Do Humans Possess an Internal Clock with Scalar Timing Properties?

J. H. WEARDEN

University of Manchester

The question of whether humans possess an internal clock with scalar timing properties was discussed with reference to three sorts of experiments. First, evidence from experiments on interval production was presented which showed that interval production by humans that is not based on chronometric counting exhibited approximate scalar properties as (i) mean times produced increased linearly with real time, and (ii) coefficients of variation of times produced varied unsystematically (in the range 0.10 to 0.16) as the time produced varied. Second, an experiment measuring the intercount units used by subjects in chronometric counting revealed that count units exhibited approximate scalar properties, with their coefficient of variation remaining roughly constant (in the range 0.10 to 0.14) as intercount interval changed. A simple model of chronometric counting based on scalar properties of count units was shown to be compatible with the declining coefficients of variation with increases in interval produced found in the overall time judgments based on chronometric counting. Finally, two experiments with normal humans which were analogs of animal experiments on interval bisection and temporal generalization were discussed. Here, human performance differed from that found in animals in both cases, suggesting that the way humans make judgments about intervals differs systematically from the judgmental processes of animals, but models of behavior consistent with underlying scalar representations of duration could be developed. Overall, all three sorts of experiments produced evidence suggesting that human timing can, in some circumstances, be described well by a scalar timing mechanism. © 1991 Academic Press, Inc.

Recent years have seen an immense growth of interest in temporal regulation of behavior in animals, particularly the possibility that a common mechanism, scalar timing (Gibbon, Church, & Meck, 1984), might underlie performance in a wide variety of different experimental situa-

tions. Scalar timing is discussed in more detail in other articles in this issue; for present purposes it is sufficient to note that it is a quantitatively precise theory in the "internal clock" tradition; that is, temporal regulation of behavior is explained by supposing that animals possess a timing-specific internal clock-like mechanism, which can be started and stopped, and the values produced by which bear orderly relations to real time.

An obvious comparative question which arises is posed by the title of this article. The idea that humans possess a "biological clock," which they can use to perform timing tasks, has a long history, going back at least to Francois (1927) and Hoagland (1935). The purpose of the present article is to discuss evidence that humans possess an internal clock with scalar timing properties, and data from three different types of studies will be considered. These are (i) interval production, (ii) chronometric counting, and (iii) experiments with human subjects which are explicit analogies of those carried out with animals.

As will be seen later, the question of whether humans possess an internal clock with scalar timing properties can be answered, in the present author's opinion, in the affirmative, as a body of evidence can be amassed which supports, or at least does not directly contradict, this idea. However, a caveat is in order. This is that the claim that humans probably do possess a scalar timing mechanism that they share with animals does not imply that all temporal judgments made by humans depend on this mechanism. The most casual perusal of a general review of human timing behavior (e.g., Fraisse, 1964, 1984) will suggest to the reader that the study of human timing is a field of extraordinary complexity, and it is highly likely that different sorts of timing behavior are mediated by different mechanisms. Although scalar timing may be one of these it may not be used in all circumstances and indeed, as will be seen later, perhaps the clearest and simplest evidence for scalar timing in humans comes from experiments in which durations are produced without the involvement of chronometric counting.

INTERVAL PRODUCTION

Experiments on interval production with human subjects require the production of an interval of specified duration (e.g., 0.9 or 10 s), with the interval being defined by two responses, or by the duration of a single response such as a button-press. After each production, feedback, which the subjects may use to alter their performance, may or may not be given. Furthermore, different sorts of feedback may be provided, for example, giving specific information about the clock time produced (e.g., Wearden & McShane, 1988), or merely indicating that some minimum required time value has been exceeded (Zeiler, Scott, & Hoyert, 1987).

Interval production experiments bear an obvious resemblance to certain sorts of reinforcement schedules used with animal subjects, such as dif-
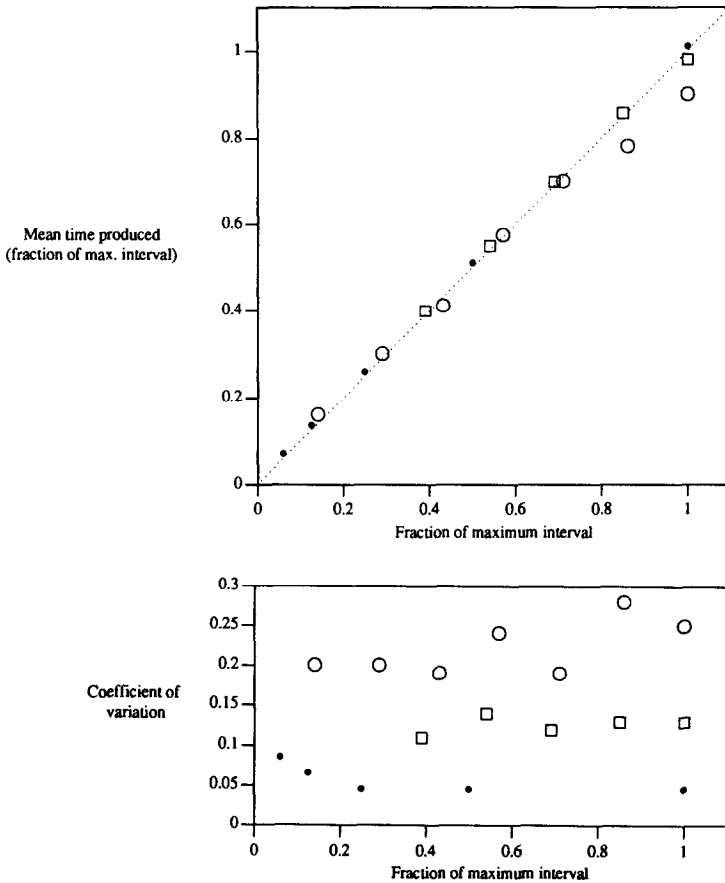
FIG. 1. Data from interval production experiments involving DRL schedules with pigeon subjects emitting a perching response (open circles), humans producing short intervals without chronometric counting (squares), and humans producing longer intervals with chronometric counting (small filled circles). Upper panel: mean interval produced plotted against required interval (or DRL value for pigeons). All values are scaled as proportions of the maximum interval used which was 70 s for pigeons, 1.3 s for short interval production, and 32 s for longer interval production. Lower panel: coefficients of variation from the three studies, plotted against interval (or DRL) value. The points from the chronometric counting experiment were obtained by simple calculation (standard deviation/mean), the other coefficients by curve-fitting.

ferential-reinforcement-of-low-rate (DRL) schedules, where each response exceeding some minimum value is followed by reinforcer delivery. This similarity permits certain animal-human comparisons to be made, and Fig. 1 shows two of these.

Figure 1 presents data from three experiments, one with pigeons, and two with normal adult humans. The experiment with pigeons (Jasselette,

Lejeune, & Wearden, 1990) involved DRL values of 10, 20, 30, 40, 50, 60 and 70 s, with a *perching* response used as the operant. Three birds were run under the schedules described above, and at each DRL value the aggregated relative frequencies of interresponse-times produced were collected. Gaussian curves were fitted to these distributions (cf. Wearden & McShane, 1988), and this analysis yielded a peak location, and a coefficient of variation, of the fitted curve. Following Roberts (1981) we might regard the location of the peak of the interresponse-time frequency distribution to directly reflect the subjects' average representation of the DRL schedule parameter, and the coefficient of variation to represent a measure of sensitivity to time, a type of Weber fraction (Gibbon, 1977).

