

# Chapter 2: Applying the scalar timing model to human time psychology: Progress and challenges

JOHN H. WEARDEN

## Abstract

Scalar timing (or scalar expectancy) theory, SET, was originally developed as an explanation of the performance of animals on temporally-constrained reinforcement schedules. In the last decade, however, it has had a growing influence on the study of human timing, and may now even be the dominant approach. The present paper reviews its successes, and points to some challenges for the future. Among the successes are (1) a re-invigoration of the old idea that some aspects of timing in humans depend on an “internal clock,” (2) the provision of a framework for developmental studies of timing in humans (both of children and the elderly), (3) the development of precise quantitative models of timing in humans, which depend on an interaction of clock, memory, and decision processes. In spite of these successes, however, many problems remain. Some of these concern details of how the SET system itself works, particularly questions concerning the roles of memory and decision processes. Data from recent experiments which manipulate memory and decision mechanisms in the SET model, in an attempt to clarify their operation, will be presented. Another set of problems concerns the application of SET-related ideas to “classical” timing tasks such as production, reproduction, and verbal estimation. Behavior on this “classic trio” of tasks often seems at variance with the scalar model, but it will be shown that the incompatibility may be more apparent than real, and that SET-based models may be used to explain many aspects of performance on these classical tasks. Thus, not only can SET-based models account for recently-collected data on many timing tasks, but they may be able to provide the first rigorous models of behavior on procedures known for more than 150 years.

## Introduction

The application of *scalar timing* theory (or *scalar expectancy* theory, SET) to human time psychology is one of the (rather few) success stories of the long history of attempts to apply models developed initially with non-human animals to the behavior of humans. SET was developed initially to explain the striking regularities in the behavior of rats and pigeons both on classical schedules of reinforcement with some sort of temporal constraint, and on specially designed timing tasks such as

bisection (Church & DeLuty, 1977), and temporal generalization (Church & Gibbon, 1982).

An article by Wearden and McShane (1988) was probably the first to analyze human timing performance using ideas derived from SET, although the possibility of doing this had been mentioned before by others. Since then, perhaps as many as 100 journal articles and chapters have applied ideas derived from SET to human timing, and a recent review by Allan (1998) discusses many of them. Lack of space precludes a complete exposition of SET here, but non-technical discussions of the operation of SET are available in Wearden (1994, 1999, 2001).

After this initial introduction, the remainder of this chapter is organized into three main sections. The first (*Progress*) discusses what seem to me to be examples of unalloyed progress in understanding human timing made possible by SET. The second (*Progress and Challenges*) discusses some areas where, although SET has made major contributions to increasing understanding, important problems remain unsolved. The final section (*Challenges*) discusses areas which appear problematical for SET, and focuses particularly on one of them, the application of SET to “classical” timing tasks.

Although some aspects of the detailed mechanics of SET will be mentioned at various points in this chapter, two features of the SET model need to be mentioned in this preface. Firstly, when we say that some behavior conforms or does not conform to SET, what do we mean by this? SET requires that time representations have two properties. The first of these is *mean accuracy*: the requirement that internal “estimates” of some real time duration,  $t$ , are on average accurate. The second is the *scalar property* which gives the theory its name. This is essentially a form of Weber's law, which can be exemplified in various ways. Perhaps the simplest of these is the requirement that the standard deviation of judgments of time grows as a linear function of the mean, or alternatively that the coefficient of variation statistic (standard deviation/mean) remains constant as the interval to be timed varies. Another way of testing the scalar property is to inspect the data for the property of *superimposition* (called *superposition* in U.S. English), the requirement that measures of timing obtained when different absolute durations are timed superimpose when plotted on the same relative scale. Thus, behavior which does not conform to mean accuracy or the scalar property seems, at least at first sight, incompatible with SET.

The other feature of SET which needs to be mentioned here is the structure of its well-known information-processing variant. This embodies the SET system in a tripartite structure. The first component is a pacemaker-accumulator internal clock which provides “raw” representations of durations. The second is a memory system consisting of a short-term working memory reflecting accumulator contents, and the second is a reference memory component containing some kind of representation of “important” times such as those associated with reinforcement in experiments with animals or used as standards in experiments with humans. Finally, a decision component is needed to produce observed behavior, and this usually involves comparisons of the contents of working memory with standards stored in reference memory. As will be seen below, some progress has been made isolating these different parts of the SET system, although many things currently remain obscure.

## Progress

I would like to concentrate on two aspects of the application of the SET system to human timing that I personally consider to have been particularly fruitful. The first is the fact that SET has provided coherent quantitative models of behavior on some timing tasks that have proved useful in understanding the ways in which different groups (e.g., children, student-age adults, and the elderly) differs. The second is the fact that SET has revitalized the old idea that humans might perform on some timing tasks by using an internal clock.

Pre-SET time psychology is filled (one might almost say “littered”) with studies comparing group X and group Y on some timing task and finding that they behave differently. So, children and young adults differ, people with brain damage differ from their intact controls, things happen to timing performance as people get older....but so what? Merely demonstrating that two subject groups differ in timing performance is almost spectacularly uninformative, as it gives us no insight whatsoever into what mechanisms might underlie the performance differences observed. By providing quantitative models of performance on (some) timing tasks, SET lets us go beneath the skin of the subject and at least make reasonable conjectures as to what mechanisms might be involved in performance differences.

