Psychol. Belg.
1993, 33-2, 241-253.

DECISIONS AND MEMORIES IN HUMAN TIMING

John H. WEARDEN

Department of Psychology
Manchester University, U.K.

Some recent work has applied the clock-comparison model of scalar timing theory, originally used as an account of timed behaviour in animals, to human timing. As well as an internal clock, the scalar timing system also involves memory processes (e.g., storage of representations of critical durations) and decision processes. The paper discusses examples of decision processes that might be involved in temporal generalization and bisection, and presents a newer task where subjects have to decide whether one stimulus duration is half the length of another. This latter case produces data which appear to pose some problems for a simple treatment by scalar theory. Some recent studies of memory for duration in humans are reviewed. An experiment on memory for short "lists" (2 or 3 items) of tone durations is described, which found that memory for duration fell to chance when 3 tones had to be remembered. This limit is consistent with the idea that humans use a phonologically-specific articulatory mechanism to encode and retain tone durations, an assertion supported by other evidence. The value of considering human timing behaviour as the result of an interaction of clock, memory, and decision processes, rather than just as a problem of the relation between subjective and real time, was emphasized.

Some of my recent research involves the application of *scalar timing* theory, currently the most influential account of animal timing (Gibbon, 1977; Gibbon, Church, & Meck, 1984), to timing in humans. Two general reviews of this work have recently been written (Wearden, 1994; Wearden & Lejeune, 1993), and Wearden (1991a) also provides an earlier discussion. The present article focuses specifically on the role of decision and memory mechanisms in timed behaviour in humans.

Scalar timing theory in its most developed form (Gibbon et al., 1984) is a clock-comparison model involving different processes arranged in three sequential layers (Wearden, in press). The first layer is an *internal clock* which produces representations of durations (e.g., the durations of presented stimuli). The proposed clock is of the pulser-accumulator type, where the number of pulses arising from a pulse-generating mechanism in some period to be timed is accumulated,

Thanks are due to Andre Ferrara, Mark Benmore, Vicki Culpin, Ben Hubbard, David Donaldson, Caroline Walker, Oliver Truckle, and Raphael Brandon for providing some of the data discussed in this article.

and this accumulated number forms the "raw material" for subsequent time judgements. The second layer consists of two *memories* of times; one, a working memory or short-term store, which contains the memory representation of the most-recently timed event (or elapsing time in some cases), another, a reference memory or long-term store, which stores representations of "important" times, such as those associated with reward in animal experiments. Behaviour is output only as the result of processes in the third layer of the system, a *decision* mechanism which involves comparisons between the two memories; for example, is the most recently-timed stimulus sufficiently close to the "important" duration to necessitate an overt response?

These three layers, pulse emission and accumulation, memory, and decisions, can be used as the framework for more general theories of timing, not necessarily involving the idea of an internal clock. For example, the idea that duration estimates can be based on the number, type, or density of "events" (e.g., memories, changes of context, see Block, 1990, 1992) can be incorporated within a similar framework. Accumulation of event number might be the analogy of the pulser-accumulator clock mechanism, this number is stored temporarily and compared with some reference store, and the result of this comparison process can be used as the basis for decisions about time. Thus, although commentators quite properly distinguish clock-based and more "cognitive" mechanisms for timing (Block, 1990), these different approaches have some formal similarities, and may also share some common problems.

The classical psychophysical approach to time has been predominantly concerned with timing processes per se, for example, how hypothesized internal clocks might operate, what relations should be proposed between internal, psychological, time and physical time. Clock-comparison models, of which scalar timing is perhaps the best-developed example, further propose that timed behaviour arises from interactions of timing, memory, and decision processes. My own recent work has been more concerned with memories for duration and decisions relating to timed behaviour than it has with properties of internal clocks, as will be seen below.

Timing and decision processes in analogue experiments

Initial work on scalar timing in humans (Wearden, 1991a, b; Wearden & McShane, 1988) tried to develop analogues for human subjects

of animal experiments whose data formed the experimental support for the theory. The analogy between time interval production by humans (Wearden & McShane, 1988) and some animal timing tasks motivated the discovery that the intervals produced by humans could conform to the requirements of scalar timing theory if subjects did not count. More elaborate analogue experiments involved *temporal generalization* (Wearden, 1992; for the original animal experiment see Church & Gibbon, 1982) and *temporal bisection* (Wearden, 1991b; for a related animal experiment see Church & Deluty, 1977).