Figure 1 also shows data from Wearden and McShane (1988). Here, four normal adults produced durations ranging from 0.5 to 1.3 s, and were given feedback, accurate to two decimal places in seconds, after each production. Gaussian curves were fitted to the aggregated relative frequency distributions of times produced, as described above. Finally, Fig. 1 includes data from an experiment on counting-based production of time intervals, conducted in my laboratory in collaboration with I. P. Christensen. The intervals to be produced were 2, 4, 8, 16, and 32 s. Subjects were required to employ chronometric counting, and each production was followed by feedback accurate to two decimal places in seconds. The data collected here were the mean time produced, and the coefficient of variation (standard deviation/mean) of the times produced, from 14 subjects.

To facilitate comparison between experiments using very different time intervals, all time values were expressed as fractions of the maximum interval used (this was 70 s for the pigeon experiment, 1.3 s for the short interval production, and 32 s for the counting-based interval production experiment). The upper panel of Fig. 1 shows mean time produced (or peak location) plotted against interval value. The dotted line drawn in Fig. 1 shows a perfect coincidence of time produced and required time (or DRL value in the pigeon experiment). Obviously, apart from some underestimation by the pigeons at the longest DRL values, mean times produced and required times coincided virtually perfectly. The data in the upper panel of Fig. 1 thus exhibit the first requirement of data that conform to scalar timing theory, namely that mean measures of behavior are accurate reflections of real time (Gibbon, 1977).

The lower panel of Fig. 1 shows coefficients of variation, derived either from Gaussian curves or by simple calculation, from the three studies, plotted against target time to be produced (or DRL) value. A critical requirement of data which conform to scalar timing theory is that coefficients of variation should remain constant (or at least not vary systematically) with changes in the interval timed. As can be seen in the lower panel of Fig. 1, data from the short interval production experiment

(squares) conformed to this requirement almost perfectly, the other two cases more approximately. In data from pigeons (open circles), coefficients of variation remained roughly constant up to DRL 50 s, then increased, perhaps indicating some loss of temporal control at the longer schedule values. In counting-based timing (filled circles), coefficients of variation tended to decrease as the interval produced increased, although this decline was most marked between 2 and 8 s, with little change thereafter. However, the most striking feature of the coefficients of variation from the three different experiments was their very different absolute values. DRL data from pigeons yielded values roughly between 0.20 and 0.30; short interval production values between 0.11 and 0.14; counting-based interval production the lowest values, all less than 0.10, and most below 0.05. In terms of empirical description of behavior all three experiments yielded data in which the mean times produced were approximately accurate, but the variability around the mean was greatest in the animal experiment, next largest when humans produced short intervals (for which chronometric counting was not spontaneously used), and smallest when counting-based timing was used.

An examination of a variety of experiments using DRL and other temporal differentiation schedules with animal subjects suggests that a coefficient of variation value of 0.20 to 0.30 is a common result (Jasselette *et al.*, 1990; Platt, 1979). We might thus suppose that such a value represents a kind of "signature" of a timing mechanism which commonly underlies animal performance on a variety of timing tasks. For humans, on the other hand, short interval production, which does not depend on chronometric counting, apparently has a different signature, a coefficient of variation of between 0.11 and 0.14. I propose tentatively that non-counting-based timing in humans will often produce coefficients of variation close to these values. Further evidence on this point is presented below. It is also clear from Fig. 1 that counting-based interval production by humans is different again, exhibiting both mean accuracy and a very small coefficient of variation. At the very least, the data presented in Fig. 1 suggest that interval production in humans may exhibit different characteristics depending on whether or not it is counting-based.

Although interval production is a "classical" psychophysical method of investigation in human timing research, and has been commonly used, few studies have employed a range of interval values necessary for tests of scalar predictions; obviously, the requirement of constancy of coefficient of variation as interval value changes cannot be evaluated if only the production of a single interval value has been investigated. However, Zeiler *et al.* (1987) provide data which are generally suitable for a scalar analysis. The three experimenters produced target interval values of 0.5, 1, 2, 4, 8, 16, 32 s under different feedback conditions. One of these ("no upper bound") provided feedback (an illumination of a light marked
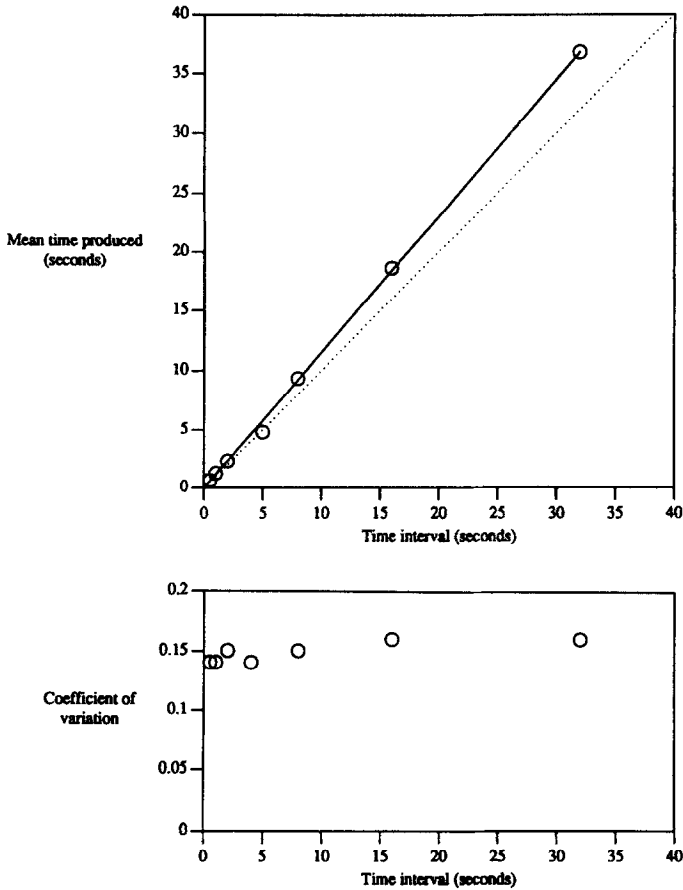
FIG. 2. Mean times produced (upper panel), and coefficient of variation (lower panel) from Zeiler, Scott, and Hoyert (1987, Table 1). Open circles show mean (upper panel) and coefficient of variation (lower panel). The solid line in the upper panel is the best-fitting linear regression line. The dotted line shows perfect correspondence between mean time produced and interval, where interval value in both panels is the lower bound of the reinforced duration.

"correct") when the required value was exceeded (lower values produced illumination of an "error" light). Subjects sometimes had not been informed verbally of the target interval value. Mean times produced, and coefficients of variation (calculated from Zeiler et al., 1987, Table 1, data aggregated over all three subjects), are shown in Fig. 2.

The upper panel of Fig. 2 shows data points for mean times produced, the regression line fitted to them, and a dotted line indicating mean accuracy. Obviously, mean times increased linearly with the lower bound of the intervals to be produced, although "overestimation" of this lower

bound occurred for most intervals. This may not be surprising given that the intervals associated with "correct" evaluations in this experiment could not be lower than the lower bound, and were necessarily greater. Another possible cause of the mean overestimating the schedule value is that such overestimation produces a higher rate of reinforcement on DRL-like schedules than positioning the mean at the schedule value (Wearden, in press).