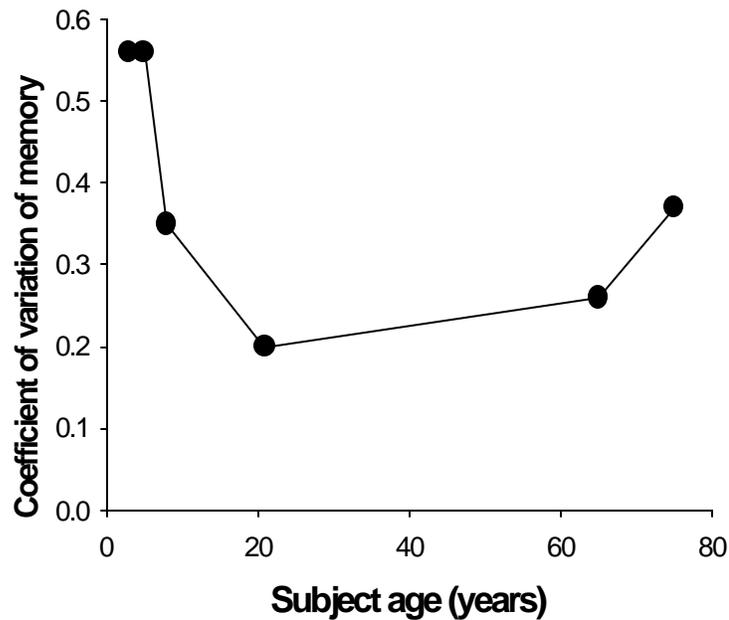
Perhaps the simplest examples come from the study of temporal generalization performance in different groups. Temporal generalization, developed originally as a method for studying timing in rats by Church and Gibbon (1982), has been extensively applied to humans, usually in variants of a method developed by Wearden (1991a, 1992). Humans are initially presented with some duration identified as a standard (e.g., a tone 400 ms long). They then receive other durations, shorter or longer than the standard, or equal to it, and simply have to judge whether or not each duration was the standard by responding YES or NO. The proportion of YES responses plotted against stimulus duration is a temporal generalization gradient. The conventional analysis of human behavior on this task is the “modified Church and Gibbon”, MCG model, developed by Wearden (1992) from Church and Gibbon (1982).

On temporal generalization, the standard duration ( $s$ ) is assumed to be stored in reference memory, and represented as a Gaussian distribution with a mean  $s$  and some coefficient of variation  $c$ , essentially representing the precision with which the standard duration is stored. Each comparison duration,  $t$ , is assumed to be timed without error, and the MCG model produces a YES response when  $abs(t - s^*)/t < b^*$ , where  $s^*$  is a sample drawn from the reference memory (which differs on each trial), and  $b^*$  is a sample drawn from a threshold distribution with mean  $b$ , and some standard deviation which is usually  $0.5b$  ( $abs$  indicates absolute value).

If different groups differ in performance on temporal generalization performance then, according to the MCG model, they might differ in  $c$ , the “fuzziness” of the representation of the standard, or in  $b$ , the decision threshold. Wearden, Wearden, and Rabbitt (1997a) tested the performance of elderly subjects (60-80 years) on temporal generalization, and looked at effects of both age and IQ. More recently, Droit-Volet, Clément, and Wearden (2001) examined the temporal generalization

performance of children of 3, 5, and 8 years, and several studies (Wearden, 1992; Wearden, Denovan, Fakhri, & Haworth, 1997b) have obtained data from student-age subjects. In general, “precision” of performance in terms of the width of the temporal generalization gradient was greatest in students, and poorer both in children and in the elderly. The MCG model suggested that the principal cause was the coefficient of variation,  $c$ , of the memory representation of the standard, and Figure 1 shows  $c$  plotted against subject age, with student data averaged over a number of conditions and assuming a subject age of 21 years.

The effects are obvious: children of 3 and 5 had the highest  $c$  values, these dipped markedly at the age of 8, and continued to dip towards student age, only to slowly rise again in the groups averaging 65 and 75 years. To be strictly accurate, some other developmental trends were present as well, such as a tendency for the youngest children to respond YES or NO randomly (Droit-Volet et al., 2001), but the essential point here is how the quantitative model derived from SET enables theoretically meaningful comparisons of the behavior of different groups, thus going much further than the usual final conclusion that different groups differ in performance.....but we can't say any more. The old adage that “nothing is so practical as a good theory” seems highly applicable here, and if SET-compatible tasks were used with, say, patients with brain damage then we might be able to draw conclusions stronger than those presently available.



**Figure 1:** Coefficients of variation of memory representation of standard durations obtained from fits of the MCG model to temporal generalization performance, plotted against subject age. Data come from Droit-Volet, Clément, & Wearden (2001), Wearden (1992), and Wearden et al. (1997a, b)

The second area of progress for SET has been a revitalization of the old idea, traceable at least back to François (1927) that humans possess some sort of *internal clock*. Once again, space constraints mean that I cannot do justice to this important notion, which has spawned much diverse research. Wearden and Penton-Voak (1995) reviewed the bizarre field of body temperature and time estimation, and found general support for internal clock ideas. Luckily, Treisman, Faulker, Naish, and Brogan (1990) developed, as a by product of their work which was actually focused on something slightly different, an innocuous method of apparently changing the pacemaker rate of an internal clock in humans, the use of repetitive stimulation in the form of clicks or flashes.

Penton-Voak, Edwards, Percival, and Wearden (1996) used a variant of Treisman et al.'s method, and showed that people estimated the duration of both auditory and visual stimuli as longer when these were preceded by clicks than when presented alone. Furthermore, the click effect was multiplicative, that is, greater at longer durations than shorter ones, exactly the effect predicted by an increase in pacemaker speed. Penton-Voak et al. also demonstrated that, while the subjective duration of stimuli estimated was lengthened by clicks, intervals produced by the subjects themselves were shortened. This again is consistent with the idea that humans possess an internal clock: If some duration produced is normally associated with  $x$  ticks of the internal clock, then when the pacemaker is speeded up the  $x$  ticks required will be reached in a shorter real time, thus shortening intervals produced. Burle and Bonnet (1999) and Burle and Casini (2001) have also demonstrated this shortening of intervals produced by click trains.

A recent article by Droit-Volet and Wearden (2002) used repetitive flicker to apparently speed up the internal clock of children as young as 3 years of age. The experimental details are too complicated to be described here, but once again a multiplicative effect of flicker, consistent with a change in pacemaker speed, was obtained. As Droit-Volet and Wearden (2002) note, obtaining such an effect in young children implies that the effect is a very “primitive” one, operating on some fundamental, perhaps even biological, level, rather than affecting complex response strategies.

Internal clock models consistent with the pacemaker-accumulator model of SET have also recently been used (Wearden, Edwards, Fakhri, & Percival, 1998) to account for the well-known modality effect in timing, that “tones are judged longer than lights:” In other words, auditory stimuli appear to last longer than visual ones of the same real duration. Once again, the modality effect was greater at longer durations than shorter ones, indicating a pacemaker speed difference between the conditions.

It may be that one day the physiological mechanism of human timing will be uncovered (although my personal view is that this discovery is not imminent), and the idea of a pacemaker-accumulator internal clock shown to be invalid. However, one might say that if people don't have a pacemaker-accumulator internal clock, they certainly *behave* as if they do, so any model which replaces this clock idea will have to account for the data which seems to support pacemaker-accumulator clocks so compellingly.

