship from the Leverhulme Trust awarded to the first author.

Address for correspondence: j.h.wearden@keele.ac.uk

Abstract

Data relevant to the “filled duration illusion”, the claim that filled intervals appear to last longer than unfilled ones of the same real duration, are reviewed. A distinction is made between divided-time studies (where an empty interval has one or more than one brief dividing stimulus inside it) and filled-duration studies (where the filled intervals are filled with some continuous event). Divided durations appear to last longer than empty ones, and the effect grows with the number of dividers, although it may be restricted to short durations. The best current explanation appears to involve the weighted summation of the different sub-intervals of which the total duration is composed. When intervals with simple fillers are contrasted with empty ones, they are usually judged as longer, and the effect may grow as the intervals lengthen, at least over short duration ranges. When complex fillers are used, fillers usually have no effect on perceived duration or shorten it. A pacemaker-counter approach can account for some simple filler effects, and division of attention for complex filler effects. Although there are some exceptions, “filled interval illusions” of all these types are normally found, but some problems, such as questions about the relative perceived variability of filled and unfilled intervals, or stimulus order effects, merit further research.

Key words: filled interval; unfilled interval; divided-time effect; pacemaker rate

The present article reviews some data and theory relevant to the “filled-duration illusion”, that is, the contention that a filled interval is judged to last longer than an empty one of the same real duration. It is unclear to us why this result should be described as illusory, when other differences in subjective duration, such as those arising from auditory and visual stimuli (Wearden & Jones, 2021) are not, so we will avoid describing the effects we discuss as illusions. The studies that we review employ two types of stimulus comparisons which appear quite distinct, hence this article’s title. One compares filled intervals, that is, intervals continuously filled by a stimulus or some activity, with an empty interval, such as one bounded by brief markers at its start and end. A *filled-duration effect* would be the result that the filled interval (e.g., a continuous tone) appeared to last longer than the unfilled one (e.g., the same interval started and ended with brief clicks). The second experimental arrangement compares empty intervals with “divided” ones, for example, intervals with one or more brief stimuli (such as clicks or flashes) presented between the start and end. Ten Hoopen, Miyauchi, and Nakajima (2008) distinguish these from filled-interval studies proper, calling them “divided-time” experiments, and we will follow this distinction, keeping the two types of studies separate. In this latter case we will refer to a *divided-time effect*, where an interval that contains one or more brief stimuli is judged to last longer than one just bounded by brief stimuli at its start and end.

Experiments on the “filled-duration illusion” amount to a substantial body of work, larger in fact than that of studies comparing judgements of the duration of auditory and visual stimuli (reviewed by Wearden & Jones, 2021). However, as will be obvious later, different articles have tended to focus on different aspects of the effects that can be obtained, so some questions may be hard to answer systematically because of the sparseness of the data. For example, Wearden and Jones (2021), reviewing subjective time differences arising from the use of auditory and visual stimuli, were able to draw conclusions about variability differences between time judgements of stimuli in the different modalities. Such a conclusion is much harder to draw with the material reviewed in the present article, but some evidence is available.

To review the material here we will proceed as follows. First, divided-time studies, which were historically prior, and filled-interval studies will be separated into two main sections. Then, within each section, we will review results obtained, generally providing an exposition in historical order of publication. This will be followed by a short summary and discussion of how the data discussed might be explained theoretically.

Divided-time studies

This review begins with divided-time studies because these were the first experiments on the “filled-duration illusion” to be reported, in 1886 by Hall and Jastrow. In fact, the earliest researchers may have found it easier to conduct divided-time studies than ones with continuously filled stimuli, as various sorts of rotating devices, where the speed of rotation could be controlled electrically or by other means, could be modified so that their rotation produced brief clicks (see Hall & Jastrow, 1886, for an example). Varying the speed of rotation could vary the time between the clicks in a precisely controlled way.

In most divided-time studies, the duration of an empty interval, started and ended by brief stimuli, is contrasted with the duration of a divided one, in which other stimuli occur between the start and end, with the dividers usually being brief stimuli, although sometimes different ones from the stimuli which start and end the interval. One study discussed below (Roelofs & Zeeman, 1951/1952) contrasted divided intervals with completely filled ones, but this is unusual. The divider stimuli themselves usually are content-free, for example, being clicks or flashes, although one study uses slides with content as dividers (Schiffman & Bobko, 1977).

The simplest case is that where a single divider is presented, although some studies use more in some conditions, and the possible effect of the number of dividers has been a focus of research, with another being the effect of the interval duration itself. Almost all studies focus on the central issue of average subjective duration differences between a divided interval and an empty one and, to anticipate results to be discussed below, the divided interval is very frequently judged as longer than the empty one: the *divided-time effect*.

Hall and Jastrow (1886) should be credited with the first divided-time study, although it was presented only briefly at the end of an article mostly about the counting of clicks. Their experiment on divided time is not described in detail, but appeared to involve presenting an interval containing clicks, followed by a “vacant” interval started and ended with a single click. Dividing the interval made its apparent duration longer, as “With 10 clicks the following vacant interval must be extended to the time of 14 to 18 clicks… [to appear equal]” (p. 62). Hall and Jastrow further noted that the effect appeared to decrease as the intervals became longer, although it was also not manifested at the shortest times they used and, furthermore, they reported that the effect appeared to be dependent on the order in which the vacant and divided intervals appeared. “This entire illusion…is reduced to a minimum…if the vacant interval precedes [the divided one]” (p. 62).

Figure 1 about here

A much more elaborate early investigation of the divided time effect was published by Meumann (1896). There were six series of experiments involving, as was normal at the time, the judgements of single participants, or “observers”. A detailed discussion of all the findings in this article, which is more than 120 pages long, is beyond the scope of the present review, but we will mention only some of Meumann’s most striking results. In the first series of experiments, empty intervals (i.e. bounded by brief auditory stimuli) were contrasted with intervals with a single filler, or three. The technique used involved finding the smallest difference between the divided and the empty interval so that the empty interval was perceived as either longer than the divided interval, or shorter than it. It is important to note that in Meumann’s experiments the “longer” and “shorter” judgements came from different experimental series, so the “shorter” and “longer” judgements were not based on the same trials.

Data from a single participant are shown in Figure 1. The upper panel shows the differences between the stimuli that resulted in them being perceived as either longer or shorter. The lower panel shows the differences only for the “longer” judgements, this time coming from conditions with a single dividing stimulus or three. As in Hall and Jastrow (1886), the data suggest that the perceived difference between the divided and empty interval decreased as the intervals increased in duration, and that the effect may even invert. The “wearing off” of the effect is clearest in the lower panel, which shows the ratio between the empty and divided interval needed for the judgement that they are equal. Clearly, this declines towards 1.0 (exact equality) as interval length grows. The potential inversion is shown in the upper panel. When the intervals are short, the empty interval is always judged as shorter and some quantity needs to be added to it to produce the longer judgement, but this quantity decreases as the intervals lengthen. In addition, as the durations lengthen, some quantity must be subtracted from the empty interval in order to produce the “shorter” judgement: in other words, the empty interval appears to be longer.

Meumann then explored the idea, common in early timing studies (see Munsterberg, 1894, and Curtis, 1916, for example) that the timing of short intervals was mediated by small muscular movements or breathing. He asked people to vary both finger movements and respiration, but found little effect. A third series of experiments used either visual stimuli (flashes) or brief tactile stimuli, and found results similar to those obtained when auditory stimuli were employed. A fourth series compared fully filled intervals with empty ones, and is discussed later in this review. A fifth series of experiments varied the type and number of dividers used. In general, increasing the number of dividers appeared to make the divided time effect larger, but grouping repeated dividers together rhythmically tended to make the effect slightly smaller.

As will be seen below, later researchers often found results similar to those of Meumann, but with more modern methods and appropriate statistical analysis. Although Hall and Jastrow (1886) deserve historical priority in discovering the divided time effect, Meumann (1896) surely merits an important place in time perception history by virtue of the extent and scope of his investigations.