The lower panel shows coefficients of variation, aggregated over all three subjects, as a function of the lower bound of the interval to be produced. The constancy of the coefficient of variation (over a 64-fold range of time values) is striking, and the absolute values (ranging from 0.14 to 0.16) are approximately the same as those found by Wearden and McShane (1988; see Fig. 1). In Zeiler et al.'s experiment, subjects "reported that they had not been aware of counting or performing other collateral behavior during any session" (p. 193). In view of the ubiquity of chronometric counting (Gilliland & Martin, 1940) this statement would be viewed with some scepticism were it not for the fact that the experimenters themselves served as subjects. It appears therefore that, once again, interval production by humans which is not mediated by counting conforms well to the requirements of scalar timing theory, with approximate mean accuracy of interval production, constancy of coefficient of variation, and coefficient of variation values in the range from about 0.10 to 0.16.

Ironically, in view of the results shown in the lower panel of the present Fig. 2, Zeiler et al. (1987) conclude that "Weber's law [i.e., constant coefficient of variation] did not fit the results" (p. 197, my addition in brackets). Their conclusion was probably based on an examination of individual differences, for different subjects did indeed exhibit differing trends in changes of coefficient of variation with interval produced; values for one subject increased markedly, another decreased and then increased, for a third, coefficient of variation remained roughly constant. In view of the lack of intersubject agreement as to trend, averaging the data (as was done to produce the lower panel of Fig. 2) does not seem unreasonable.

Given that coefficients of variation are Weber-fraction type measures (Gibbon, 1977), their constancy with changes in interval timed suggest approximate conformity to Weber's law in noncounting-based interval production by humans, a conclusion that is at variance with the usual assertion (Allan, 1979) that human timing is nonWeberian. In some cases, however, this assertion may depend on a confounding of counting- and noncounting-based timing. For example, the classic article by Treisman (1963), a landmark in the development of internal clock theory, presents some data from interval production by humans which appear, at first sight, to clearly violate Weber's law and, by implication, scalar timing theory. In Treisman's Experiment 1, three subjects produced intervals of

0.5, 1, 2, 3, 4, and 9 s. It is not clear whether feedback as to time produced was given (presumably it was not), nor is chronometric counting discussed, so presumably no steps were taken to discourage this. Weber fractions (calculated by a method somewhat different from coefficient of variation measures, see Treisman, 1963, p. 3 for details) declined systematically as the interval to be produced increased, taking values of 0.17, 0.16, 0.1, 0.09, 0.07, and 0.07 at the intervals described above.

Obviously, such a consistent decline in Weber fraction with time value produced represents a serious departure from scalar predictions. However, it is possible that Treisman's experiment has confounded noncounting-based and counting-based timing. In fact, the Weber fraction values found support this interpretation. According to the data in Figs. 1 and 2, non-counting-based timing should produce a coefficient of variation of 0.11 to 0.16, and the Weber fractions found by Treisman (1963) at 0.5 and 1.0 s (values for which chronometric counting may not spontaneously occur) are close to this range. On the other hand, coefficients of variation from interval production with chronometric counting are lower (see Fig. 1) and are close to the Weber fraction values (0.1 or less) found by Treisman at longer intervals. It appears, therefore, that Treisman's results may not, after all, run counter to the argument advanced in the present article that interval production by humans which is not based on chronometric counting often conforms well to scalar predictions. It is possible, however, that evidence supporting scalar timing in humans can be found even from the study of chronometric counting.

CHRONOMETRIC COUNTING

True chronometric counting is obviously an activity which is unique to humans who, alone in the animal kingdom, are able to represent numbers with any degree of accuracy although nonhuman animals can exhibit rudimentary numerical ability (see Davis & Memmott, 1982, for some examples). As such, the study of chronometric counting might seem to be barren ground for a search for timing processes which humans share with animals. However, if chronometric counting involves the reproduction of some basic count unit (e.g., the subject's representation of 1 s), the question arises of how the basic count unit itself is timed, since the units used in chronometric counting cannot themselves be produced by conscious counting. One possibility is that humans use a biologically based timing mechanism, possibly shared with animals, to produce the "raw material" for chronometric counting. If we identify this underlying process with the hypothetical scalar timing mechanism in humans, discussed above, we might expect that, whether or not the overall intervals produced by chronometric counting conformed to scalar timing requirements, some evidence for scalar processes in humans could be obtained by considering the count units themselves.

An experiment, conducted in my laboratory by Bairbre McShane, supports this argument. Four normal adults were required to produce an interval of 6-s duration, and were given accurate posttrial feedback as to the interval produced. The technique for producing the overall interval enabled the measurement of the intercount intervals the subjects used. To produce the target time subjects were required to repeatedly press a button. The first press produced a display "count 0" on a computer screen and started the interval, the next press the display "count 1," the next "count 2," and so on, until a required number of counts was produced. The number of counts required varied between conditions, although the target time remained constant at 6 s, and the numbers of intercount intervals required were 5, 6, 8, 10, and 12. That is, in the count 5 case, the subject would ideally produce an interval of 6 s by 5 productions of 1.2 s (6/5), in the count 12 case, the target interval would be produced by 12 productions of 0.5 s, and so on. This experiment thus kept the overall target time constant, while systematically varying the count unit. The subjects were instructed to respond at a "regular rate," but no feedback was provided for intercount intervals, only for the overall time produced.

Each subject produced the 6-s target time 10 times at each count requirement, so there were more intercount intervals produced with a 0.5-s count than with 1.2 s. Overall, the aggregate of four subjects produced from 200 (at 1.2 s) to 480 (at 0.5 s) intercount intervals. The experiment attempted to vary the count rate, that is, to change the "ideal" count unit used in chronometric counting. If the count units are derived from a scalar source then (i) mean intercount interval should increase linearly, and accurately, with changes in "ideal" count unit, (ii) the standard deviation should increase linearly with the mean, and (iii) the coefficient of variation of the intercount interval should remain constant as the count rate varies.

Linear regression was used to evaluate the first two predictions, and it was found that (i) the mean count interval increased linearly with the "ideal" value (slope $= 0.861$, intercept $= 0.04$, $r^2 = 0.99$), and (ii) the standard deviation of the intercount intervals increased linearly with their mean (slope $= 0.105$, intercept $= 0.04$, $r^2 = 0.94$). Slope value from the mean versus ideal count regression (0.861) was close to 1.0, and the intercept close to zero, so the results conformed to a reasonable first approximation to the scalar prediction that mean count should accurately reflect the ideal count value.

The third prediction of scalar timing theory was evaluated by a curve-fitting analysis. In their study of interval production by humans, Wearden and McShane (1988) fitted Gaussian curves to the relative frequency distributions of times produced and found that (a) the peak of these fitted curves was close to the target time, as this varied, and (b) the coefficient

of variation of the fitted curves remained approximately constant (in the range 0.11 to 0.14) as the target time changed. Intercount intervals in the present experiment were subjected to an identical analysis, and Fig. 3 shows the results.

The intercount intervals produced at each "ideal" value were grouped into relative frequency bins 0.05 s in width. Standard nonlinear regression (BMDP statistical package) fitted Gaussian curves to these distributions. The thoretically relevant statistics from this analysis are the (i) location of the peak of the curve, (ii) curve coefficient of variation, and (iii) proportion of variance accounted for by the curve. These three statistics are presented in each panel of Fig. 3, as well as the obtained data points and the fitted curve.

Overall, intercount intervals showed reasonable conformity to scalar predictions in that (i) peak location was generally close to "ideal" count value (although there was some deviation in the direction of underestimation at longer ideal count values), and (ii) coefficient of variation of fitted curves remained approximately constant (range 0.10 to 0.14) as the count unit changed. Proportion of variance accounted for by the fitted curve was high (minimum 86%) at all ideal count values.