## Progress and challenges

In the second section of this article, I will discuss some areas where ideas compatible with SET have stimulated research, but in which serious outstanding problems remain. These relate to the tripartite clock-memory-decision structure of SET, and some critics (e.g., Staddon & Higa, 1999) have complained that this structure allows SET too much flexibility in fitting data, rendering it almost undisprovable.

An important aspect of SET is that it has stressed how behavior on all timing tasks results from more than just properties of an internal clock. Even if some clock produces “raw” representations of time, these must be temporarily stored and compared with various sorts of standards by using some decision process: Behavior does not necessarily directly reflect the properties of underlying representations of time. But, as I pointed out elsewhere (Wearden, 1999), it is difficult to know when the fit obtained by varying the memory and decision mechanisms of SET so that the predictions of the model converge on data is a reasonable one, as different memory and decision process could always be posited.

To partly address this charge of undisprovability, Wearden (1999) suggested the program of systematically trying to “isolate” the different components of the SET system, to see if they behaved as standard versions of SET require. Part of the program of isolating the internal clock component had already been carried out with animals (e.g., Meck, 1983), and the section above showed how definite evidence for the sort of pacemaker-accumulator clock proposed by SET had been obtained from studies with humans. The other parts of the SET system have received far less attention.

The memory level of SET consists of two memory stores, one is a short-term *working memory*, assumed to represent the contents of the accumulator more or less accurately. The second is a *reference memory*, which stores “important” times necessary for the task in hand.

Considering the working memory first, it is currently unclear what relation the working memory of SET bears to other accounts of working memory, such as those proposed influentially by Baddeley (1986). A second problem is that there is some evidence that short-term memory of the duration of events is susceptible to an unusual form of forgetting, “subjective shortening,” a tendency for the duration represented to become subjectively shorter (rather than to degrade randomly) as the memory ages. Spetch and Wilkie (1983) popularized this effect using data from pigeons, but Wearden and Ferrara (1993) and Wearden, Parry, and Stamp (2002) have demonstrated it in humans and in unpublished studies with elderly people, children, and patients with Parkinson’s Disease. I have replicated their effects with non-student subject populations. However, in spite of the apparent robustness of subjective shortening in humans it is currently unclear (1) why subjective shortening occurs at all, (2) how it relates to “conventional” short-term memory (see discussion in Wearden et al., 2002), and (3) how it might affect performance on the sort of tasks used to evaluate SET.

The reference memory component of SET plays a more important role than working memory in most versions of SET. This is mainly because Gibbon, Church

and Meck (1984) proposed that the reference memory is the source of the scalar property (conformity to Weber's law, constant coefficient of variation, etc.) so often observed in timed behavior. So, for example, data from temporal generalization in rats superimposes when plotted on the same relative scale when the standard duration is varied over values of 2, 4, and 8 s because the reference memory representations of 2, 4, and 8 s have scalar variance (Church & Gibbon, 1982). This contrasts with the commonsense idea that the raw time representations produced by the internal clock itself might have scalar variance, possibly caused by variations in pacemaker speed from trial to trial. Lack of space prevents a full discussion of the issues here (although Jones & Wearden, submitted for publication, provides one), but some recent studies with humans (Allan & Gerhardt, 2001; Wearden & Bray, 2001) have provided simple demonstrations of the existence of the scalar property of behavior in situations where it is unlikely that reference memory was ever used by the subjects (particularly Wearden & Bray, 2001, Experiment 3, where the use of reference memory seems impossible). This does not, of course, prove that reference memory is not a source of scalar variance when it is used, only that scalar variance can apparently arise from other sources.

The issue of what is actually stored in reference memory is also problematical. Timing experiments with animals often involve extensive training with "important" times, e.g., those associated with reinforcement, and it was natural for early work to assume that the repeated presentations of these important times were individually stored, each being slightly distorted, to form an extensive "memory distribution" in reference memory. In tasks with adult humans, training lasts seconds or minutes at most, so it seems unlikely that such a memory distribution could be formed. In addition, there is evidence from experiments with animals that once trained on some critical duration,  $t_1$ , animals can shift extremely rapidly to some new duration,  $t_2$ , suggesting that the "reference" memory can be rapidly overwritten (Higa, 1996; Lejeune, Ferrara, Simons, & Wearden, 1997).

Jones and Wearden (in press; submitted) attempted to manipulate the number of distinct "items" in reference memory in humans, using methods too complicated to be described here. Their conclusion was that temporal reference memory in humans, and possibly also in animals, did not consist of any extensive distribution of stored time values, but possibly only of a single "standard" which had upper and lower limits. So, for example, when resending a person with instances of a 400-ms "standard" resulted in the following processes. The first 400 ms value,  $s$ , was multiplied by a constant with a mean of 1.0 and some variability, and the resulting value  $s^*$  stored. A subsequent presentation of the 400-ms standard would generate another potential representation of the standard, call this  $s'$ . If  $s'$  was within some percentage of the original  $s^*$ , say within 10% above and below, then essentially nothing happened, but if  $s'$  was outside this range, the reference memory was "perturbed" as  $s'$  was substituted for  $s^*$ . This "perturbation model" not only fitted data from humans well, but also accounted for many effects in experiments with animals.

Whether or not the perturbation model is correct, it serves to focus interest on the important issue of what temporal reference memory contains and how it changes

when the duration to be timed changes, but remains constant when the duration remains constant.

Many other unresolved issues centre around the role of decision processes in the SET model. As Wearden (1999) pointed out, the decision processes specified for a particular timing task are usually conjectured ad-hoc, albeit in a very plausible way, and will necessarily vary from one task to another. One problem is that many different models can fit the same data more or less equally well. For example, the MCG model outlined above works well for temporal generalization gradients in humans, but even the original paper proposing it (Wearden, 1992) showed that models assuming that the threshold had no variance, or that variance was presented in the just-presented duration,  $t$ , as well as the representation of the standard, worked just as well. Of course, these models are small variants of the original MCG model, but sometimes models which differ more radically (and which have quite different psychological implications) are hard to distinguish.