Temporal generalization is perhaps the simplest procedure to describe. Church & Gibbon's (1982) rats were rewarded for lever-presses when a just-presented stimulus had some critical duration (e.g., 4 sec), but not when it was longer or shorter. The probability of a lever press was highest at the target (i.e., 4 sec) duration, and declined symmetrically in real time as the difference between the just-presented stimulus and the target increased; for example, if the target was 4 sec, stimuli 3 and 5 sec long supported about the same level of responding. Wearden's (1992) analogue for humans used much shorter stimuli (to prevent chronometric counting, a common precaution in experiments with human subjects) with, for example in his Experiment 1, a 400 msec stimulus being identified as the standard. Subjects then had to decide whether stimuli ranging from 100 to 700 msec were or were not the standard. The proportion of identifications of a stimulus as the standard peaked at the real-time standard, but the function was asymmetrical, with a 500 msec stimulus being more confused with the standard than a 300 msec one. The upper panel of Figure 1 shows averaged data from 20 subjects tested on a variant of this task, where on each trial subjects received two short tones (the first one defined as the sample, the second one the comparison), with a delay (either 1, 2, 5, or 10 sec) between them. The sample duration changed randomly on each trial, and was drawn from a uniform distribution ranging from 400 to 600 msec. The comparison either had the same duration, was 100, 200, or 300 msec longer, or 100, 200, or 300 msec shorter. Subjects simply responded SAME or DIFFERENT (i.e., they performed an identity judgement on the sample and comparison). As the upper panel of Figure 1 shows, this variant of the task produced the same main result as the original from Wearden (1992); subjects were more likely to confuse the two stimuli when the comparison was *longer* than the sample than when it was *shorter* by the same amount.

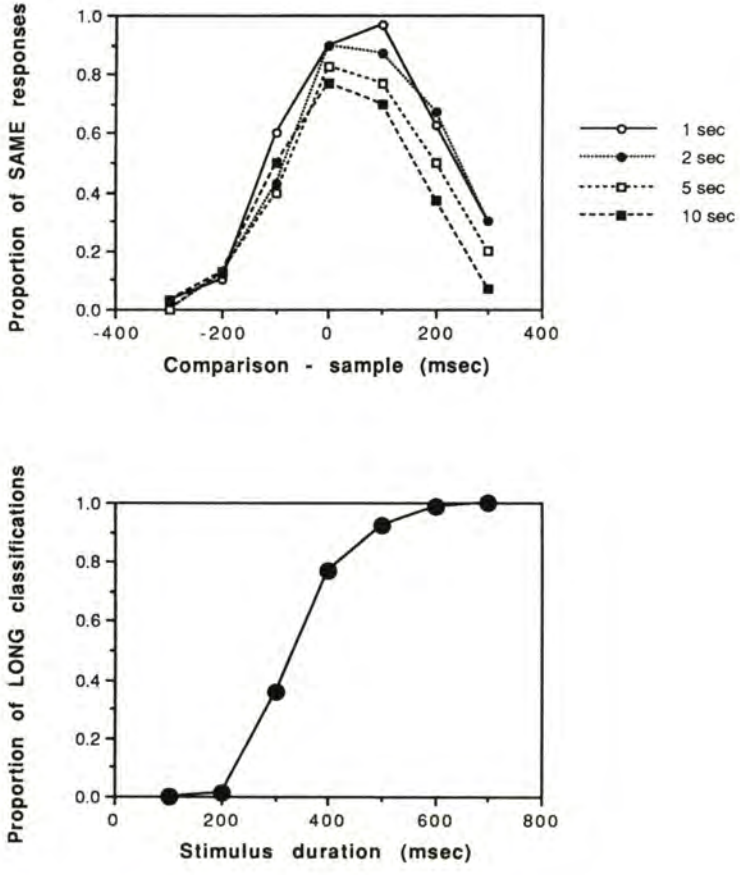

Figure 1: Upper panel. Proportion of SAME responses (i.e., judgements that the two durations were equally long) as a function of comparison - sample difference in the temporal generalization experiment described in the text. Positive difference values come from trials on which the comparison was *longer* than the sample, negative differences from trials on which the comparison was *shorter*. Data are shown separately for the four different inter-stimulus intervals used (1 [open circles], 2 [filled circles], 5 [open squares], and 10 [filled squares] seconds). Lower panel: Data from the bisection study described in the text. Proportion of LONG classifications (i.e., judgements that a presented stimulus was more similar to the 700 msec standard than the 100 msec standard) are plotted as a function of stimulus length.