Roelofs and Zeeman (1951) made the unusual comparison of judgements of the duration of filled stimuli with judgements of divided ones. All stimuli were square patches of light (11.5 x 11.5 cm) projected onto a screen. The divided intervals started and ended with a 177.8 ms projection, with a total duration of 3200 ms. These were then alternated with a continuous projection of a square, the duration of which could vary between 3200 and 1777.8 ms. The divided and continuous stimuli alternated, with a gap between them of 3200 ms, until “the test subject had established his judgement [of their relative durations]”, but how long this took was not reported. Each combination of empty and filled intervals was repeated 20 times. Following this, an identical procedure was used, except that there were 1, 3, or 7 177.8 ms projections equally spaced between the start and end stimuli of the divided interval. In all cases, the participant’s task was to judge which stimulus lasted longer.

The number of participants used is not reported, nor are any statistical analyses given. Instead, a rather complex chain of reasoning was used to describe the results. This involved a previous finding by the authors that the first stimulus in a sequence of two visual stimuli was judged as 20.6% shorter than the second one, and in the present study the divided interval was presented first, so the authors sometimes used the 20.6% figure to "correct" the data obtained.

A comparison of the empty with the divided interval showed that the empty interval was judged as 31.4% shorter than the filled one, and adding a single divider resulted in a 33.4% shortening, but the authors "corrected" both these values by subtracting 20.6%, resulting in a 13.1% shortening with one divider, and 5% with three. When there were seven dividers, the "apparent shortening" was 14.4%, but when the 20.6% was subtracted, the conclusion was that the divided interval was subjectively 6.2% *longer* than the continuous one (Roelofs & Zeeman, 1951, p. 98). A reader may be sceptical about the correction factor used in this article, but one finding that seems clear is that increasing the number of dividers made the divided time subjectively longer.

A very similar result was reported in Buffardi (1971). All durations were either empty or divided, starting and ending with a 96 ms stimulus, with 35 ms stimuli used as dividers within the interval. Stimuli were either auditory (1200 Hz tone), visual (red light) or tactile (stimulation of right index finger). The constant 1056 ms interval was divided into 11 parts, and from zero to five dividers were presented. In some arrangements these occurred in early positions in the interval, in others at the end of the interval, and in others they were evenly distributed throughout the interval. Pairs of durations were presented in an ABBA order, with a 1,920 ms gap between them, and all 91 duration pairs were presented over 108 participants, although it is not clear whether all participants received all the possible pairings. Participants were required to judge which stimulus lasted the longer, and an equal response was not allowed. Buffardi's simplest result is presented in Figure 2 which shows the percentage of times that stimuli with different numbers of dividers were judged as the longer one of the pair. Obviously, this increased as the number of dividers increased. Note, however, that the results come from all comparisons, not just comparisons of divided intervals with empty ones: presumably, that sort of comparison would have produced an even more striking effect.

Figure 2 about here

Buffardi (1971) found no effect of modality on the effects he reported, similarly to Meumann (1896) who likewise found no effect of modality between auditory, visual, and tactile stimuli. Buffardi also found that positioning dividers early in the interval produced longer subjective durations than positioning them late. A more complex study of the positioning of fillers, by Adams (1977), is discussed below.

Both Roelofs and Zeeman (1951), Buffardi (1971), and Meumann (1896) found evidence that increasing the number of events presented during an interval makes the interval seem longer. Schiffman and Bobko (1977) reported a similar result. Intervals of 5, 9, 13, and 17 s were used with a reproduction method. The intervals were filled with 3, 5, 7, or 9 slides, which also varied in familiarity to the participant. There was no effect of familiarity, but a larger number of slides produce longer reproductions, particularly at the shorter durations. But varying the number of elements presented during a time period does not always produce such simple results. Fraisse (1961) required people to reproduce intervals of 5 or 10 s, while viewing the presentation of squares at the rate of 2/s or 6/s. There were no significant differences in reproduced duration even though there were three times the number of squares presented in the 6/s case than the 2/s one. However, in a second experiment, Fraisse varied the rate of presentation of stimuli during the stimulus to be reproduced and during the reproduction. Reproductions were longer for both 5 and 10 s when the rate during the stimulus to be reproduced was 6/s and the rate during the reproduction 2/s, compared with the reverse case. Similar results were found with auditory stimuli, and with a sensori-motor task where people pressed a key at different rates.

In Thomas and Brown (1974) the durations to be timed started and ended with 10 ms 900 Hz tones. The durations were composed of two sets, a “1-sec” set consisting of durations of 750, 1000, 1500, and 1750 ms, and a “5-sec” set consisting of 4500, 5000, 5300, and 5500 ms. The empty intervals consisted only of the start and end tones. There were two filled interval types, one with 3 regularly spaced 10 ms 480 Hz clicks between the start and end tones, the other with 3 randomly spaced tones of the same sort. Participants reproduced the duration of the stimulus event by pressing a button, and received the 10 presentations of each of the 12 different stimulus types

(4 durations, empty, regular or irregular divisions). Half of the participants received the 1-sec set the other half the 5-sec set. ANOVA found that stimulus duration and stimulus type were highly significant for both sets, with the divided intervals of both types resulting in longer reproductions than the empty intervals, but there was no overall difference between regular and irregular division. Inspection of their Figure 1 (p. 451) suggests that the divided/empty difference was around 500 ms for the 1-sec set and perhaps 600-700 ms for the 5-sec set, representing a much larger percentage difference for the 1-sec than 5-sec sets, so a relatively smaller effect at the longer durations. For the 1-sec set the standard deviation of the divided intervals appeared higher than for the empty ones, although no statistical test was used, but this was not true for the 5-sec set.

Adams (1977) reported three rather complex experiments. The first one compared divided and empty intervals, but in some cases the dividing stimuli (1000 Hz 20 ms tones) could also appear during the interval between trials. Durations of 800, 1000, and 1200 ms were presented, and rating on a category scale (ranging from 1 [short] to 5 [long]) was the response measure. The divided time intervals were judged as longer than the empty ones at all three duration values, but only in the case where the dividing stimuli were presented between the start and end tones of the durations, that is, when the inter-trial interval was silent. In a second experiment, either one or 5 dividers were used, and these occurred either regularly spaced or spaced at random in the interval. The only significant effect was that random placement of one divider produced the largest divided/empty difference. A third experiment varied placement of either one to four dividers and four time intervals from 840 and 1200 ms. Dividers were distributed symmetrically, or clustered either at the beginning or the end of the interval. Generally, symmetrical placement produced the highest category ratings, followed by beginning, followed by end.

Ihle and Wilsoncroft (1983) used intervals ranging from 1 to 60 s. All were bounded by 10 ms 595 Hz tones. There were 3 interval types: empty, filled with 10 ms 950 Hz tones, or filled with “proportional” tones, e.g. 600 ms long at 60 s. In the proportional case there were four equally spaced tones in the interval. It is not clear from the report how many 10 ms tones were used to fill the interval. A verbal estimation method was used. Estimates of the duration of both sorts of divided intervals were significantly longer than for the empty interval only for durations of 1 and 3 s, and only for the 1 s interval did the 10 ms dividers produce the longest estimates.

Horr and DeLuca (2015) compared standard intervals of 1000 ms, with comparisons ranging from 500 to 1500 ms. Empty intervals were defined by 10 ms 1000 Hz tones, and there were two sorts of divided intervals, both containing five 10 ms dividers. One type was what they call “synchronous”, with regular divider spacing, the other was entitled “asynchronous”, with random spacing. A filled interval continuous tone was also used. A pair-comparison method was used. Divided and filled intervals were judged as longer than empty ones, but the divided/empty difference was greater than the filled/empty difference for both sorts of divided intervals, although comparison of filled intervals with both sorts of divided ones was not significant. Intervals with synchronous divided spacing were perceived as longer those with asynchronous spacing. Horr and DeLuca (2015) is a rare divided-time study to report effect size, and they found partial eta-squared values of .49 and .58, indicating a large effect size, for their discrimination and duration comparison tasks respectively, with these numbers coming from a comparison of all the 4 types of stimuli they used, not just divided versus empty intervals.