Bisection also furnishes a number of examples. On a bisection task, people initially receive examples of standard *short* and *long* durations ( $S$  and  $L$ , e.g., tones 200 and 800 ms long). They then have to make judgments about other durations,  $S$  and  $L$ , and values in between, usually in terms of their similarity to  $S$  or  $L$  (e.g., “was that duration more similar to the standard *short* or the standard *long* duration”). SET-compatible models developed by Allan and Gibbon (1991) and Wearden (1991b) both fitted the data well, but employed different rules for judging the similarity of some duration  $t$  to  $S$  and  $L$ . However, worse was to come, as Wearden and Ferrara (1995) suggested that people were not actually making any kind of comparison between  $t$  and  $S$  and  $L$ , but instead comparing each  $t$  with the *mean* of all the durations experienced, including  $S$  and  $L$ , but awarding no special status to these “standards.” Wearden, Bajic, and Brocki (submitted for publication) have recently shown that a model incorporating a variant of this idea fits the overall body of data on temporal bisection in humans (derived in their article from 77 data sets) better than any published competitor. So, what is actually happening in temporal bisection, what is being used as a “reference” for timing decisions made, and what sort of decisions are operative? These questions still have no definite answers.

In psychology, we are so accustomed to theories accounting for data poorly that we are embarrassed when a theoretical system like SET generates not one model that fits data excellently, but any number of them! How can we decide between the many possibilities that changing the memory and decision processes of SET can throw up? Perhaps the only way is to look at variants of experimental procedures, and to see if the models proposed predict behavior correctly when some aspect of the experimental situation is changed. For example, Wearden and Ferrara (1995) showed that changing the spacing of intermediate durations in bisection (from linear to logarithmic spacing, for example) between  $S$  and  $L$  changed bisection performance (as did Allan, 2002). Although virtually any model can fit any single set of bisection data, it is not the case that all models can simulate such stimulus spacing effects. In fact, any model which simply assumes that the decision about some time value,  $t$ , depends simply on a comparison between  $t$  and  $S$  and  $t$  and  $L$  cannot in principle simulate such effects without adding additional factors to the model.

In this vein, Wearden and Grindrod (2003) tried to manipulate the decision processes made in temporal generalization by encouraging or discouraging people to make YES responses by the award and deduction of points. In their study, durations remained identical between different conditions, and the manipulation was intended to change decision processes alone. Unsurprisingly, the manipulation worked: people made more YES responses overall (particularly to durations close to the standard duration) when the contingencies encouraged this, but the central question is how the SET-based model which fits temporal generalization in humans, the MCG model discussed earlier, modelled data. If the standard interpretation according to SET is correct, Wearden and Grindrod's manipulation should change the response threshold, leaving everything else unchanged, and this prediction was supported reasonably well, in that the main effect was a change in the threshold value, rather than the coefficient of variation of memory representation of the standard.

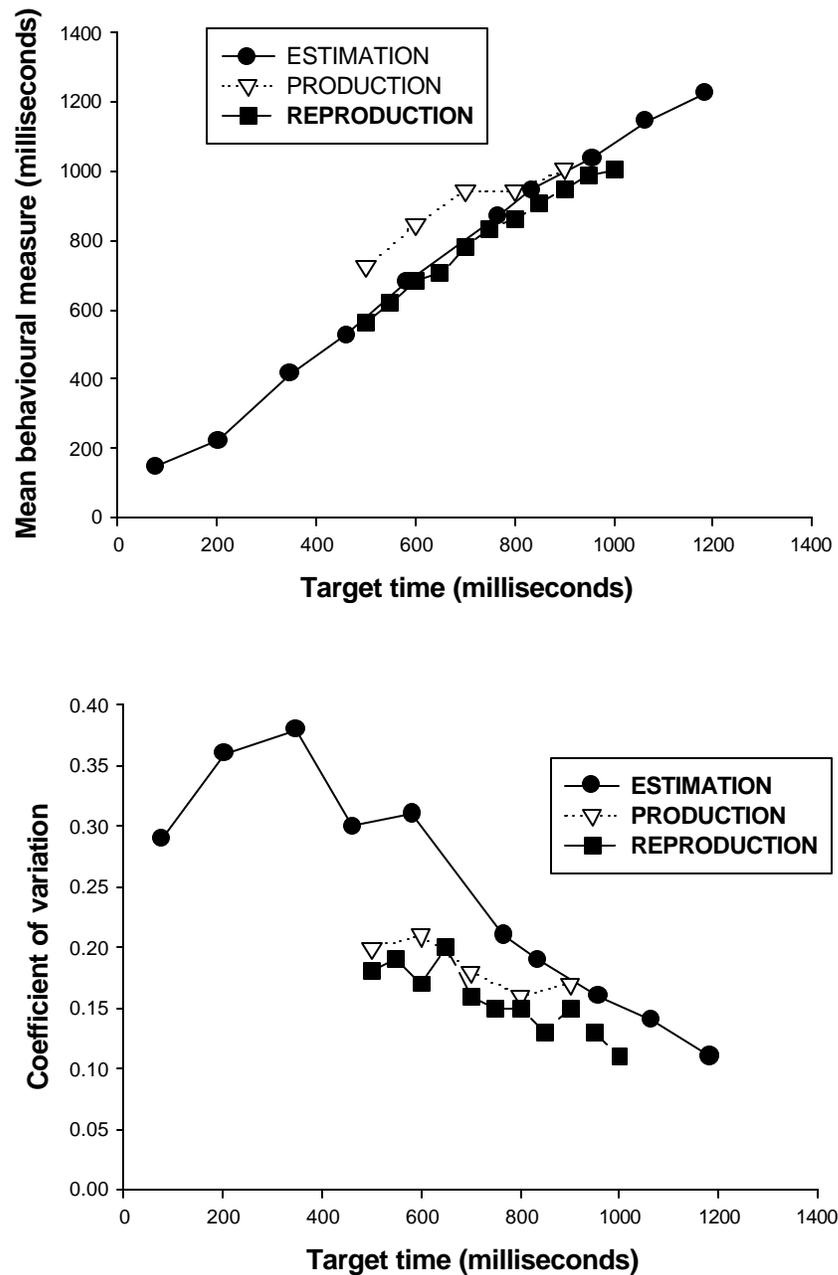
In spite of the success of this simple manipulation, much more work needs to be done on variants (both small and large) of standard SET-compatible tasks like temporal generalization and bisection before we can be completely sure as to what people are doing when they perform on these tasks. Just because a simple model fits the (equally simple) data well does not mean that the model fitted is correct.