Obviously, the results shown in Fig. 3, as well as the results of the regression analyses discussed above, provide a degree of support for the assertion that humans possess a scalar clock, albeit from an unexpected source, chronometric counting. Not only were the intercount intervals produced in reasonable conformity to scalar predictions, but the coefficients of variation from the curve-fitting analysis (Fig. 3) were usually within the range found in interval production by humans when counting does not occur.

Inspection of Figs. 1 and 3, where coefficients of variation from non-counting-based and counting-based timing are presented might at first sight suggest a paradox. I have argued above that noncounting-based timing in humans has scalar properties with a coefficient of variation of 0.10 to 0.16, whereas counting-based interval production produces lower values. It might seem strange that coefficients of variation of the *overall* intervals produced by counting-based timing are *lower* than the coefficients of variation of the count units of which they are made up. In fact, this apparent paradox is easy to resolve by developing a simple theory of chronometric counting.

Suppose, for example, that subjects produce an integer number of seconds (n seconds), by randomly sampling n times from a representation of 1 s which has the form of a Gaussian distribution with an accurate mean and a coefficient of variation within the range proposed above for noncounting-based timing (e.g., 0.14). To produce the required interval, the subject simply adds together in sequence the time values derived from randomly sampling the underlying representation of 1 s. In the first case
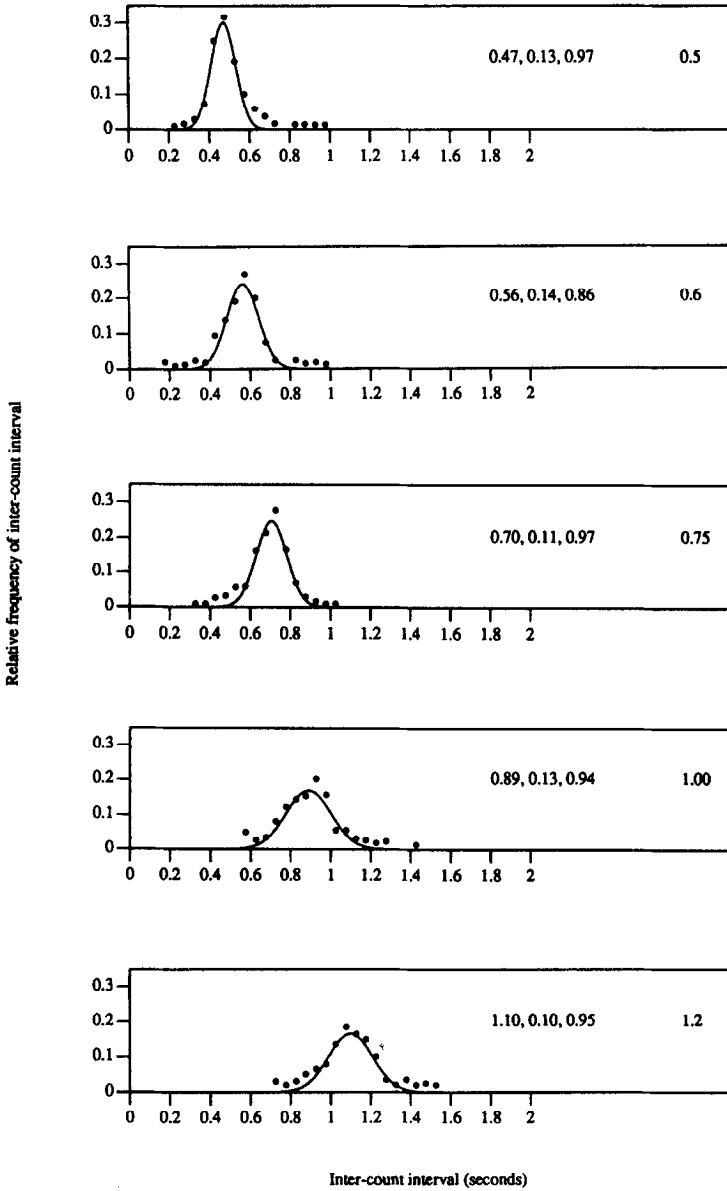
Inter-count interval (seconds)

FIG. 3. Relative frequency of counts of different duration plotted against intercount interval in an experiment which measured the duration of counts used during chronometric counting. Within each panel small filled circles show data points, the curve is the best-fitting Gaussian curve. The three numbers to the right of the curve (e.g., 0.47, 0.13, 0.97 in the upper panel) give, respectively, the peak location, the coefficient of variation, and the proportion of variance accounted for by the fitted curve. The number at the far right (e.g., 0.5 in the upper panel) shows the "ideal" count unit, the total duration to be produced divided by the number of counts.
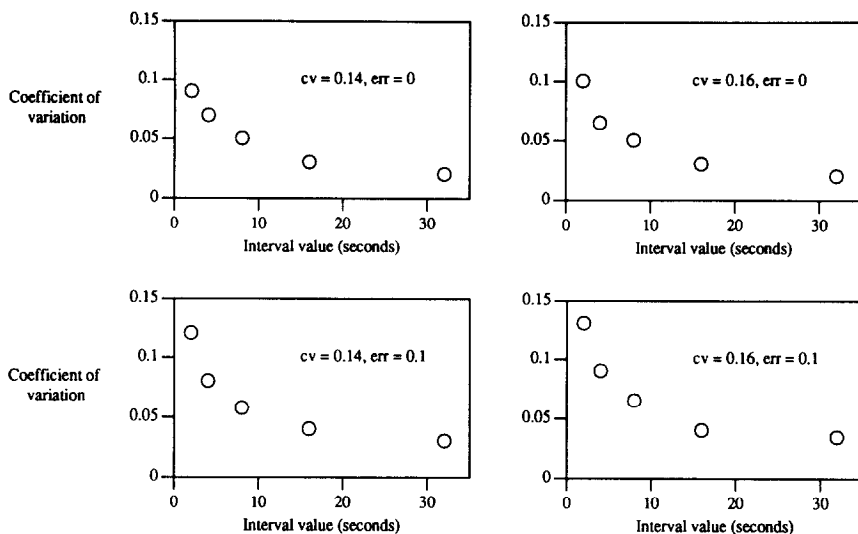
FIG. 4. Results simulated by the "scalar counting" model discussed in the text. All simulations involve productions of 2, 4, 8, 16, and 32 s. Each panel shows the coefficient of variation of the resulting productions (open circles), plotted against interval value. Different panels show the effect of varying the parameters of the model which are the coefficient of variation of the underlying count units (cv, taking values of 0.14 or 0.16), and random error (err, taking values of zero or plus or minus 0.1).

to be considered, this addition is performed without any error (such as might arise from motor factors).

This simple model, a sort of "scalar counting theory," was simulated as a Pascal program, which produced 100 "estimates" of 2-, 4-, 8-, 16-, and 32-s intervals. In all cases the mean times produced were almost perfectly accurate (a result intuitively derivable from the accurate mean count unit), and coefficients of variation are shown in Fig. 4.

The results in the upper panels of Fig. 4 come from simulations in which the underlying count unit coefficients of variation were either 0.14 (left panel) or 0.16 (right panel). It is immediately obvious from Fig. 4, however, that the coefficients of variation of the overall intervals produced (i.e., 2 to 32 s) were lower than the coefficient of variation of the underlying count unit, and that they declined systematically with increases in the overall time produced. The second of these properties, and the generally low coefficients of variation predicted by the model, have both been found in the intervals produced by counting-based timing in my laboratory (see Fig. 1 for an example).