Divided time studies: summary and theoretical explanations

Most studies find that a divided interval seems to last longer than an empty one, although the effect may reduce, possibly to near zero, as durations lengthen (Meumann, 1896; Ihle & Wollsencroft, 1983). More dividers appear to make durations longer (Buffardi, 1971; Schiffman & Bobko, 1977). Placing dividers early in the divided interval makes the divided intervals longer than with a late placement (Adams, 1977), and synchronous placements may produce a stronger divided time effect than asynchronous ones (Horr & Di Luca, 1985).

How can these effects be explained? Perhaps the simplest explanation of divided time effects comes from Nakajima’s (1987) “supplement hypothesis” (originally proposed by Nakajima in 1979 in an article in Japanese). This states that the subjective duration of an empty interval is related to its real duration plus a constant time of around 80 ms. This idea accounts for a number of results. Firstly, an interval with a single divider should be perceived as longer than an empty interval of the same duration, because of the “supplement” of 80 ms added to the divided time interval. Secondly, increasing the number of dividers should increase the divided time effect, as each divider provides an additional supplement of 80 ms. Thirdly, the divided time effect with a single divider should be easier to observe at short intervals than long ones, as the “supplement“ becomes a smaller and smaller proportion of the time intervals used and thus more likely to be washed out by general timing variance, which will increase as intervals lengthen. As noted above, all these effects are present in data. Ten Hoopen et al. (2008) furthermore show that Nakajima’s proposal can account quantitatively for data from Buffardi (1971), a sample of which is shown in our Figure 2.

Although Nakajima’s simple idea can predict the divided time effect itself, and the effect of increasing the number of dividers, it cannot deal with any effects of placement of dividers within the intervals: only the number of dividers makes any difference. To address potential grouping effects, something else is needed, and a plausible candidate is the *weighted sum* model of Matthews (2013). Matthews conducted experiments which procedurally differed from those discussed above, in that his stimuli were composed of sequences of continuous tones, so they were what might be described as “divided filled intervals”. For example, a sequence could consist of tones of different pitches which divided an interval into equal segments (a regular spacing), accelerated (so the lengths of different pitches became shorter as the interval progressed), or decelerated (the reverse). These conditions resemble, but are not exactly the same as, divided time intervals with regular divider spacing, or clustering of dividers at the beginning (decelerated) or the end (accelerated).

A complete discussion of Matthews’ data and model is beyond the scope of the present article, and we will give just a brief introduction to it here. The model basically rests on two principles. The first is the growth of subjective time as a time interval passes is a negatively accelerated function of duration (in fact, he used a power function). The total subjective duration of a divided interval is the sum of the subjective durations of its constituent parts, so with a negatively accelerated growth function, this is sufficient to account for the simple form of the divided time effect. For example, suppose we contrast an empty interval of length *t*, with a divided interval with a single divider half-way through, making the two components each of duration *t/2*. Given that the growth function is negatively accelerated, the sum of two segments each of length *t/2* is greater than *t*, so a divided-time effect occurs. This can be extended simply to more dividers, so adding more dividers makes the interval seem longer the more that are added, as in Buffardi (1971), see our Figure 2.

Matthews also had a second principle which is needed to account for spacing effects, or what was acceleration and deceleration in his experiments. This is that the weight of an interval in the sum is determined by how long ago the interval ended, and he assumed exponential decay over time. Along with the negatively accelerated subjective time growth function, this second principle can produce a range of results, and we briefly describe some of them here. For example, if we have a single divider early in the interval, the short initial segment may have decayed substantially by the end of the whole stimulus, so the subjective duration of the whole duration will be determined principally by the longer second segment, which is not subject to any decay. In contrast, if the single divider is towards the end of the interval, the longer first segment may be subjected to decay, where as the (shorter) final segment is not. This reasoning tends to suggest that placement of dividers early in the interval should tend to make durations seem longer than when the dividers are later, something which is generally found (Adams, 1977).

Adams varied both the placement and number of dividers, and generally found that more dividers reduced the effect of placement, but also the divided-time effect itself (e.g. Adams, 1977, Figure 3, p. 533). The first of these results seems broadly consistent with Matthews' model. When Adams placed more dividers early in the interval, the length of the final undecayed segment was reduced, so this contributed less to the total subjective duration, thus reducing the advantage of early placement of dividers.

The decay process may also provide a different explanation of why the divided-time effect diminishes as intervals lengthen. If we consider a condition with a single centrally-placed divider, as the interval lengthens the initial component's contribution will be subject to decay, which will be more pronounced as the interval lengthens, so the effect of the divider lessens as the interval gets bigger, something reported in a number of articles (e.g. Meumann, 1896, see our Figure 1, also Ihle & Wilsoncroft, 1983). Another effect predicted by Matthews's model is that regular spacing of fillers will produce longer subjective durations than irregular spacing, and this has also been found (Adams, 1977; Horr & Di Luca, 2015).

However, depending on the exact growth and decay functions, a range of different outcomes is possible (see Matthews, 2013, pp. 267 for discussion), so the model has some flexibility to account for a range of potential results. This weighted-sum model clearly has a lot of potential as an explanation of divided-time effects, although one finding noted in the early literature appears outside its scope. This is the claim that, when an empty and divided interval are presented successively, the magnitude of the divided-time effect depends very strongly on the order of presentation. There appear to be no modern data relevant to this issue, probably because recent researchers have systematically counterbalanced order in tasks involving presentation of two stimuli of different types (e.g., Horr & Di Luca, 2015).

Filled-duration studies

In a filled-duration study, one time-interval is completely filled by a stimulus or some other activity, and the duration of this filled interval is, most commonly, contrasted with that of an empty interval, started and ended with brief stimuli. Many different sorts of stimuli and behaviours have been used to “fill” the filled interval, so studies of this type are very heterogeneous and it may be difficult to compare one with another in a simple way. It is perhaps surprising that what would seem to be the simplest sort of comparison, for example, contrasting a continuous tone with the same interval started and ended with brief clicks, has been infrequent, although Wearden et al., 2007, provide an example of this. Some “fillers” have been much more elaborate, for example, writing instructions displayed in a mirror, taking dictation (Gullikson, 1927), or even listening to “two of the most amusing pages from Mark Twain’s *Huckleberry Finn*” (Swift & McGeogh, 1925). An obvious problem with such elaborate filler material is that it may result in a division of attention between timing the interval and processing the material in it, resulting in what is in effect a “dual-task” procedure. Most research (briefly discussed in Wearden 2016) suggests that dividing attention between a timing and non-timing process makes duration judgements shorter, and this may be a factor in some filled-duration studies, as will be seen below.

To facilitate exposition, we will divide discussion of the studies into three sections. The first discusses studies which report what we judge to be simple fillers (like continuous tones or continuous visual stimuli), the second focusses on more complex fillers, and the third is a section discussing experiments which have compared the discriminability of filled and unfilled intervals rather than their duration.

Simple filler effects

Probably the earliest comparison of a completely filled interval with an empty one comes from Meumann's (1896) experimental series 4. Figure 3 shows one of his results. A continuous sound (filled interval) was contrasted with an interval started and ended with two brief sounds (empty interval). The duration of the empty interval was varied to the point where it was perceived as longer than the filled one. In Figure 3, the ratio of the empty/filled interval is always greater than 1.0 (that is, there is a filled-interval effect) but, as in Meumann's divided time study (see the lower panel of our Figure 1), the effect declined as the time intervals lengthened, and is most marked for intervals less than 1 second long.

Figure 3 about here

Curtis (1916) was mostly concerned with the introspections produced by 5 participants (one of which was E.G. Boring, then an instructor at Cornell University) while comparing pairs of stimulus durations, 1.2 and 1.8 s in length, including some filled/empty comparisons. She reported (p. 14) "When a filled time is followed by an empty time, the empty time is usually underestimated."