## Challenges

There are some areas in which SET remains either untested, or highly problematical, even within the domain of the laboratory-based timing of short durations. I will mention three, but concentrate only on the last one, a personal favorite. The first challenge is to link SET with more conventional cognitive psychology, in particular that part which deals with memory and attention. Numerous studies have shown that attention plays a critical role in timing behavior (e.g., Brown & West, 1990; Macar, Grondin, & Casini, 1994), but it is unclear how what is basically a mechanical system like SET can accommodate the notoriously tricky concept of attention. SET, reflecting its roots in behaviorist psychology, regards animals, and humans, as processing systems making timing decisions according to the operation of a specified internal clockwork. The subject is just a vessel responding to stimuli and processing them to produce output. This characterization of SET, while perhaps bleak, is not intended to be in the least pejorative, as personally I completely agree with this conception, but how the ghostly concept of "attention" can be defined so as to be incorporated into quantitative models remains hard to imagine.

Another major challenge is to try to understand the physiological processes underlying timing. Recent articles (e.g., Mattel & Meck, 2000) have criticized the idea of a pacemaker-accumulator clock on physiological grounds. As I note above, timing behavior often conforms well to the mathematical predictions of a pacemaker-accumulator clock, so if it is physiologically implausible its successor must be able to account to data which seem so consistent with it. Oscillator-based models (Church & Broadbent, 1990; Mattel & Meck, 2000; Wearden & Doherty, 1995) are said to have the cachet of great physiological plausibility that the pacemaker-accumulator clock,

but whether then can predict behavior as well as the much-maligned clock remains to be seen.



**Figure 2:** Upper panel: Mean measures of behavior plotted against target duration from 80-83 subjects performing on verbal estimation of tone durations, production of intervals without feedback, or temporal reproduction. Lower panel: coefficient of variation (standard deviation/mean) of measures of behavior from verbal estimation, production, and reproduction

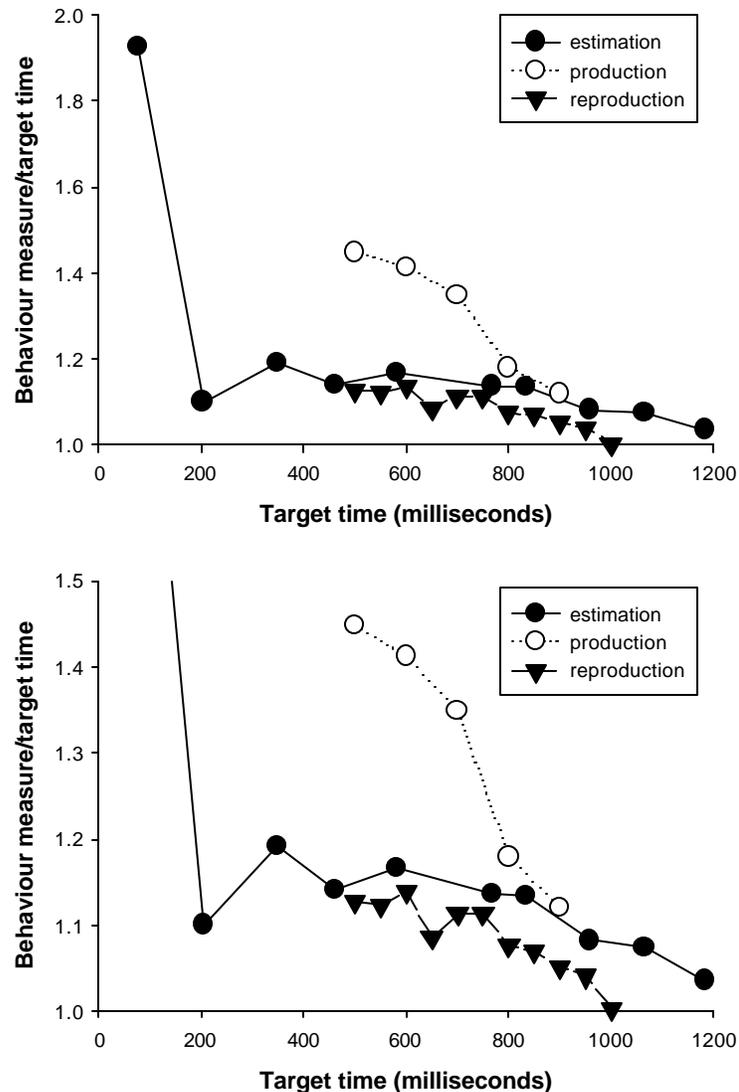
My choice of a problem for the remainder of this section is neither of the two previously-mentioned, but rather than application of SET to “classical” timing procedures, in particular the production, reproduction, and verbal estimation of duration. It is perhaps ironic that while simple SET-compatible models can deal with what appear to be quite complicated experimental procedures (such as categorical timing, which involves making judgments about 18 or 24 different durations, see Wearden, 1995), we have few or no properly-developed models of how people produce, reproduce, or assign verbal labels to durations, in spite of the fact that such procedures have been used since the earliest days of time psychology (Fraisse, 1964). In fact, data from such procedures seem at first sight clearly incompatible with SET. Figures 2 and 3 show data from an experiment where 80-83 female undergraduate students (the number depending on the procedure used) produced short durations without feedback, reproduced durations, or assigned verbal labels to them.

The upper panel of Figure 2 shows the mean measures of behavior plotted against target time. Close to linear relations are evident in all the three cases, but the lower panel of Figure 2 shows that the scalar property of constant coefficient of variation (standard deviation/mean) is violated, sometimes quite severely, in data from all three procedures. Figure 3 shows the behavior measures expressed as a fraction of the target, with the lower panel showing the same data as the upper one but on a more sensitive scale. Obviously, measures of timed behavior from all three procedures exceeded target times at short target times, but became closer to targets at longer times. This is a form of Vierordt’s law, the principle that short durations are “overestimated” whereas longer durations are “underestimated” with some indifference point between these two values. The data in Figure 3 never show “underestimation,” although some other data from reproduction, to be discussed later, do.

Behavior on all three classical timing tasks thus appears to conform to neither of the properties of behavior required by SET by violating mean accuracy (and instead showing Vierordt-like effects), and also violating the scalar property of variance by showing decreasing coefficients of variation as the intervals timed get longer. Given that mean accuracy and the scalar property are very commonly found in human timing (e.g., Wearden, 1992, 1995; Wearden et al., 1997b; Wearden, Bajic, & Brocki, submitted) this state of affairs is surprising, even shocking. Can SET rise to the challenge of modelling data which seem at first sight to violate it in important ways?