In the original animal experiment on temporal bisection (Church & Deluty, 1977), rats were trained to make one response after a short stimulus (e.g., 2 sec), another response after a longer one (e.g., 8 sec). After this discrimination was mastered, rats received a range of stimulus durations, ranging between and including the short and long standards, and the response to each stimulus was noted. A focus of theoretical interest was the location of the bisection point, the stimulus duration giving rise to 50% of responses appropriate to the short and long standards. In most animal studies the bisection point is located at the geometric mean of the short and long standards (the square root of their product, 4 sec in the example given above). The lower panel of Figure 1 shows results from an unpublished experiment on temporal bisection in humans (for other data see Wearden, 1991b), where a 100 msec tone was identified as the *short* standard, a 700 msec tone as the *long* standard, and durations ranging from 100 to 700 msec, in 100-msec steps, were presented. Subjects were required to classify each stimulus in terms of its similarity to either the long or short standards. The number of *long* classifications increased systematically with increasing duration, but the stimulus giving rise to 50% long classifications was clearly above the geometric mean of the short and long standards (265 msec in the example shown), but somewhat below the arithmetic mean (400 msec), the result typically obtained in my laboratory (Wearden, 1991b, but see Allan & Gibbon, 1991).

Why do humans produce the type of behaviours shown in Figure 1? The answer proposed by recently-published work is that the *decision* processes that humans use are different from those used by animals, rather than there being differences in the type of time representations that animals and humans possess. The simplest illustration of this comes from the analysis of temporal generalization. Suppose that s^* is the internal representation of the standard duration, s, which is on average accurate but variable from trial to trial, t is the duration of the stimulus most recently presented, which is assumed to be timed without error, and b^* is a threshold variable from trial to trial. Church and Gibbon (1982) assumed that animals responded when

$$\frac{abs(s^* - t)}{s^*} < b^*$$

where *abs* is the absolute value of the bracketed quantity, a decision process which will produce symmetrical temporal generalization gradients (i.e., the same response level at durations equal temporal distances above and below the standard), whereas Wearden (1992) derived his asymmetrical gradients from a decision process of responding "Yes" when

$$\frac{abs(s^* - t)}{t^*} < b^*$$

where all terms are as for the previous equation. Thus timing processes are supposed to have the same basic properties in humans and animals, and decision differences are the hypothesized cause of differences in behaviour. Why decision processes should differ in animals and humans remains to be discovered (although a decision rule like the one proposed above appears to operate in many different sorts of timing tasks when humans are used, see Wearden & Ferrara [1993] for another example), but the essential point here is that behaviour differences between species, or conditions using the same species, need not necessarily imply fundamental differences in time representations.

In discussion of data from bisection experiments with humans, Wearden (1991b) presented a computer model which was in good quantitative accord with data. This is too complex for exposition here, but its basic assumption was the notion that humans employed different types of temporal comparison processes from animals. In other words, once again, animal/human differences are explained by different *decision* processes, not by fundamentally different *timing* processes.

A recent study in my laboratory involved human subjects making a more complicated decision about durations than those required by temporal generalization or bisection, and in this case it is unclear whether the underlying principles of animal models can be successfully applied to humans. The task was a type of *fractionation* task, where subjects had to judge whether one stimulus duration was equal to *half* of another. Subjects received an 800 msec tone as the first or second stimulus of a pair (which were separated by a brief and variable gap) and the other stimulus ranged in length from 100 to 700 msec in 100-msec steps. The decision required was whether the shorter stimulus was *half* the length of the longer one, and no feedback was given, neither was any example of the duration that really was one half (400 msec) explicitly identified as such.