Triplett (1931) used reproduction methods, with filled auditory intervals (sometimes of different pitches, although pitch had little effect), click-defined intervals, and also unfilled intervals which were defined by the duration between the offset of a tone and a flash. Intervals varied, but were most commonly 1.5 s. No experiment presented more than one sort of stimulus on a single trial, so the different trial types were never directly compared. Different participants could receive different stimuli and different numbers of trials, but Triplett's Table XIII (p. 245) summarizes the results. 11 of 17 participants reproduced the filled intervals as longer than the empty ones, for 11 of 15 the filled interval was judged longer than the click-defined interval, and for 10 of 15 the unfilled interval was judged longer than the click-defined one. Overall, then, Triplett confirmed a weak filled-duration effect with auditory stimuli.

Goldfarb and Goldstone (1963) used filled visual intervals, or empty intervals bounded by brief visual stimuli. Interval durations varied, and the task was to judge whether each interval was more or less than one second. The durations corresponding to one clock second were calculated, and median values were .81 s for the filled interval, and 1.0 s for the empty one, indicating the presence of a filled-duration effect, such that filled intervals were judged as longer than unfilled ones.

Goldstone and Goldfarb (1963) conducted a much more elaborate investigation. Two category rating scales were used, a “Social standard” where the rating varied from 1 (very much less than 1 s) to 9 (very much more than 1 s), and a “subjective standard” ranging from 1 (very very short) to 9 (very very long). Stimulus durations varied from 0.15 to 1.95 s, and the stimuli were either auditory (1000 Hz tones [filled] or clicks [unfilled]) or visual (fluorescent light [filled] or flashes [unfilled]). The order of presentation of filled and unfilled durations was also varied.

The first experiment found that auditory filled intervals were rated as longer than unfilled ones, but this was not true of visual intervals. Furthermore, the auditory filled/unfilled effect was only found when the filled interval came after the unfilled one, and only when the “Social standard” measure was used. A second experiment with a slightly different design replicated all these findings.

Craig (1973) employed an unusual procedure involving presenting two successive stimuli of equal duration with an unfilled gap between them. The participant’s task was to adjust the gap so that it equalled the duration of the first stimulus. He found that the length of the gap considerably overestimated the target stimulus duration, when the stimuli were vibrotactile, auditory, or visual, indicating that the unfilled gap needed to be considerably longer than the target stimulus to be perceived as equal in duration to it. The duration of the target stimulus was varied over values of from 100 to 1200 ms, and it appeared that the additional difference needed to make the gap subjectively equal to the target was constant as the target varied, although the difference varied with modality, being 596, 637, and 436 ms for vibrotactile, auditory, and visual stimuli, respectively.

Wearden, Norton, Martin, and Montford-Bebb (2007) compared filled and empty auditory intervals. Their first experiment used a temporal generalization method (Wearden, 1992), and found that their filled intervals (500 Hz tones) were judged as longer than their unfilled ones (bounded by 10 ms clicks). Durations ranged from 200 to 800 ms. A second experiment used stimuli ranging from 77 to 1181 ms in duration, and a verbal estimation method. There were two sorts of filled/unfilled comparisons. One involved comparison of tones with the same durations bounded by clicks, the second a comparison of filled intervals with gaps in tones. In both cases, very marked filled-duration effects were found, with the filled intervals being judged as 40-50% longer than the unfilled ones at the longest stimulus durations. In general, the filled/unfilled differences for both click and gap comparisons increased as the intervals lengthened: that is, the filled/unfilled difference was manifested as a slope difference when mean verbal estimates were plotted against stimulus duration. The increasing size of the filled/unfilled effect as time intervals became longer contrasts with the results of Meumann (1896), shown in our Figure 3. This also contrasts with the filled/unfilled effects obtained by Craig (1973) which appeared constant as the time intervals to be judged varied. In Wearden et al,'s study, the standard deviations of filled and unfilled estimates of both types did not differ significantly, although coefficients of variation (standard deviation/mean) were much lower for filled intervals, presumably as a result of the mean estimate being longer.

Williams, Yuksel, Stewart, and Jones (2018) conducted a near-exact replication of Wearden et al.'s comparison between filled and click-defined intervals, and obtained the same results. The filled durations were estimated as longer, and the filled/unfilled difference increased as the durations lengthened, a slope effect as in Wearden et al. (2007). Williams et al. reported effect sizes. For the filled versus unfilled estimate measures, the partial eta-squared was .37 for the main filled versus unfilled effect, and .281 for the interaction, both indicating a large effect size. Likewise, when slopes of the filled and unfilled conditions were compared, a Cohen's d of .85 was reported, again consistent with a large effect size.

Hasuo, Nakajima, and Ueda (2011) compared filled and unfilled auditory intervals, using an adjustment method where the second of two stimuli (the comparison, always unfilled) was adjusted so that it subjectively equalled the standard duration, for example a filled interval. Durations were all short, ranging from 20 to 180 ms. The filled-duration effects was not found overall, but subgroups of participants could be identified, with one subgroup judging filled intervals longer than unfilled ones and another, slightly smaller, subgroup judging the unfilled intervals as longer.

A similar study by Hasuo et al. (2014), with slightly longer durations (up to 520 ms) used both an adjustment and a magnitude estimation method. Once again, the filled-duration effect was not clearly found, particularly with the adjustment method, although subgroups could be identified who showed this effect, particularly when the magnitude estimation procedure was used.

These two studies are rare failures to obtain a filled-duration effect with simple fillers (tones). The existence of subgroups of participants who reversed the effect is also unusual. Few other studies report individual differences, but Williams et al. (2018) did, and they found only four people, out of thirty, who judged unfilled intervals longer than filled ones when a verbal estimation method was used. This contrasts with their findings when auditory and visual durations were compared, where a sizeable subgroup judged visual stimuli to last longer than auditory ones, the reversal of the usual result (see Wearden & Jones, 2021, for discussion).

Bratzke, Birngruber, Durst, and Schroter (2017) examined the filled-duration effect with auditory stimuli, using a reproduction method in two experiments. A novelty of their procedure was that the reproduction itself could either be “filled” (holding down a button continuously), or “empty” (involving two presses starting and ending the reproduction). Intervals from 400 to 1200 ms were used. In both of their experiments, a substantial filled-duration effect was obtained overall, with partial eta-squared values of .62 and .70, indicating a large effect size. There were also effects of reproduction method, with the filled-duration effect being 77 ms larger for empty than filled reproductions. In both experiments, Bratzke et al. reported that coefficients of variation of the reproductions (standard deviation/mean) did not differ significantly as a function of whether or not the stimulus or the reproduction was filled or unfilled. At first sight, this suggests no differences in variability between filled and unfilled conditions. However, this interpretation is complicated by the difference in mean reproduction reported, as the coefficient of variation is a relative, rather than absolute, measure of variability. If condition A has a higher mean than condition B, but the same coefficient of variation, then one possible conclusion is that the absolute variability (for example, as assessed by the standard deviation) must have been greater for condition A. In the case of the Bratzke et al. experiments, this would suggest greater absolute variability for filled rather than unfilled intervals, but this remains speculative, given that standard deviations are not presented.

Complex filler effects

A number of early studies used complex events to fill the filled interval. For example, Spencer (1921) compared reproduction by 10 participants of an empty 30 s interval with the same durations reproduced with “interpolations” of various sorts: the participant or the experimenter reading poetry or prose, or dictation. The empty interval was on average overestimated (at 38.23 s). The readings produced little change (from 37.25 to 40.89 s), but dictation reduced the reproduction value to 32.41 s. Two other studies (Swift & McGeogh, 1925, and Gullikson, 1927) likewise used complex tasks to fill intervals, and estimates of elapsed time performing these was contrasted with “doing nothing” or “rest”. In Swift and McGeogh’s first experiment, people copied nonsense syllables, “dry material” from a catalogue of laboratory apparatus, or passages from Huckleberry Finn. In a final condition they listened to “amusing pages” from the same novel being read to them. Intervals of 30 s, 1, 2, and 5 minutes were used, with verbal estimates of duration being provided. Judgements were contrasted with a period of rest, without interpolated activity. When the judgements from the different activities were averaged together, duration judgements resulting from filling the interval with activity did not appear to differ from those obtained with no activity. A second study used a 10 minute interval, but no rest condition, and found that listening to Huckleberry Finn lengthened estimates compared with copying other material.