In order to model data from the “classic trio” of timing tasks, some very difficult problems need to be solved, at least in part. One of these, relevant both to production and verbal estimation, is the problem of “scaling,” that of how conventional time units such as seconds or milliseconds are related to subjective time, in both directions (time units to behavior in the case of production, subjective duration to time units in the case of verbal estimation). Another problem relevant to verbal estimation is that of “quantization,” the fact that when people make judgments (e.g., in ms) of the duration of short stimuli, they do not use all possible values (in fact, around 90% of the estimates end in “00,” and almost all the rest in “50”). Although we know that quantization occurs, we have no detailed models of how it is done, or how quantization affects time estimates produced. For example, one attractive possibility is that “raw” representations of stimulus durations exhibit both mean accuracy and the scalar

property, but some quantization effect intervenes to produce the non-scalar character of estimates shown in verbal estimation data. This possibility is easy to propose, but producing an accurate model of verbal estimation may be difficult, as many different means of quantizing verbal estimates can be imagined and simulated.



**Figure 3:** Mean behavior measures shown in Figure 2 divided by target time. Lower panel shows same data as upper one on a more sensitive scale

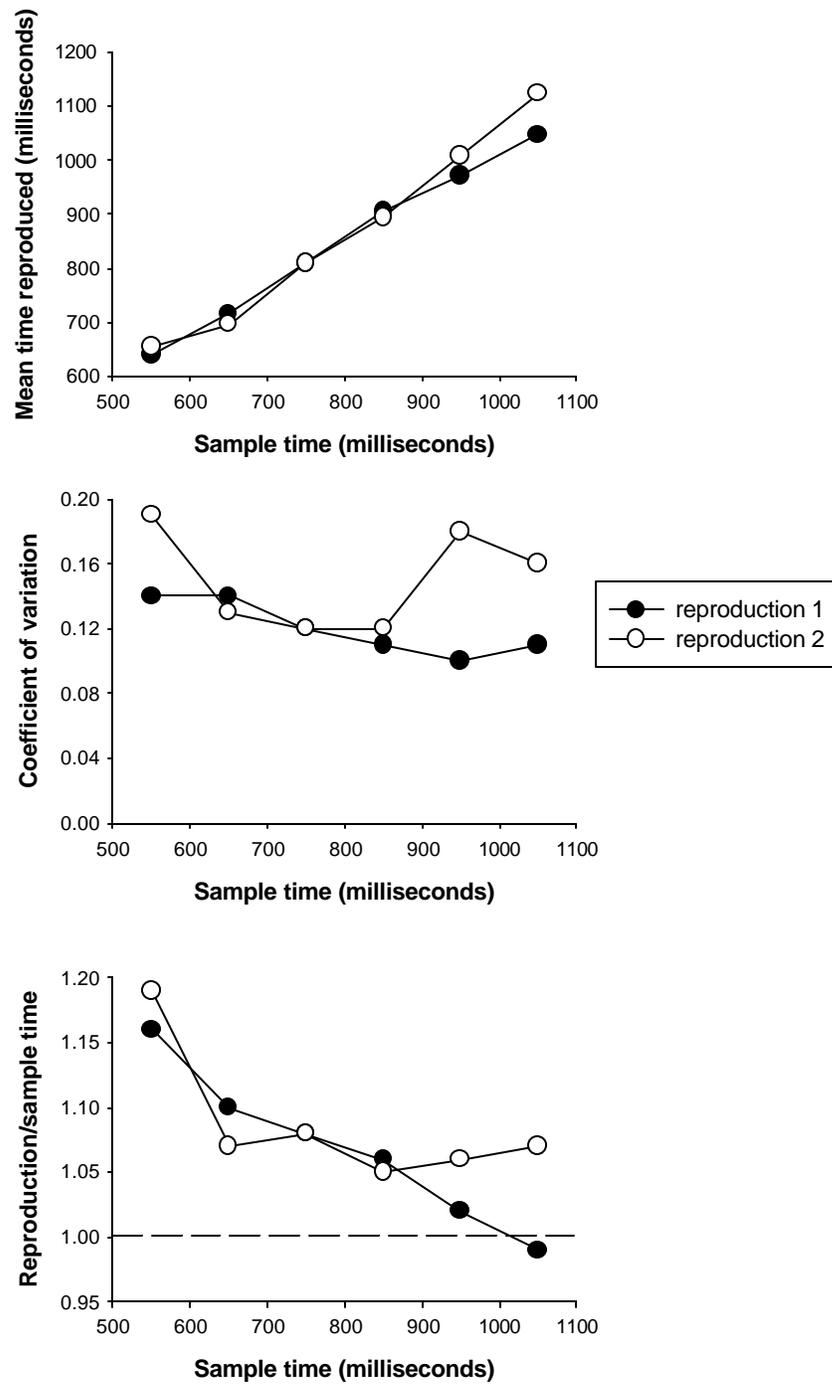
The task of temporal reproduction offers a starting point for trying to develop SET-compatible models of performance that seems to violate SET, as people do not use conventional time units in reproduction (so the scaling problem is [probably] avoided), and the response, being motor, is not quantized in the same way as verbal estimates are. For the rest of this chapter, I will sketch out a potential model of temporal reproduction that reconciles performance with principles of SET.

First, some definitions. On temporal reproduction we will suppose that the subject initially receives some sample duration,  $s$ , in the form of a presented stimulus. The task is then to make some motor response, what I will call the reproduction of  $s$ ,  $r$ , so that  $s$  and  $r$  are equal so far as the subject is concerned. The method used in my laboratory is what I call “reproduction by waiting” and involves just a single motor response. So, for example,  $s$  is defined by the time between two brief clicks, then a short random gap follows, then a third click. The subject’s task is to wait after the third click for a time equal to  $s$ , the sample.

Figure 4 shows data from two independent replications of reproduction of sample durations ranging from 550 to 1050 ms. The upper panel of Figure 4 shows the mean times reproduced, which clearly vary linearly with sample duration, the centre panel shows that (unlike the reproduction data presented earlier) the coefficient of variation remained more or less constant, or declined only slightly, with increases in sample duration, and the bottom panel shows a Vierordt-type plot, where mean time reproduced is divided by the sample time.