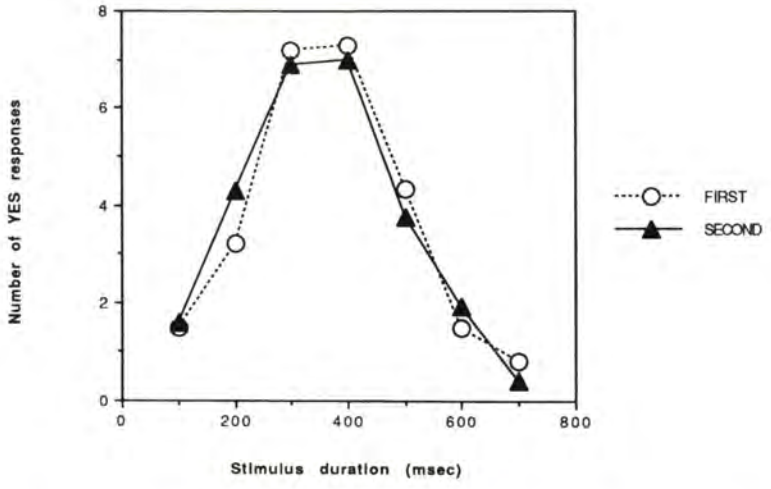

Figure 2. Proportion of YES responses (i.e., decisions that the shorter stimulus of a pair was equal to half the longer, 800 msec, stimulus), as a function of stimulus length.

Figure 2 shows the results, where a YES response represents a judgement that a presented duration *was* half of 800 msec. The FIRST or SECOND conditions (using 15 subjects each) were distinguished by whether the longer stimulus came first or second, but no effect of order was found. It can be seen that although the real-time half value (400 msec) tied for maximum number of YES responses, the data suggest that the underlying half-point was *less* than 400. For example, 500 msec produced fewer YES responses than 300 msec, and overall the function shown in Figure 2 appears symmetrical around a value of about 350 msec. The data obtained on the fractionation task show a number of unusual features; for one thing, subjects make more YES responses to stimuli shorter than the real half-point than they do to stimuli longer by the same amount, an asymmetry which is roughly the mirror-image of that obtained in temporal generalization, where longer stimuli are more often confused with a stimulus in the middle of the stimulus range than shorter ones are (see Figure 1, and other examples in Wearden, 1992). Another peculiarity is that the data at first sight appear to violate an important property of scalar time representations, namely that the estimate of some time, t, is on average accurate. If such perfectly veridical time representations are operating in the present case, an exactly accurate subdivision of the 800 msec

tone would be expected, whereas in fact the subjective half of this value appears to be about 350 msec. In contrast to this result, Gibbon and Church (1981) report data from animals on a task which resembles fractionation and find, in accordance with the tenets of scalar timing theory, that fractionation is accurately performed on average. Finally, even if we assume that for some reason the subjective half-point of 800 msec is about 350 msec, and that subjects perform on the fractionation task by deciding whether each stimulus length is or is not equal to this subjective value, the generalization functions obtained are roughly symmetrical (i.e., about equal levels of YES responses at 300 and 400 msec, and at 200 and 500 msec, see Figure 2), implying a decision rule like that appropriate to temporal generalization in animals (Church & Gibbon, 1982) rather than humans (Wearden, 1992).

The data in Figure 2 are presented here only as an illustration of the kind of results which can be obtained, and the kind of theoretical challenges such results might pose, when subjects are required to make more complicated decisions about the relation between two durations than their identity (as in temporal generalization) or similarity (as in bisection). Whether the theoretical models derived from animal studies will continue to be applicable in such cases remains to be seen.

Memory for duration

As mentioned above, the clock-comparison model of scalar timing involves clock, memory, and decision processes. Of the three, the study of temporal memory processes is by far the least advanced yet, as Wearden and Ferrara (1993) point out, many timing tasks implicitly involve memory for duration. The animal literature on memory for duration (reviewed by Spetch & Rusak, 1992) suggests that duration memories are prone to an unusual form of forgetting called *subjective shortening*. This simply means that if some stimulus duration is stored in memory, its internal representation becomes progressively *shorter* as the memory ages. Subjective shortening provides indirect evidence for what Wearden (1994) refers to as *quantity encoding* of time, where longer times are represented as larger quantities of something, for example as larger numbers of stored pulses. If forgetting occurs by quantitative loss of stored contents, then subjective shortening will necessarily follow. It should be noted, however, that some proposals for how durations are represented employ *qualitative* encoding (e.g.,

registration of oscillator states, see Church & Broadbent, 1990), where subjective shortening would not be so readily expected.