Gullikson (1927), in a rather similar study, contrasted a 200 s period of rest with a variety of other activities, including listening to a slow or fast metronome, carrying out arithmetical divisions, inducing fatigue by holding an arm in the air, pain by pressing on a pin, or receiving dictation from a Psychology text. All filled periods were judged as shorter than the rest period, sometimes markedly shorter. For example, the rest period was on average judged to last 241.7 s, but the intervals filled with dictation and division 174.6 and 168.9 s, respectively.

Whitely and Anderson (1930) compared judgements of the durations of three types of intervals, empty click-defined intervals, an interval filled with a non-rhythmical buzzer, and an interval filled with music (Johann Strauss’s *Radetsky March*). In their first experiment intervals of 15, 30, and 45 s were used, with a verbal estimation method. The interval filled with music was judged shortest, with little difference between the other two. In a second experiment pairs of intervals were presented, and the task was to judge if the second was longer or shorter than the first, or equal to it. Durations were 5, 10, and 15 s. The same tendency was present in this experiment, with the interval filled with music judged as shorter than the other two, although the effect was clearest at the 10 and 15 s durations.

Warm, Smith, and Caldwell (1967) compared judgements of the duration of intervals of 6, 12, 24, and 48 s, which were filled by gripping a dynamometer at 0, 10, 20, or 40% of the participant’s maximum grip strength. Verbal estimation and reproduction methods were used for the time judgement. Variation in grip force had small effects, but the zero load always produced the longest judgements, particularly when the 48 s interval was used. Comparison of verbal estimation and reproduction judgements showed that verbal estimation tracked the longer intervals more closely than reproduction.

Two more recent studies of filler effects, by Gomez and Robertson (1979) and Robertson and Gomez (1980), used very different duration ranges and filler types compared with the earlier studies. In Gomez and Robertson (1979) two different duration sets were used: a 15 ms set (values of 15, 30, and 45 ms) and a 70 ms set (values of 15, 85, and 155 ms). Large or small visual figures were presented during the time intervals, in a rather complex design involving both within- and between-subject effects. In their first experiment, they found an effect of stimulus size, with larger size being judged as longer, but only for the 70 ms set. In a second experiment, they used durations of 155, 170, and 185 ms, and found an effect of stimulus size. In a second study, Robertson and Gomez (1980) varied visual stimulus size and also “goodness” (e.g., symmetry versus asymmetry) with 15, 30, and 45 ms presentations. There were small effects of both size and goodness, with larger and “poorer” figures being judged as longer, but this was only a “slight tendency” which was not statistically significant.

Discrimination studies

In what we call discrimination studies the principal focus of interest is how discriminable filled and unfilled intervals are, rather than whether one type is subjectively longer than another one. Some of these studies used an adaptive technique, which can be illustrated with reference to the work of Rammsayer and

Lima (1991). These authors compared short 1000 Hz tones, with intervals defined by 3 ms clicks. Pairs of durations were presented and the participant's task was to decide which lasted longer. One stimulus had a constant duration (in this case 50 ms), and the other was variable. On the first few trials, the durations of the stimuli were substantially different, with the variable stimulus being clearly longer, and a correct response, which usually occurs on the first few trials, made the difference between the durations smaller on the next trial, that is, the variable duration was reduced on the next trial, by some step value. An incorrect response increased the duration of the variable stimulus by some step value. In this sort of procedure, the step values are often reduced as the experiment proceeds. The mean duration difference between the stimuli on the last 20 trials of the experiment defines the difference threshold for discrimination for that stimulus. Note that in this procedure filled and unfilled durations are not directly compared: both stimuli on each trial are either filled or unfilled.

Rammsayer and Lima found that the difference thresholds were 6.6 ms for filled intervals but 20.9 ms for unfilled ones, so the unfilled stimuli were much more difficult to discriminate (i.e. the difference between them needs to be bigger for a correct response to occur) than the filled ones. A second experiment by Rammsayer and Lima (1991) imposed a cognitive load at the same time as the discrimination task was being performed. This increased thresholds for both stimulus types, but there was still a marked filled/unfilled difference.

However, work by Grondin (1993) produced very different results. The first two experiments used a single stimulus method, where one stimulus was presented on each trial, which was either short or long, and the task was to judge which. In Experiment 1 filled and empty auditory intervals were used, with short and long values of 241 and 259 ms, respectively. In addition, filled and empty visual intervals with durations of 225 and 275 ms were employed. In both cases, performance was more accurate with empty rather than filled intervals. A second experiment used intervals around 50 ms, as in Rammsayer and Lima (1991), and found no difference between discriminability of filled and unfilled intervals when the stimuli were auditory, but a difference favouring the empty intervals when they were visual. A third experiment compared the single stimulus method with a forced choice procedure involving the presentation of two stimuli on each trial, with all stimuli being auditory.

Both methods produced higher discriminative accuracy for empty intervals than filled ones. A fourth experiment used an adaptive procedure like that of Rammsayer and Lima (1991) with stimuli around 50 or 250 ms long, and in both auditory and visual modalities. In both cases, filled intervals produced larger thresholds (i.e., were less discriminable) than empty ones when they were visual, but not when auditory.

Rammsayer (2010) used a very similar method to Rammsayer and Lima (1991). In the first experiment, the duration of the constant stimulus was either 50 or 1000 ms. A slightly different measure of performance (a Weber fraction) was used, with smaller values indicating better discrimination performance. As in the studies above which used an adaptive procedure, data from the last 20 trials of the discrimination series were used. The values were then converted into estimates of the 25% and 75% correct estimates, although the article does not explain exactly how this was done, and the difference between these values was used to define a *difference limen*. The difference limen divided by the standard duration produced a Weber fraction. Weber fractions were significantly different (filled 0.15, unfilled 0.33) when the standard duration was 50 ms, but not when it was 1000 ms (filled 0.12, unfilled 0.13). In a second experiment Rammsayer varied the duration of the constant stimulus over values of 50, 100, 200, 400, 600, 800, and 1000 ms. There was a significant effect of interval type overall (with filled producing smaller Weber fractions than unfilled), but the effect was limited to the durations of 100 ms or less. Obviously, this suggests that the filled/unfilled interval discrimination difference is restricted to very short durations, but Williams et al. (2018), using an almost identical method with a standard duration of 700 ms, found very large filled/unfilled threshold differences, with the filled threshold value (107.33 ms) being just slightly over half that obtained with unfilled intervals (213.62 ms).

We measured data in Rammsayer's (2010) Figure 1 (p. 1595) to determine effect size, via Cohen's d, for the 50 and 100 ms conditions where significant filled/unfilled effects were obtained. Values were .82 at 50 ms and .54 at 100 ms, indicating moderate to large effect sizes.

Horr and Di Luca (2015) used a different discrimination technique, which involved pairs of stimulus durations (1000 ms standard, comparisons from 500 to 1500 ms, and the task was to decide which of the two stimuli was the longer one. Stimuli were filled intervals (1000 Hz tones), empty intervals started and ended by clicks, and two sorts of divided intervals, one with five equally spaced 10 ms tones in the interval and another with the same 5 tones randomly spaced. For each comparison, the stimuli were always of the same type. In terms of just-noticeable differences, filled intervals were discriminated best, followed by intervals with equal filler spacing, followed by empty ones, with intervals with random filler spacing being discriminated most poorly.