The objection might be raised that the kind of error-free generation of count sequences proposed by this simple model is implausible, and that count unit production would be "contaminated," for example, by motor

errors in repetitive production of intervals. To investigate this possibility, the simple model outlined above was complicated by the addition of an "error" added to each count produced. In the simulation results shown in the lower panels of Fig. 4, this error was 0.1 or −0.1 s, with both values occurring with equal probability at random. The results in the lower panels of Fig. 4 obviously closely resembled those shown in the upper panels, apart from having slightly higher coefficients of variation, as a result of the addition of random error.

It seems clear, therefore, that results from counting-based production of time intervals may not in fact be incompatible with the notion that subjects have an underlying scalar representation of count units with a coefficient of variation in the range 0.10 to 0.16. Overall therefore, in spite of the fact that chronometric counting is a uniquely human activity, the idea that underlying count units have scalar properties, resulting from a timing mechanism analogous to that apparently possessed by animals, is not only largely consistent with actual measurements of the intervals used in chronometric counting (Fig. 3), but also leads to a simple model of chronometric counting which seems in reasonable quantitative accord with obtained data (Fig. 4).

Although chronometric counting is a near-ubiquitous phenomenon in experiments with humans involving the timing of durations of more than a second or so (Gilliland & Martin, 1940), it is only recently that it has received any sort of experimental and theoretical analysis. Killeen and Weiss (1987), for example, present an elaborate mathematical analysis of counting processes which is sufficiently general to cover counting too rapid to be accomplished by motor apparatus, such as counting of very rapid neural processes. That is, Killeen and Weiss analyze the process of count-ing *in principle*, they do not necessarily model the behavior of chrono-metric counting as it occurs in normal adults. In general, their work illustrates various ways in whch counting improves timing accuracy, and how the frequent result of underlying counting mechanisms is coefficients of variation (Weber fractions) which decrease as the interval to be timed increases (e.g., Figures 2, 3, 4, and 5 from Killeen & Weiss, 1987). They further discuss extensively the question of "optimality," i.e., what count unit value should be chosen to optimize timing accuracy. The very much simpler, and more restricted, model of chronometric counting advanced above is concerned with the somewhat different problem of how to model the actual behavior of human subjects while they explicitly count, whether or not such behavior is optimal. The above model incorporates the com-monplace observation that when humans are required to produce an in-teger number of seconds they count up to that integer value, and is also consistent with the results of measurement of intercount intervals, as shown in Fig. 3. It is further consistent with one of the very few previous attempts to manipulate intercount intervals by Getty (1976), who likewise

found that the "timing of subintervals is accomplished by a mechanism whose variance grows more as the square of time than as a linear function of it" (Killeen & Weiss, 1987, p. 460), in other words, a mechanism with scalar properties like those observed in Fig. 3. The study of explicit chronometric counting, as well as the "counting" of very rapid underlying events, such as neural pulsers, appears to be a field ripe for new developments. In particular, explicit chronometric counting, apparently a natural form of temporal regulation of behavior in normal humans, at last seems to be emerging from decades of mysterious neglect.

ANALOG EXPERIMENTS

Interval production, with or without chronometric counting, bears some clear procedural resemblances to standard training conditions used with animal subjects such as fixed-interval (Lowe & Wearden, 1981; Wearden & McShane, 1988) or temporal differentiation schedules (Jasselette et al., 1990). It is not, however, a very direct procedural analog of animal experiments, and has the additional difficulty that the task is contaminated, to some unknown degree, by motor factors. This section discusses results from my laboratory which have tested the performance of normal adult humans on tasks designed to be analogues of two procedures commonly used with animals, bisection (Church & DeLuty, 1977) and temporal generalization (Church & Gibbon, 1982). A complete report of the bisection experiments is available elsewhere (Wearden, in press).

Both sorts of analog experiment used a similar procedure. The durations presented to subjects were always short (less than 1 s), to avoid chronometric counting, and were delineated by two very brief clicks presented by a computer.

Bisection

A typical bisection experiment with animals (e.g., Church & DeLuty, 1977) proceeds as follows. During initial training, subjects are presented on each discrete trial with one of two standard stimuli (e.g., tones 2 or 8 s in length). Following tone presentations, subjects have the opportunity to respond on one of two different manipulanda such as levers. A response to one lever (e.g., the left one) is reinforced after one of the two durations (e.g., the longer one); a response on the other lever is reinforced after the other duration. After a number of sessions of training, high levels of temporal discrimination is achieved, with subjects responding on the "long" lever after the 8-s stimulus, and on the "short" lever after the 2-s one. Following this initial training, stimuli of various durations (usually ranging between the two standards) are presented without reinforcement, and the response of the animal is noted. In general, probability of a "long" response increases systematically with increasing interval duration.

A measure which has attracted considerable attention is the "bisection point," the stimulus duration after which 50% of responses occur on each manipulandum, and this is often considered to represent the animal's subjective midpoint of the two training durations. Experiments with animals generally find that the bisection point is located at the geometric mean (the square root of the product) of the two standard intervals rather than their arithmetic mean (Platt & Davis, 1983; Stubbs, 1976). In the case discussed above, the bisection point would thus be closer to 4 rather than 5 s.

In my laboratory, 10 normal adults participated in 2 experimental sessions, which began with 5 presentations of stimulus durations of either 0.2 and 0.8 s, or 0.1 and 0.9 s. These durations were described as Short or Long standards. Following this, subjects received stimuli of varying durations (ranging in 0.1-s steps from the Short to the Long standard) in a random series, and were required to classify each stimulus in terms of its similarity to the Short and Long standards, by pressing buttons marked appropriately (i.e., Short and Long responses). Subjects received two "refresher" presentations of the standard Short and Long stimuli, then another series of classifications, then more "refreshers," and so on until they had made 10 classifications of each duration. Each subject participated in two experimental sessions; for one of these the standard stimuli were 0.2 and 0.8 s, for the other 0.1 and 0.9 s. The session given first was counterbalanced between subjects.

Figure 5 shows the proportion of Long responses, plotted against stimulus duration. The upper panel shows data from the 0.2- to 0.8-s condition, the lower panel results from the 0.1- 0.9-s condition. It is clear from Fig. 5 that the experimental procedure used resulted in good temporal regulation of responding, even though the "training" phase of presentation of the standard durations was extremely brief relative to the training period needed with animals. In addition, the extremely low level of "errors" obtained (classification of the Short standard as Long, and vice versa) indicates a high level of attention to the task.

In the 0.2 to 0.8 condition, the bisection point (obtained by the method used by Maricq, Roberts, & Church, 1981) was located at 0.44 s, a value closer to the geometric mean (0.4 s) of the training stimuli than their arithmetic mean (0.5 s). Such a result appears at first sight to be evidence for animal-like interval bisection with humans. However, the bisection point from the 0.1 to 0.9 condition was located at the same value, whereas here the geometric mean is lower (0.3 s), suggesting instead that, in humans, the bisection point is located just below the *arithmetic* mean.