Wearden and Ferrara (1993) investigated whether something like subjective shortening could be observed in humans' memory for short-duration tones (average 400 msec). Their experimental technique involved presenting subjects with two tones (the first defined as the sample, the second as the comparison), with delays ranging from 1 to 16 sec between them. After the comparison tone, subjects had to judge whether it was longer, shorter, or the same length as the sample. Wearden and Ferrara found evidence for subjective shortening of the representation of the sample tone. For example, if the internal representation of the sample becomes progressively shorter as the time between sample and comparison presentation increases, performance should improve with increasing sample-comparison delay on trials on which the comparison is actually *longer* than the sample, as subjective shortening makes the stimuli more discriminable as the time between them increases. This unusual effect, representing an *improvement* in performance accuracy with increasing retention time, was obtained.

A complicating factor in Wearden and Ferrara's experiment was the presence of a positive time-order error (TOE), in the form of a tendency to judge that the sample was longer than the comparison, regardless of its actual length. Time-order errors can be found when decisions are made about stimuli in many sensory modalities, and a small literature on time-order errors for duration judgements exists (e.g., Allan, 1977; Jamieson & Petrusic, 1975a, b). By definition, TOEs are present when some judgement (such as equality) is made about two successively-presented stimuli S_1 and S_2, and when the result of the judgement depends on which stimulus comes first. Positive TOEs mean that S_1 appears to be weighted more than S_2, negative TOEs the reverse; in the case of TOEs for duration, for example, a positive TOE would mean that the first stimulus would tend to be judged longer than the second. If the S_1S_2 sequence is taken as an "item list", parallels emerge between the different TOEs and memory effects like "primacy" (better memory of the early items, like a positive TOE), and "recency" (better memory of later items, like a negative TOE).

Exploring this view of TOEs led me to try to set up "lists" of durations to see whether primacy and recency effects could be obtained. If a number of durations are presented, for example, are subjects better able to make judgements about the first, the last, or some other one?

An unpublished study from my laboratory tried to explore this, using what appeared at first to be a simple technique. Different subjects received "lists" of tones of different duration (with either 2 or 3 "items"), then a retention interval of 5 sec, then presentation of one of the tones in the list. The subject had to judge which one it was (first or second [2 item list] or first, second, or third [3 item list]). The tones in the list differed by 100 msec around a mean of 400 msec, and the "target" tone was, of course, equally likely to be in all serial positions in the list.

In the case of both the 2- and 3-item list, more correct judgements were made when the "target" duration was in the first serial position, but the difference was never statistically significant. The most striking feature of performance on this study was, however, that accuracy remained close to chance levels at all serial positions. Unlike memory for words, where 7 or more items can be routinely remembered, the storage of the duration of short tones seems much more limited, possibly to as few as 2 items. One implication of this limitation is that experiments requiring the judgement of several sequentially presented stimulus durations, all of which must be retained at once, are probably impossible to carry out, and one might ask why memory for duration is so limited.

In other unpublished work, Wearden and Culpin drew explicit parallels between the psychological processes that might operate in Wearden and Ferrara's memory for duration task and those used for short term verbal memory, and they suggested that a *phonological loop* memory mechanism (Baddeley, 1986, 1992), usually employed for short-term retention of words, might serve to preserve a copy of the sample duration during the sample-comparison delay in the memory for duration task. The phonological loop is a subsystem of the working memory model (Baddeley, 1986), and is a limited-capacity repetition mechanism where contents are periodically refreshed by rehearsal. If such a mechanism is implicated in the retention of tone durations, then manipulations which affect the phonological loop in verbal memory studies should also exert analogous effects in memory for duration tasks. One important result supporting this view came from a condition in which subjects had to remember the duration of a tone, with concurrent storage of a word presented in writing. Subjects made significantly more errors on the memory for duration task on trials when they also had to remember a word compared with those on which no word was presented but, more strikingly, having to remember a word with 3 syllables disrupted the memory for duration task significantly

more than remembering a 1-syllable word. This was true even though the 1- and 3-syllable words were drawn from a population with the same word frequency, and even though both types of words were remembered almost perfectly. The number of syllables in a word affects articulatory processes, so the effect of syllable number on memory for tone duration suggests that subjects are retaining the tone in a phonologically-specific mechanism with which articulation of other material competes.