Discrimination methods may provide information about the relative variability of filled and unfilled intervals. Variability measures are rarely reported in studies of filled and unfilled intervals (but see Wearden et al., 2007), or only mentioned in passing (Thomas & Brown, 1974), so data are scarce. If we assume that the principal cause of differences in discriminability of stimuli in discrimination tasks is variability of their representations, then the evidence from some studies suggests that unfilled intervals have more noisy representations than filled ones, the results of Grondin (1993) being the exception. If filled intervals are more precisely represented from one trial to another, then their representations will overlap less and they will be easier to discriminate from one another than stimuli with greater variability. However, this is an assumption and it may be that differences in discriminability between filled and unfilled intervals have different causes, or causes in addition to differences in variability. Wearden et al. (2007) in fact found that standard deviations of verbal estimates of filled and unfilled intervals did not differ significantly, although coefficients of variation (standard deviation/mean) differed very markedly, presumably as a result of differences in means. This lack of an effect contrasts with the implication derived from the majority of discrimination studies, which favour considerably higher variability for unfilled compared with filled intervals. It is unclear why this discrepancy exists, although Wearden’s (2015) model of verbal estimation implies that variability data from verbal estimation procedures can, at least in some cases, misrepresent the “raw” underlying variability of the stimuli judged. However, the misrepresentation is unlikely to completely mask the large differences in variability implied by the results of most discrimination experiments.

Filled-duration studies: Summary and theoretical explanations

When simple fillers are used (such as continuous tones), duration judgements are usually longer for filled intervals than unfilled ones (the articles by Hasuo and colleagues are an exception), but when two successive stimuli are presented for comparison the effect may depend on the order of presentation. When verbal estimation of short auditory stimuli is used, the filled/unfilled difference increases as the durations become longer, but research by Meumann (1896, see our Figure 3) suggests that the effect diminishes at longer durations, albeit with another method. Complex fillers often produce no effect or the reverse effect, with the filled interval being judged as shorter, although experiments with complex fillers tend to use much longer durations than those with simple fillers, so there may be a confounding effect of duration when studies of different types are compared. Most studies find that durations of simply filled intervals are easier to discriminate than durations of unfilled intervals, although the difference may disappear at intervals longer than a few hundred milliseconds. The few studies that have reported effect sizes find the filled/unfilled difference to be associated with at least a moderate effect size, and usually a large one.

Given that complex and simple fillers often appear to generate the opposite pattern of results, it is perhaps sensible to treat them separately from the theoretical point of view. Complex fillers resemble dual-task procedures, for example, where a timing task is performed concurrently with another task (usually not involving timing). The general consensus is that the non-timing task distracts attention away from timing, and thus shortens time representations (e.g., Brown, 2008; Fortin & Rousseau, 1987). This seems the most likely explanation of the early research involving complex fillers, which find that such fillers shorten perceived durations compared with conditions without a filler, in the majority of cases.

One explanation of the effect of simple fillers comes from Grondin (1993), his *internal marker* hypothesis. This suggests that detecting the offset of a stimulus takes longer than detecting its onset. This means that the subjective duration of an empty interval is based on the time between detection of the onset of the first stimulus and the onset of the second one, whereas the duration of a filled interval is based on the time between its onset and detection of its offset, which is, according to this hypothesis, a longer period of time. This implies that the difference between the subjective durations of filled and empty intervals of the same real length should be a constant amount. As reviewed above, some data (Craig, 1973) support this idea, whereas others (Bratzke et al., 2017; Wearden et al., 2007; Williams et al., 2018) find that the filled/unfilled difference grows with stimulus duration. Although Craig's data are consistent with Grondin's idea, if we suppose that the difference between the filled and unfilled durations is simply down to the delay in judging offset of the filled interval, then the differences between the subjective durations that Craig found (between 436 and 637 ms, depending on stimulus modality) may seem implausibly large.

Grondin (1993) also used his internal marker account as an explanation for his finding that empty intervals are better discriminated than filled ones (but see Rammsayer & Lima, 1991, and Williams et al., 2018). The explanation used the idea that filled intervals would be on average judged as longer than unfilled ones of the same duration. Given that timing variance increases with duration, this would lead to higher timing variance in filled intervals, which would make them more difficult to discriminate than unfilled ones.

Another explanation of the effect of simple fillers comes from Wearden et al. (2007) who used an idea derived from internal clock, or pacemaker-counter, theories. This supposes that time representations arise from a pacemaker-like mechanism, which feeds into a counter or accumulator. If two different sorts of stimuli are timed (such as filled and empty intervals), the pacemaker may run at different speeds in the different cases, thus the duration will seem longer for one stimulus type than another. In the simplest mathematical form of pacemaker/counter systems (e.g. Wearden, Edwards, Fakhri, & Percival, 1998) pacemaker rate multiplies real duration, so the perceived difference between the conditions with different pacemaker speeds will increase as the intervals lengthen, exactly the effect found with verbal estimates by Wearden et al. (2007) and Williams et al. (2018), assuming that the pacemaker runs faster for filled rather than unfilled intervals.

Although the results from Williams et al. (2018) were in accord with this idea, they questioned it on rather complex grounds. In their study, as mentioned above, they performed verbal estimate and discrimination tasks with the same participants, and the same filled and unfilled intervals, so within-subject comparisons on the two tasks could be made. Their argument proceeded in two stages. Firstly, if pacemaker rate is responsible for the difference, then individuals with faster pacemakers should produce higher slopes when their estimates on the verbal estimation task were regressed against real duration. Secondly, individuals with faster pacemakers should show more sensitive timing on the discrimination task. These two assumptions lead to the conclusion that there should be a strong correlation between estimation slope and threshold from the discrimination task, whereas in fact such a correlation was not obtained.

There are some problems this argument, even if we assume that faster pacemakers would result in higher estimation slopes. The first is the assumption that faster pacemakers should give rise to more sensitive timing on the discrimination task. It is true that if the pacemaker is of a Poisson type, higher pacemaker rates produce relatively more sensitive timing, and thus should lead to lower thresholds. But the pacemaker may not operate as a Poisson mechanism: pacemakers can be devised where the rate and variability of the pacemaker are largely decoupled (e.g., Gibbon, 1992). If the pacemaker is not a Poisson emitter, then Williams et al.'s argument is weakened. Another issue is that pacemaker rates, even if a partial explanation of the filled/unfilled difference found, may not be the *only* source of difference between filled and unfilled intervals. Onsets and offsets of the stimuli may be more difficult to detect in one case than another, thus increasing variability. It might also be that, in the discrimination task, issues of confusion or interference arise. For example, when two click-defined durations with an interstimulus interval are compared, people may confuse the empty interstimulus interval with the click-defined durations, whereas this confusion may be less common when filled intervals, with a silent interstimulus interval between them, are compared.

Conclusions

Our review suggests that there are “filled-interval illusions” in the plural rather than a single effect. Although divided time and filled-duration studies produce a similar result, that the filled or divided interval appears longer than an empty one, they are procedurally distinct, focus on different aspects of their respective effects, and produce data which seem to require quite different theoretical explanations. So far as we know, no study has used a verbal estimation procedure, like that employed by Wearden et al. (2007), to contrast divided time intervals with empty ones. Given that this method seems sensitive to changes in the filled/difference as intervals lengthen, which provides evidence for the pacemaker-based account of the effect, verbal estimation might be helpful in confirming that the divided-time effect is based on something other than pacemaker rate, as we concluded in an earlier section.

A further distinction that may be useful is between simple and complex fillers in filled-duration studies. Studies with complex fillers are procedurally distinct from those using simple fillers, usually employ much longer durations, often obtain no effect or a shortening of perceived duration compared with empty intervals, in contrast to the lengthening usually obtained with simple fillers, and their results have different theoretical explanations from those using simple fillers.