Empirically, therefore, this analog experiment (as well as another presented in Wearden, in press) suggests differences between animals and humans in location of the bisection point, even though in humans the intervals used were too short to make chronometric counting very useful
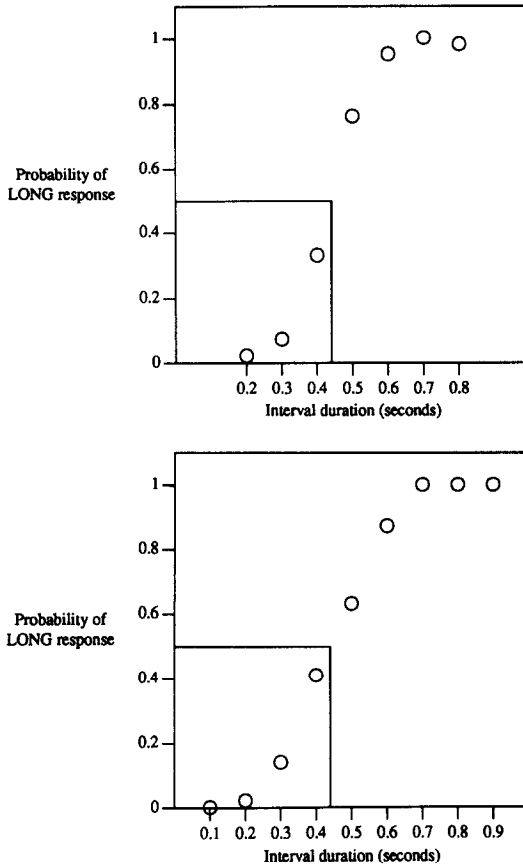
FIG. 5.   Results from the interval bisection experiment discussed in the text. Standard durations were 0.2 and 0.8 s (upper panel), 0.1 and 0.9 s (lower panel). Open circles show the probability of a Long response as a function of interval duration. The small square within each panel shows the bisection point (0.44 in both cases).

to subjects, and where, accordingly, "animal-like" timing processes might be expected to manifest themselves.

Gibbon (1981, pp. 175–179) provides a detailed analysis of various sorts of decisions that subjects might make on bisection tasks, based on the idea that subjective similarity of the current stimulus to the short and long durations is the basis of bisection decisions. In all cases, the starting point for analysis is the *difference* between the current stimulus to be classified, and the Short and Long standards, which are all assumed to have accurate mean representations. Arithmetic mean bisection can be derived from a simple difference decision (i.e., the arithmetic mean of the Short and Long durations is the point at which the absolute difference

of some value from the Short and Long standards is equal), or by expressing this difference as a proportion of the current time value. Thus, if $D(S,T) = T - S$, and $D(T,L) = L - T$, where T, S, and L are accurate mean representations of the stimulus to be judged (T) and the Short (S) and Long (L) standards, the point where $D(S,T) = D(T,L)$ is the arithmetic mean and, furthermore, the point where $D(S,T)/T = D(T,L)/T$ is also the arithmetic mean. However, when $D(S,T)$ and $D(T,L)$ are both expressed as fractions of their respective endpoints, these fractions [$D(S,T)/S$ and $D(T,L)/L$] are equal at the *harmonic* mean. Finally, the ratios $D(S,T)/T$ and $D(T,L)/L$ are equal at the *geometric* mean, the result routinely obtained with animals. Thus various possible locations of the bisection point are consistent, according to Gibbon's analysis, with different normalizations (i.e., divisions by different denominators) of the quantities $D(S,T)$ and $D(T,L)$, the real-time difference between the stimulus to be judged and the Short and Long standards.

The results shown in Fig. 5, in which the bisection point is located just below the arithmetic mean, are compatible with simple difference decision biased towards responding Long. The subject responds Long when the quantities $D(S,T)$ and $D(T,L)$ are similar; that is, when it is difficult to decide which is the smaller of $D(S,T)$ and $D(T,L)$ the subject responds Long. When $D(S,T)$ and $D(T,L)$ are not similar, the subject responds according to the standard value (S or L) which is the closer to T. A computer model using this decision as its basis provides a good fit to the values shown in Fig. 5, providing that it incorporates scalar representations of the Short and Long standard durations by embodying such representations in the form of Gaussian distributions with accurate means and constant coefficients of variation. Thus, overall, although the results from bisection experiments in humans and animals appear to differ (at least when the procedure reported above is used; see Allan and Gibbon, 1991, for a different result), it appears that it is the decision process that differs in humans and animals, not the fundamental properties of the representations of durations, which are compatible with scalar predictions. A similar point follows from another analogue experiment, this time on *temporal generalization.*

Temporal Generalization

In the initial temporal generalization experiment by Church and Gibbon (1982) responses by rats were reinforced if preceded by a stimulus of a standard duration (e.g., 4 s), not if preceded by other durations, both shorter and longer than the standard. Responding declined symmetrically as deviations from the standard increased, and this decline was such that equal real time deviations from the standard supported the same level of responding.

In my laboratory, 20 normal humans were allocated arbitrarily to one of two equal-sized groups, both of which received a single experimental session. A standard stimulus of 0.4 s was initially presented five times. Following this, subjects experienced a number of trials in which various durations were presented. Subjects were required to classify the durations as either Same or Different from the standard duration, and each response was followed by accurate feedback on a computer screen (e.g., "Yes, that was the standard duration"). The procedure for the two groups was identical except for the nonstandard durations used. For one group, Linear, nonstandard stimuli were spaced linearly around the standard in 0.1-s steps, for the other (Log group), the nonstandard stimuli were logarithmically spaced around the standard. For the Linear group, stimuli 1 through 7 in Fig. 6 were 0.1 to 0.7 s, respectively; for the Log group, stimuli 1 through 7 were 0.19, 0.24, 0.31, 0.4, 0.51, 0.66, and 0.84 s.

Figure 6 shows the probability of a Same response as a function of stimulus number for the Linear group (upper panel) and the Log group (lower panel). Both groups showed maximum probability of Same at the standard duration (stimulus 4), and both groups showed little or no tendency to classify remote stimuli 1 and 7 as Same, indicating a high level of attention to the task. The principal difference between the groups was in their response to stimuli 3 and 5, relative to the standard, 4. In the Linear condition, subjects were more likely to classify the 0.5-s stimulus as Same than the 0.3-s stimulus ($p < 0.01$ by Wilcoxon test). That is, equal real-time deviations from the standard did *not* support equal levels of responding, unlike the results from rats in Church and Gibbon (1982). In the Log group, on the other hand, there was no significant difference in probability of Same responses to stimuli 3 and 5, which were spaced around the standard in equal logarithmic steps, although a nonsignificant asymmetry was still present. These results were, once again, opposite to those obtained from rats (e.g., Church & Gibbon, 1982) which showed asymmetries in logarithmic but not real time.