Assuming that subjects are remembering tone durations in this way may also explain why such memory is so limited in the list experiment described above. According to standard theory, the phonological loop can store material which can be repeated in a total span of about 1.5-2 sec, so it can only contain at most 2 seconds worth of material, of whatever sort. Adding together the lengths of the tones, together with the intervals between them, suggests that performance would fall to near chance levels with between 2 and 3 tones, as these fill up the available time, and this result is close to the one obtained. At first sight, it would seem that this explanation could be easily tested by using tones of shorter duration (so more could be encompassed in the critical 1.5 to 2-sec period), but in practice subjects may find very short duration tones difficult to discriminate if their durations differ by small amounts. If length differences are greater, on the other hand, subjects may verbally label the stimuli (e.g., distinguishing between click-like stimuli and more temporally-extended tones), and use the verbal labels as the basis for their judgements, rather than duration per se (see Wearden & Ferrara [1993], p. 165-166, for discussion). These considerations suggest that some new method may be needed if the possibility of serial-position-like effects in memory for duration is to be properly investigated.

Concluding comments

The above brief illustration of some recent results shows how my own work has evolved from simple demonstrations of behaviour from humans exhibiting properties found in animal timing (e.g., in Wearden & McShane, 1988) to a fuller consideration of the interaction of different sorts of mechanisms (clocks, memories, and decision processes) in determining observed behaviour. Traditional internal clock theories may have treated clock properties in a sophisticated way, but they

tended to regard the decision and memory mechanisms as completely transparent (i.e., assuming that observed behaviour directly reflects underlying clock properties), or simple transforms (e.g., linear ones) of underlying subjective time. Scalar timing theory is considerably more developed, but may still not be sufficiently elaborate to encompass the range of phenomena found even in restricted domains of timed behaviour in humans. In particular, linking together studies of timing with methods and theories developed in research on human memory, or other areas of conventional cognitive psychology, may be especially fruitful.

Even if internal clocks exist in animals and humans (and not everyone regards their postulation as very useful, see Block, 1990), knowing what type of time representations the clock produces may not be sufficient to understand timed behaviour. Detailed consideration may need to be given to the way in which event durations are encoded and retained, and the way in which duration representations form the raw material for decision processes which intervene between timed events and observed behaviour.

REFERENCES

Allan, L.G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology, 31*, 24-31.

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

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

Baddeley, A.D. (1992). Is working memory working? *Quarterly Journal of Experimental Psychology, 44A*, 1-31.

Block, R.A. (1990). Models of psychological time. In R.A. Block (Ed.), *Cognitive models of psychological time* (pp. 1-35). Hillsdale, NJ: Erlbaum.

Block, R.A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas, & W.J. Friedman (Eds.), *Time, action and cognition* (pp. 141-152). Dordrecht: Kluwer.

Church, R.M., & Broadbent, H.A. (1990). Alternative representations of time, number and rate. *Cognition, 37*, 55-81.

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.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279-325.

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.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception* (pp. 52-77). New York, NY: New York Academy of Sciences.

Jamieson, D.G., & Petrusic, W.M. (1975a) Presentation order effects in duration discrimination. *Perception and Psychophysics, 17*, 197-202.

Jamieson, D.G., & Petrusic, W.M. (1975b). The dependence of time-order error direction on stimulus range. *Canadian Journal of Psychology, 29*, 175-182.

Spetch, M.L., & Rusak, B. (1992). "Time present and time past". In W.K. Honig & J.G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 47-67). Hillsdale, NJ: Erlbaum.

Wearden, J.H. (1991a). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation, 22*, 59-83.

Wearden, J.H. (1991b). Human performance on an analogue of an interval bisection task. *Quarterly Journal of Experimental Psychology, 43B*, 59-81.

Wearden, J.H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 18*, 134-144.

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

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

Wearden, J.H., & Lejeune, H. (1993). Across the Great Divide: Animal psychology and time in humans. *Time and Society, 2*, 87-106.

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.

Oxford Road
Manchester M13 9PL
United Kingdom