In spite of the considerable number of experiments conducted under the general rubric of the “filled duration illusion”, their very heterogeneity means that some issues which are mentioned from time to time in the literature have not yet received any kind of systematic exploration, so definite conclusions are difficult to draw. One unresolved question is about variability. Do unfilled intervals give rise to more variable time representations than filled intervals? At present, evidence from most discrimination studies suggests a positive answer, whereas data from verbal estimation implies no difference. The alleged dependence of filled-duration effects on the order in which stimuli are presented when two are successively delivered likewise merits further investigation as, if true, any such effect would pose a challenge to models which otherwise seem to fit data well in many respects. What little evidence there is (from Goldstone & Goldfarb, 1963) suggests that order effects may be in the opposite direction for divided-time procedures (bigger effect when divided-interval comes first) and for filled-duration procedures (bigger effect when filled interval comes second). Another issue of interest is the apparent dependence of both divided-time and filled-duration effects on the length of the interval judged. All these questions seem to be suitable topics for further research into the venerable question, which has occupied psychologists for more than 120 years, of the nature of “filled duration illusions”.

References

Adams, R.D. (1977). Intervening stimulus effects on category judgments of duration. *Perception and Psychophysics*, *21*, 527-534.

Bratzke, D., Birngruber, T., Durst, M., & Schroter, H. (2017). Filled and empty motor reproductions of filled and empty intervals: Is there also a filled-reproduction illusion. *Attention, Perception, and Psychophysics*, *79*. 2143-2152.

Brown, S.W. (2008). Time and attention: Review of the literature. In S. Grondin (Ed.) *Psychology of Time*, pp. 111-138. Bingley: Emerald.

Craig, J.C. (1973). A constant error in the perception of brief temporal intervals. *Perception and Psychophysics*, *13*, 99-104.

Curtis, J.N. (1916). Duration and the temporal judgment. *American Journal of Psychology*, *27*, 1-46.

Fortin, C., & Rousseau, R. (1987). Time estimation as an index of processing demand in memory search. *Perception and Psychophysics*, *42*, 377-382.

Fraisse, P. (1961). Influence de la durée et de la fréquence des changements sur l'estimation du temps. *L'Année Psychologique*, *61*, 325-339.

Gibbon, J. (1992). Ubiquity of scalar timing with a Poisson clock. *Journal of Mathematical Psychology*, *36*, 283-293.

Goldfarb, J.S., & Goldstone, S. (1963). Time judgment: A comparison of filled and unfilled durations. *Perceptual and Motor Skills*, *16*, 376.

Goldstone, S., & Goldfarb, J.S. (1963). Judgment of filled and unfilled intervals: intersensory factors. *Perceptual and Motor Skills*, *17*, 763-774.

Gomez, L.M., & Robertson, L.C. (1979). The filled duration illusion: The function of temporal and nontemporal set. *Perception and Psychophysics*, *25*, 432-438.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception and Psychophysics*, *54*, 383-394.

Gullikson, H. (1927). The influence of occupation upon the perception of time. *Journal of Experimental Psychology*, *10*, 52-59.

Hall, G.S., & Jastrow, G. (1886). Studies of rhythm. *Mind*, *11*, 55-62.

Hasuo, E., Nakajima, Y., & Ueda, K. (2011). Does the filled duration illusion occur for very short time intervals? *Acoustical Science and Technology*, *32*, 82-85.

Hasuo, E., Nakajima, Y., Tomimatsu, E., Grondin, S., & Ueda, K. (2014). The occurrence of the filled duration illusion: A comparison of the method of adjustment with the method of magnitude estimation. *Acta Psychologica*, *147*, 111-121.

Horr, N.K., & Di Luca, M. (2015). Filling the blanks in temporal intervals: the type of filling influences perceived duration and discrimination performance. *Frontiers in Psychology*, article 114.

Ihle, R.C., & Wilsoncroft, W.E. (1983). The filled duration illusion: Limits of duration of interval and auditory fillers. *Perceptual and Motor Skills*, *56*, 655-660.

Matthews, W.J. (2013). How does sequence structure affect the judgment of time? Exploring a weighted sum of segments model. *Cognitive Psychology*, *66*, 259-282.

Meumann, E. (1896). Beitrage zur Psychologie des Zeitbewussteins. *Philosophische Studien*, *12*, 127-254.

Munsterberg, H. (1894). Studies from the Harvard Psychological Laboratory I. *Psychological Review*, *1*, 34-60.

Nakajima, Y. (1987). A model of empty duration perception. *Perception*, *16*, 485-520.

Rammsayer, T.H., & Lima, S.D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception and Psychophysics, 50*, 565-574.

Rammsayer, T.H. (2010). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Attention, Perception, and Psychophysics*, *72*, 1591-1600.

Roberston, L.C., & Gomez, L.M. (1980). Figural versus configural effects in the filled duration illusion. *Perception and Psychophysics*, *27*, 111-116.

Roelofs, C. O., and Zeeman, W.P.C. (1951/1952). Influence of different sequences of optical stimuli on the estimation of duration of a given interval of time. *Acta Psychologica*, *8*, 89-128.

Schiffman, H.R., & Bobko, D.J. (1977). The role of number and familiarity of stimuli in the perception of brief temporal intervals. *American Journal of Psychology*, *90*, 85-93.

Spencer, L.T. (1921). An experiment in time estimation using different interpolations. *American Journal of Psychology*, *32*, 557-562.

Ten Hoopen, G., Miyauchi, R., & Nakajima, Y. (2008). Time-based illusions in the auditory mode. In S. Grondin (Ed.) *Psychology of Time*, pp. 139-187. Bingley: Emerald.

Thomas, E.A.C., & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception and Psychophysics*, *16*, 449-458.

Triplett, D. (1931). The relation between the physical pattern and reproduction of short temporal intervals: A study in the perception of filled and unfilled time. *Psychological Monographs*, *41*, 201-265.

Warm, J.L., Smith, P.T., & Caldwell, L.S. (1967). Effects of induced muscle tension on judgment of time. *Perceptual and Motor Skills*, *25*, 153-160.

Wearden, J.H. (2015). Mission: Impossible? Modelling the verbal estimation of duration. *Timing and Time Perception*, *3*, 223-245.

Wearden, J.H. (2016). *The Psychology of Time Perception*. London: Palgrave MacMillan.

Wearden, J.H., & Jones, L.A. (2021). Judgements of the duration of auditory and visual stimuli. *Timing and Time Perception, 9,* 199-224.

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., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled duration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 716-729.

Whitely, P.L., & Anderson, J.C. (1930). The influence of two different interpolations upon time estimation. *Journal of General Psychology*, *4*, 391-401.

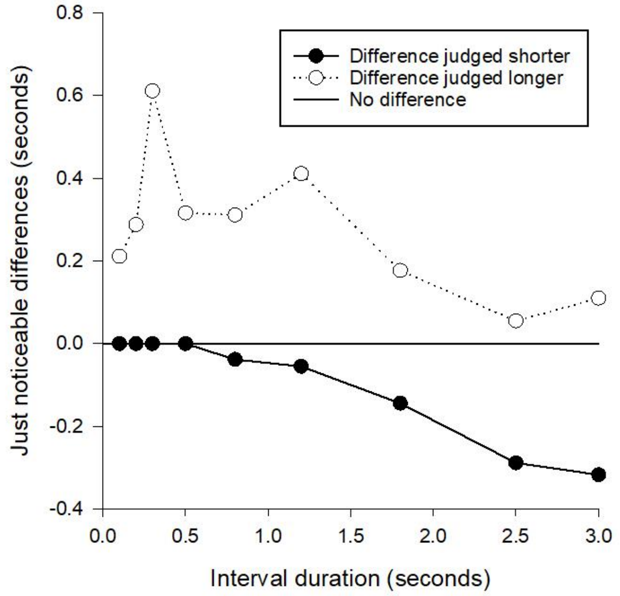
Williams, E. A., Yuksel, E.M., Stewart, A.J., & Jones, L.A. (2018). Modality differences in timing and the filled duration illusion: Testing the pacemaker rate explanation. *Attention, Perception, and Psychophysics*, *81*, 823-845.