Empirically, therefore, this second analog experiment suggested that human temporal generalization was symmetrical in logarithmic rather than linear time, whereas the reverse is true for animals (Church & Gibbon, 1982). The theoretical explanation of temporal generalization provided by Church and Gibbon (1982, p. 179) derives the symmetrical generalization gradients in real time from the decision to respond if $|S - T|/S < b$. Here, S is the standard, reinforced, duration, T is the current duration, and b is a threshold value. Although this basis for response decisions depends on a ratio calculation, the statistic calculated in fact behaves like a difference in the sense that for a constant standard, S, the only thing that distinguishes different T values is their absolute difference in real time from S. Obviously, such a linear process produces symmetrical response gradients in real time.
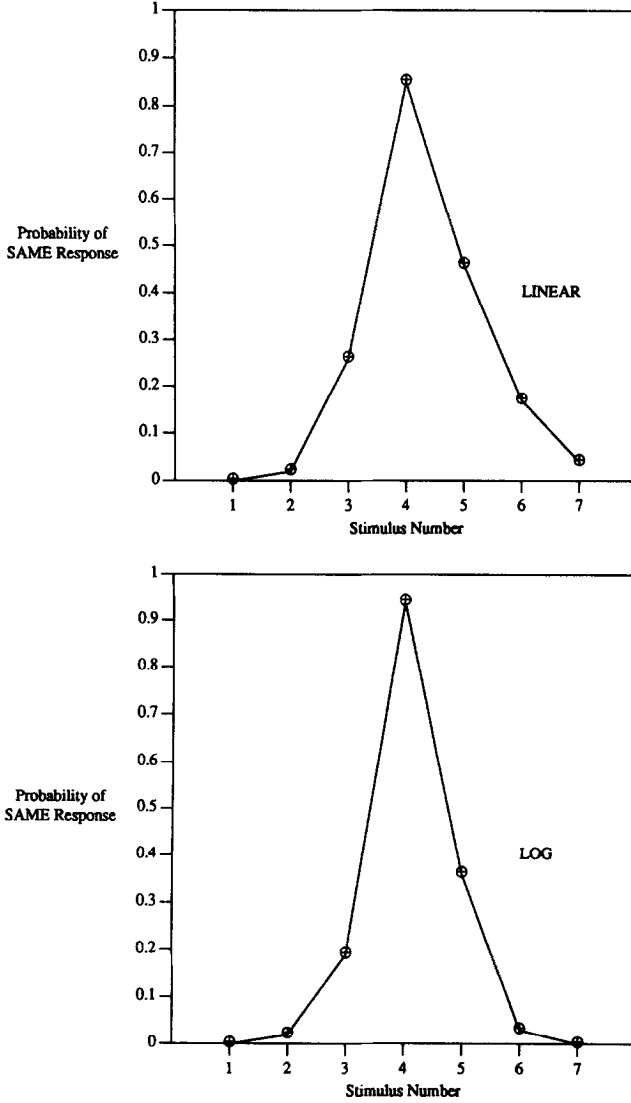
FIG. 6. Results from the temporal generalization experiment discussed in the text. Each panel shows probability (circles including crosses) of Same response as a function of stimulus number (number 4 was the standard duration). Upper panel shows results when nonstandard stimuli were linearly spaced around the standard; lower panel results when nonstandard stimuli were logarithmically spaced around the standard.

To produce the asymmetrical gradients obtained in the experiment with humans, some other type of decision process is needed, and several possibilities exist. One is to use the value $|S - T|/T$ as the basis for comparison with the threshold. This quantity does not behave like a simple difference, as different T values are distinguished by something other than just their absolute difference from S. Such a calculation would produce asymmetrical response gradients, as can easily be shown by substitution of the values used in the linear spacing condition of the experiment above. For example, for $T = 0.3$, the calculation above yields a value of 0.3, whereas for $T = 0.5$ the value is 0.2. Thus, 0.5 s is closer to the standard duration (for which the calculation above obviously yields a value of zero) than is 0.3 in terms of the calculation proposed. Another type of process producing response asymmetry is a type of *direct ratio comparison*. A complete model is too complex to be given here but the basic idea is as follows: Suppose that the problem facing the subject is to decide whether two durations are the same (i.e., to decide whether a given duration is the same as a standard). One way is to construct a *difference* between the to-be-judged stimulus and the standard and compare this with zero, responding Same if the difference is "small enough" as discussed above. Another is to construct the *ratio* of the two stimuli and compare this with unity, responding Same if the deviation from unity is less than some threshold value. A ratio comparison process which can yield asymmetrical judgments with equal real-time deviations from a standard is to construct the value $r = s/t$, where t is the interval to be judged and s is the standard. In the case of $s = 0.4$ s, then for $t = 0.5$ s, $r = 0.8$, and for $t = 0.3$ s, $r = 1.33$. Here the absolute difference between r and 1.0 is *greater* for 0.3 s than for 0.5 s; i.e., 0.5 s is nearer to 0.4 s than is 0.3 s. A model of this type, once again assuming scalar representations of standard and comparison durations, can model the asymmetry found in the Linear group in Fig. 6 and the lack of asymmetry found in the Log group.

Overall, the preliminary results from analog experiments presented above (and in Wearden, in press) suggest that ideas derived from scalar timing theory, and methodologies inspired by experiments conducted with animals, can be fruitfully applied to normal adults. Data are extremely orderly, the tasks apparently have high face-validity for subjects, and results can be obtained much more rapidly than is the case with nonhuman subjects. The findings do, however, suggest that the way humans make decisions in bisection and temporal generalization experiments is, for some reason, different from the hypothesized decision processes in animals. It remains to be seen whether the results shown in Fig. 5 and 6 are typically obtained from humans when variables such as stimulus modality (e.g., light versus sound), or type (e.g., filled versus unfilled intervals) are changed, and whether, in consequence, models based on scalar representations of time will continue to describe data. Allan and Gibbon (1991)

discuss a bisection experiment with human subjects which produces results consistent with geometric mean bisection. Their study, however, uses different intervals and a different discrimination technique from the experiment described above but, at the very least, their work implies that the results obtainable from experiments with human subjects depend critically on procedural variables. It is interesting, nevertheless, to note that Allan and Gibbon's theoretical analysis of their results incorporates scalar representations of duration with a sensitivity parameter, comparable to the Weber fraction or coefficient of variation, which is almost always in the range 0.10 to 0.16 (only 1 subject out of 11 shows a higher value), the range proposed in this article as a possible signature of scalar timing in humans.

   A possible future development in analog research might be the development of analogs for human subjects of some other animal timing tasks. The theoretically important "time-left" procedure of Gibbon and Church (1981) presents a particular methodological challenge if brief durations are to be used, as in its original form it involved pigeon subjects changing from one response manipulandum to another, something which cannot possibly be accomplished rapidly enough to develop a straightforward analogue for humans. However, overall, analog experiments like those described above seem a fruitful field of study, and may eventually lead to understanding of the type of representations that humans have of durations, as well as the ways that these representations are used in behavioral decision processes.

CRITICISMS AND CONCLUSIONS

   For clarity of exposition, the present article has taken a position of positive advocacy with respect to the question posed in its title, and has generally sought to amass positive evidence for scalar timing in humans, but how strong is that evidence? Consider, for example, the results from interval production discussed above (see also Wearden & McShane, 1988) and presented in Figs. 1 and 2. In general, results indicate that (i) mean time intervals produced generally closely match require clock time (scalar property 1), and (ii) standard deviation of response produced increases linearly with the mean (scalar property 2), often in the form of constant coefficients of variation (derived from curve-fitting analyses) as the interval to be timed varies.

   One possible criticism is that scalar properties 1 and 2 are, in fact, relatively weak predictions in the sense that they are compatible with a wide range of possible mechanisms. Certainly, property 1 is compatible with Poisson as well as scalar timing (Gibbon, 1977), and may also be derivable from various sorts of sequential decision processes when feedback is given (e.g., if interval produced on trial n is too big, decrease it by a random amount for trial $n + 1$, if too small, increase it by a random

amount). Scalar property 2 is more discriminating, but even this may be derivable from various sorts of underlying proportionalities in timing mechanisms (e.g., Killeen & Fetterman, 1988). However, more convincing evidence for scalar timing might come from examination of actual data values produced. For example, coefficients of variation from measures of animal timing are frequently within the range 0.2 to 0.3 (Platt, 1979), suggesting a mechanism common to different procedures. Likewise, in the present article, interval production by humans which is not based on counting consistently produced coefficients of variation within the range 0.10 to 0.16, again suggesting a common mechanism (see Figs. 1, 2, and 3, for example) across differing procedures.