Data from reproduction (Figures 2, 3 and 4) suggest in fact that  $r$  and  $s$  will not generally be the same, and that  $r > s$  when  $s$  is small, and closer to it when  $s$  is longer (essentially Vierordt’s law). To model reproduction, I will follow ideas used to explain repetitive tapping in the well-known model of Wing and Kristofferson (1973). The essential point is that the measured reproduction,  $r$ , is actually generated by two consecutive processes. The first process involves waiting until the elapsed time (from the third click) is “close enough” to  $s$ , at which time a response is initiated. Making the response, which takes some time,  $d$ , is the second process. So, the total reproduction,  $r$ , is made up of the time of response initiation ( $t$ ) plus  $d$ . Given that the subjects in my experiments did not receive feedback, they had no way of knowing what the relation between  $r$  and  $s$  actually was.

There are various ways of implementing the ideas sketched above. A particularly simple version was incorporated into a computer model, which simulated the reproduction of target durations ranging from 500 to 1000 ms. On each trial, the sample duration,  $s$ , was multiplied by a random value drawn from a Gaussian distribution with mean 1.0 and coefficient of variation,  $c$ , to produce an effective standard for the trials,  $s^*$ .  $s^*$  varied from trial to trial, but was on average accurate (mean accuracy) and the multiplicative transform used also gave the representations of the sample the scalar property of variance. Thus, the basic representation of the time to be reproduced was completely compatible with SET principles. To initiate responding, the model ran the internal clock from zero until some time,  $t$ , which was “close enough” to  $s^*$ , in fact until the difference between  $s^*$  and  $t$  was some threshold proportion (in the simulations between 10% and 40%) of  $s^*$ . At this moment, the response was initiated, and the response was assumed to take on average 250 ms to be emitted, with the response time being represented as a value drawn from a uniform distribution running between 150 and 350 ms, the time on each trial being  $d$ . The total reproduction on the trial,  $r$ , was just  $t + d$ , whatever values they took on the trial.

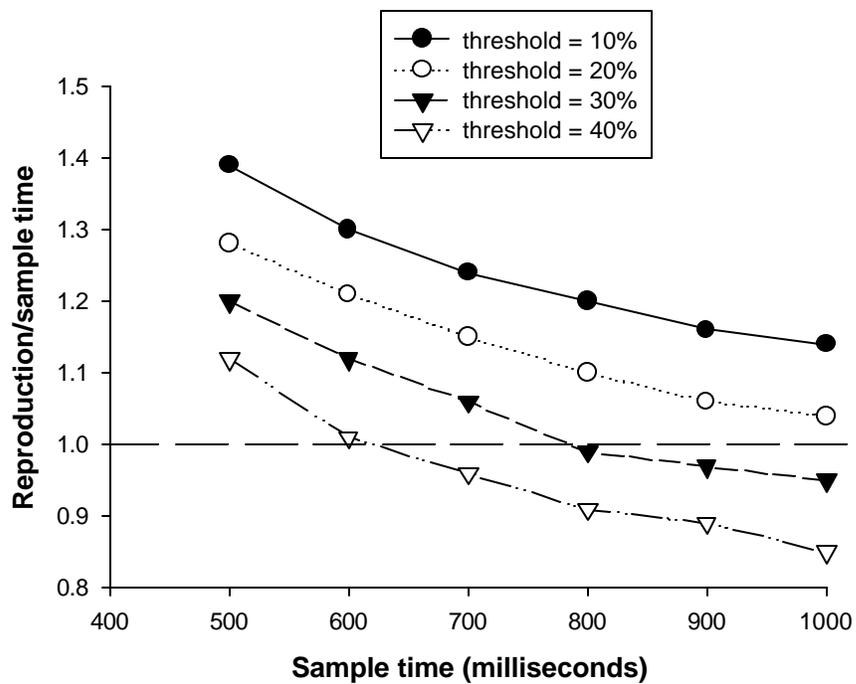
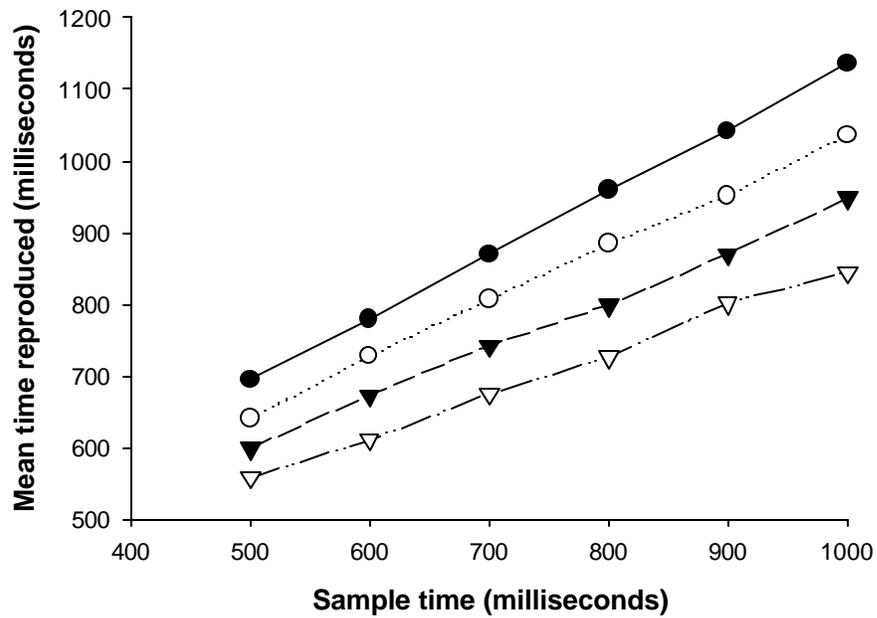


**Figure 4:** Data from two experiments on temporal reproduction. Upper panel: mean time reproduced plotted against sample (target) time. Centre panel: coefficient of variation of reproduced time plotted against sample time. Lowest panel: Mean time reproduced divided by sample time, plotted against sample time

Figure 5 shows some simulations using this model, with 1000 trials at each time to be reproduced being simulated.  $c$ , the coefficient of variation of the memory of the sample duration, was kept at 0.2, and Figure 5 shows the effect of varying the response threshold. The upper panel shows the mean time reproduced as a function of the sample duration. It is clear that in all cases the model produced mean reproductions which grew linearly with sample duration. The lower panel shows a Vierordt-type plot, where the mean reproduction was divided by the target time. Here, values greater than 1.0 show what is (rather misleadingly) referred to as “overestimation” of the target, values less than 1.0 show “underestimation.” Coefficients of variation remained roughly constant with changes in sample duration (thus resembling the data in Figure 4 but not those in Figure 2), but are not shown, to save space.