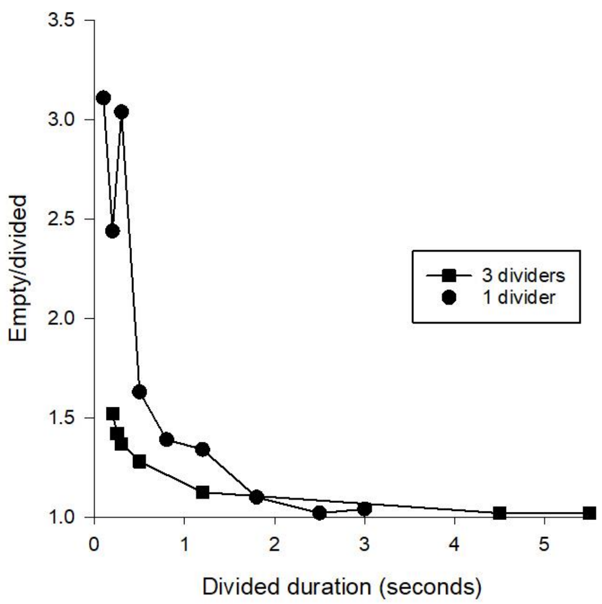
Figure legends

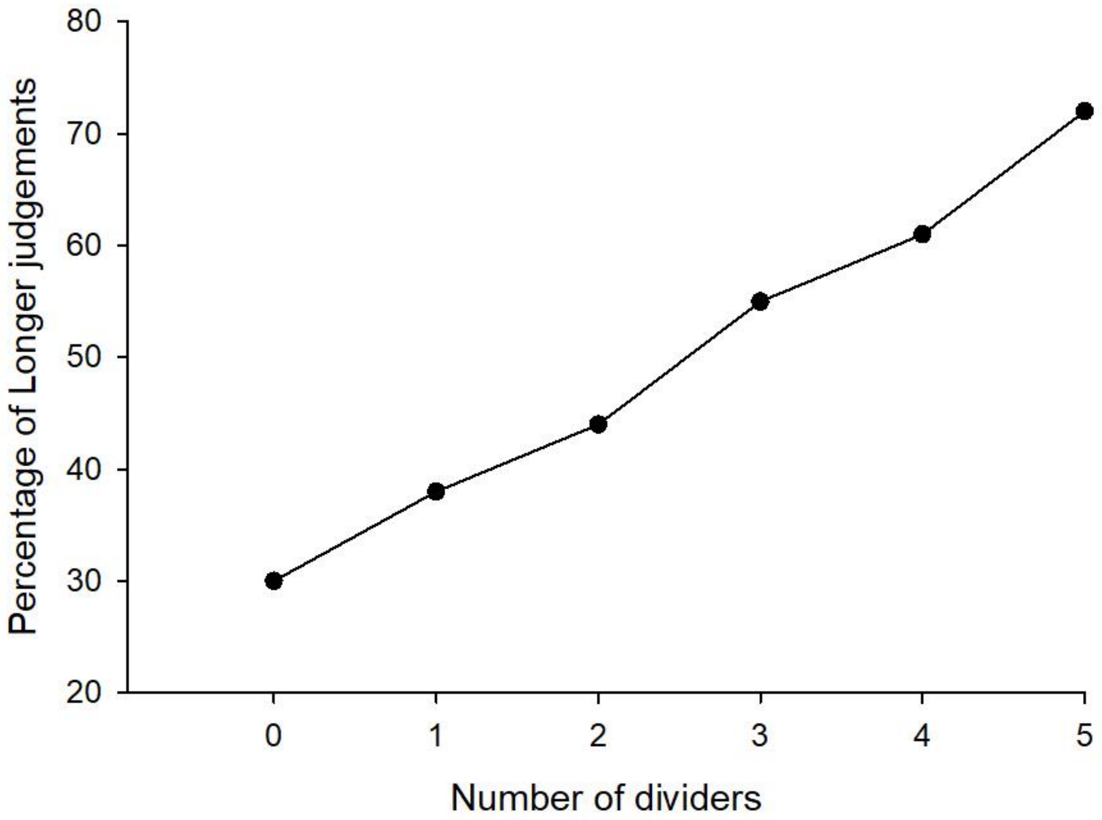
Figure 1: Data from Meumann (1896). Upper panel: Just-noticeable differences for judgements that a comparison stimulus is longer or shorter than a standard one, plotted against stimulus duration (see text for details). Lower panel: Effect of numbers of fillers on duration judgements. The measure is the ratio of the empty to the divided-time interval needed for the empty interval to be judged as longer.

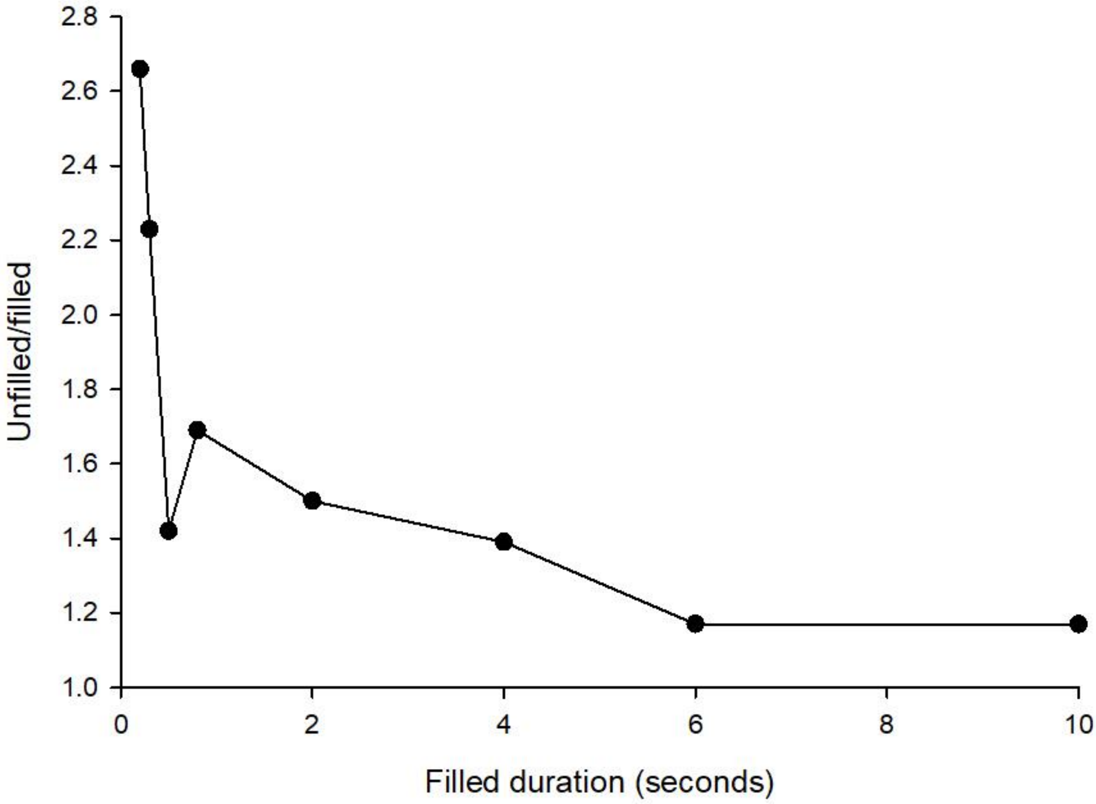
Figure 2: Data from Buffardi (1971). Percentage of longer judgements obtained with a pair comparison procedure, when the number of fillers in a comparison stimulus was varied.

Figure 3: Data from Meumann (1896). Comparison of filled and unfilled intervals as a function of interval duration. Measure shown is the ratio of the unfilled interval to the filled one needed for the unfilled interval to be judged as longer.


Just noticeable differences (seconds)
Interval duration (seconds)
Difference judged shorter
Difference judged longer
No difference
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Empty/divided
Divided duration (seconds)
3 dividers
1 divider
3.5
3.0
2.5
2.0
1.5
1.0
0
1
2
3
4
5


80
70
60
50
40
30
20
Percentage of Longer judgements
0
1
2
3
4
5
Number of dividers


Unfilled/filled
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0
2
4
6
8
10
Filled duration (seconds)