Unfortunately, inspection of a range of coefficients of variation obtained from experiments using both human and animal subjects finds that values outside the characteristic ranges often occur (for animals, see Lejeune & Wearden, 1991, and also Cantor & Wilson, 1981). In particular, some duration judgment tasks using human subjects report coefficients of variation well outside the range found in interval production, even though the intervals judged were too short for chronometric counting to have occurred (e.g., Abel, 1972; Getty, 1975). There is even a commonly used short interval production task that consistently produces coefficients of variation below the range posited in the present article for scalar timing in humans, and this is the synchronized tapping task devised by Wing and Kristofferson (1973). Here, subjects are initially presented with a sequence of brief tones with constant spacing, and are required to tap a response key in synchrony with the tones. Then the tones are discontinued and subjects make a series of self-paced taps. Coefficients of variation derived from intertap intervals are often close to 0.05, even when the intertap intervals are too short to have been themselves produced by counting. Some particularly striking examples of low coefficients of variation come from Ivry and Keele (1989), who used the tapping task with patients with various sorts of cerebral damage, as well as college-age and elderly (mean age 66.7 years) controls. In this study the intertap interval was 550 msec, and the average self-paced intertap interval was close to this value for all subjects. Coefficients of variation of intertap interval (calculated from Ivry & Keele [1989], Table 1) were 0.04 in college-age students, 0.05 in elderly controls and patients with sensory loss in the limb producing the response, and 0.06 in patients suffering from Parkinson's disease. Only patients with cerebellar damage produced a coefficient of variation (0.09) approaching the value obtained in short interval production by healthy college-age subjects used by Wearden and McShane (1988). A recent study at the Manchester University Psychology Department (I. Todd and D. J. O'Boyle, personal communication) provided a comparison of interval production and paced tapping from a patient with unilateral cerebellar damage. Production of a 500-msec duration, using a

procedure identical to that of Wearden and McShane (1988), resulted in performance with a coefficient of variation of 0.30, when the production was made with the hand affected by the cerebellar damage, whereas the synchronized tapping task, with a 550 msec intertap interval, resulted in a coefficient value of 0.09.

It is therefore clear that two commonly used interval production tasks (production with chronometric counting and synchronized tapping) produce results contrary to the assertion that coefficients of variation from production experiments with humans are in the range 0.10 to 0.16. Various explanations for the low coefficients of variation obtained in chronometric counting experiments exist (Killeen & Weiss, 1987), and it has been shown above that they can be reconciled theoretically with underlying count units with coefficients of variation in the range proposed for scalar timing in humans in the present article. The data from synchronized tapping experiments present more of a challenge, particulrly if the standard analysis of such tasks (due to Wing & Kristoffersen, 1973) is followed, whereby the observed performance is derived from addition of timer and motor variance, as this results in underlying timer coefficients of variation even smaller than the values described above. Furthermore, according to this analysis, it is the *variance* of the underlying timing mechanism which grows linearly with the mean interval produced, rather than the standard deviation as predicted by scalar timing theory (see Wing, 1980, Fig. 6, p. 479, for a clear example). A particularly puzzling feature of synchronized tapping, in the present context, is that if normal subjects are initially synchronized to tap at regular 550-msec intervals then coefficients of variation derived from their intertap intervals will be about 0.05, whereas if they count spontaneously at a similar rate the coefficients of variation of their intercount intervals are much larger (see Fig. 3, 0.5- and 0.6-s "ideal count" conditions). Obviously, these problems need more experimental and theoretical analysis.

In spite of the qualifications and apparent exceptions discussed above, the search for scalar timing processes in humans appears fruitful and interesting, and overall a body of evidence supporting scalar timing in humans can be obtained, at least from some sorts of experiments. Although the value of animal experiments in the field of timing is sometimes questioned (e.g., Fraisse, 1984), it may be that future developments in the study of human timing inspired by data and theory derived from research with animals may go a long way toward answering questions posed about the possibility of an internal clock in humans more than 50 years ago by Francois (1927) and Hoagland (1935).

REFERENCES

Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, **51**, 1219–1223.

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

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, **22**, 39–58.

Cantor, M. B., & Wilson, J. F. (1981). Temporal uncertainty as an associative metric: Operant simulations of Pavlovian conditioning. *Journal of Experimental Psychology: General*, **110**, 232–268.

Church, R. M., & DeLuty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, **3**, 216–228.

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

Davis, H., & Memmott, J., (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin*, **92**, 547–571.

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

Fraisse, P. (1984). The perception and estimation of time. *Annual Review of Psychology*, **35**, 1–36.

Francois, M. (1927). Contribution a l'etude du sens du temps: La temperature interne comme facteur du variation de l'appreciation subjective des durees. *L'Annee Psychologique*, **27**, 186–204.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception and Psychophysics*, **18**, 1–8.

Getty, D. J. (1976). Counting processes in human timing. *Perception and Psychophysics*, **20**, 191–197.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, **84**, 278–325.

Gibbon, J. (1981). Two kinds of ambiguity in the study of psychological time. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior, Vol. I. Discriminative properties of reinforcement schedules* (pp. 157–189). Cambridge, MA:Ballinger.

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 87–108.

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

Gilliland, A. R., & Martin, R. (1940). Some factors in estimating short time intervals. *Journal of Experimental Psychology*, **27**, 243–255.

Hoagland, H. (1935). *Pacemakers in relation to aspects of behavior*. New York: Macmillan.

Irvy, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, **1**, 136–152.

Jasselette, P., Lejeune, H., & Wearden, J. H. (1990). The perching response and the laws of animal timing. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 150–161.

Killeen, P., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, **95**, 274–295.

Killeen P., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review*, **94**, 445–468.

LeJeune, H., & Wearden, J. H. (1991). The comparative psychology of fixed-interval responding: Some quantitative analyses. *Learning and Motivation*, **22**, 84–111.

Lowe, C. J., & Wearden, J. H. (1981). Weber's law and the fixed-interval postreinforcement pause. *Behaviour Analysis Letter*, **1**, 27–32.

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

Platt, J. R. (1979). Temporal differentiation and the psychophysics of time. In M. D. Zeiler

& P. Harzem (Eds.), *Reinforcement and the organisation of behaviour* (pp. 1–29). Chichester, England: Wiley.

Platt, J. R., & Davis, E. R. (1983). Bisection of temporal intervals by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 160–170.

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

Stubbs, D. A. (1976). Scaling of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior, 26,* 15–25.

Treisman, M. (1963). Temporal discrimination and the indifference interval. *Psychological Monographs, 77,* 1–31.

Wearden, J. H. (in press). Maximizing reinforcement rate on spaced-responding schedules under conditions of temporal uncertainty. *Behavioural Processes.*

Wearden, J. H. (in press). Human performance on an interval bisection task. *Quarterly Journal of Experimental Psychology.*

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

Wing, A. (1980). The long and short of timing in response sequences. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 469–486). Amsterdam: North-Holland Publishing Company.

Wing, A., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics, 14,* 5–12.

Zeiler, M. D., Scott, G. K., & Hoyert, M. S. (1987). Optimal temporal differentiation. *Journal of the Experimental Analysis of Behavior, 47,* 191–200.