Two points are obvious from inspection of the lower panel of Figure 5. Firstly, Vierordt-type effects are found whatever the threshold value: mean times reproduced overshoot target times more when the targets are short than when they are long. Secondly, with some threshold values, short targets are “overestimated” while long targets are “underestimated” (e.g., threshold = 30 and 40%). Thirdly, with some threshold values there is an “indifference” point where a mean reproduction would be perfectly accurate: for the 30 and 40% thresholds this indifference point lies between 600 and 800 ms. The existence of an indifference point excited the interest of classical time psychologists to an excessive degree (e.g., Fraise, 1964), and much psychological significance was attributed to it, but the present simulations suggest that the indifference point could be anywhere with choice of an appropriate threshold, and need not imply a profound difference between durations which are “overestimated” and those which are “underestimated.”

It should be acknowledged that this simple reproduction model is not the only one that could be invented, although it is perhaps one of the simplest. In addition, it deals better with situations in which the coefficient of variation of times reproduced remains roughly constant. Marked decreases in coefficient of variation with increases in sample duration pose problems for the model, which can only simulate them by having very high (perhaps unrealistically high) variance in  $d$  from trial to trial. Nevertheless, the model performs the useful function of showing that Vierordt-type deviations from mean accuracy need not imply underlying non-scalar timing processes, but also begins the task of reconciling SET with classical timing data, and opening up a route to explaining why data collected since the earliest days of psychology have the form they do.



**Figure 5:** Upper panel: Mean time reproduced plotted against sample time derived from simulation discussed in text. Lower panel: Mean time reproduced divided by sample time, plotted against sample time, derived from simulation discussed in text

## References

- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*, 101-117
- Allan, L. G. (2002). The location and interpretation of the bisection point. *Quarterly Journal of Experimental Psychology*, *55B*, 43-60.
- Allan, L. G., & Gerhardt, K. (2001). Temporal bisection with trial referents. *Perception and Psychophysics*, *63*, 524-540.
- 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.
- Brown, S. W., & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica*, *75*, 103-121.
- Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to the temporal processing of information. *Behavioural Processes*, *45*, 59-72.
- Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 195-205.
- Church, R. M., & Broadbent, H. (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.
- Droit-Volet, S., & Wearden, J. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *Quarterly Journal of Experimental Psychology*, *55B*, 193-211.
- Droit-Volet, S., Clément, A., & Wearden, J. H. (2001). Temporal generalization in 3- to 8-year-old children. *Journal of Experimental Child Psychology*, *80*, 271-288.
- Fraisse, P. (1964). *The psychology of time*. London: Eyre and Spottiswoode.
- François, M. (1927). Contributions à l'étude du sens du temps: La température interne comme facteur de variation de l'appréciation subjective des durées. *L'Année Psychologique*, *27*, 186-204.
- Gibbon, J., Church, R. M., & Meck, W. (1984). Scalar timing in memory. In J. Gibbon and L. Allan (Eds.), *Annals of the New York Academy of Sciences*, *423: Timing and time perception* (pp. 52-77). New York: New York Academy of Sciences.
- Higa, J.J. (1996). Dynamics of time discrimination: II. The effects of multiple impulses. *Journal of the Experimental Analysis of Behavior*, *66*, 117-134.
- Jones, L. A., & Wearden, J. H. (in press). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *Quarterly Journal of Experimental Psychology*.
- Jones, L. A., & Wearden, J. H. (submitted). Double standards: Memory loading in temporal reference memory. *Quarterly Journal of Experimental Psychology*.
- Lejeune, H., Ferrara, A., Simons, F. & Wearden, J. H. (1997). Adjusting to changes in the time of reinforcement: Peak interval transitions in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 311-331.
- Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory and Cognition*, *22*, 673-686.

- Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, *22*, 94-103.
- Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 171-201.
- Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 307-320.
- Spetch, M. L., & Wilkie, D. M. (1983). Subjective shortening: A model of pigeons' memory for event duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 14-30.
- Staddon, J. E. R., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*, 215-251.
- Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705-748.
- 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. (1995). Categorical scaling of stimulus duration by humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 318-330.
- Wearden, J. H. (1999). "Beyond the fields we know...": Exploring and developing scalar timing theory. *Behavioural Processes*, *45*, 3-21.
- Wearden, J. H. (2001). Internal clocks and the representation of time. In C. Hoerl & T. McCormack (Eds.), *Time and memory: Issues in philosophy and psychology* (pp. 37-58). Oxford: Clarendon Press.
- Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory: Episodic temporal generalization and bisection in humans. *Quarterly Journal of Experimental Psychology*, *54B*, 289-310.
- Wearden, J. H. & Doherty, M. F. (1995). Exploring and developing a connectionist model of animal timing: Peak procedure and fixed-interval simulations. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 99-115.
- 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., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology*, *48B*, 289-310.
- Wearden, J. H., & Grindrod, R. (2003). Manipulating decision processes in the human scalar timing system. *Behavioural Processes*, *61*, 47-56
- 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.
- Wearden, J. H., & Penton-Voak, I.S. (1995). Feeling the heat: Body temperature and the rate of subjective time, revisited. *Quarterly Journal of Experimental Psychology*, *48B*, 129-141.

- Wearden, J. H., Bajic, K., & Brocki, J. (submitted). Temporal bisection in humans: Response strategies, biases, and the location of the bisection point. *Journal of Experimental Psychology: Human Perception and Performance*.
- Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997 b). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 502-511.
- Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why “sounds are judged longer than lights:” Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology*, 51B, 97-120.
- Wearden, J. H., Parry, A., & Stamp, L. (2002). Is subjective shortening in human memory unique to time representations? *Quarterly Journal of Experimental Psychology*, 55B, 1-25.
- Wearden, J. H., Wearden, A. J., & Rabbitt, P. (1997 a). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 962-979.
- Wing, A. M., & Kristofferson, A. B. (1973). The timing of interresponse intervals. *Perception and Psychophysics*, 20, 191-197.