ELSEVIER

Behavioural Processes 45 (1999) 3–21

BEHAVIOURAL PROCESSES

"Beyond the fields we know...": exploring and developing scalar timing theory

J.H. Wearden

Department of Psychology, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Received 20 August 1998; received in revised form 5 January 1999; accepted 5 January 1999

Abstract

The article discusses three areas that appear neglected or underdeveloped in current treatments of scalar timing theory (SET). In particular, questions about where variance in the SET system comes from, and how memory and decision processes operate within SET are discussed. The article suggests a number of possible experiments with humans, some based on pilot work which is described, that may clarify all three areas to some degree. Methods derived from conventional studies of memory are suggested as providing techniques for investigating the operation of memory and decision processes within the SET model, both areas previously considered somewhat inaccessible. In general, the tripartite division of SET into clock, memory, and decision processes is advocated as a useful general framework for studying timing, including questions related to its neurobiological basis, whether or not data always conform to SET predictions, although more needs to be known about how all three parts of the SET system operate.

Keywords: Scalar timing; Internal clock; Memory for duration; Temporal decision processes

1. Introduction

Even its most vituperative critics can surely not deny that scalar timing theory (SET) developed first by Gibbon (1977) then Gibbon et al. (1984), has transformed our understanding of timing processes in animals and (more recently) humans over the last 20 or so years. A notable feature of the more than 100 articles that use ideas from SET is their great diversity. Reports of animal studies that use SET-related analyses include material from performance on conventional reinforcement schedules (e.g. Gibbon, 1977), sometimes in animal species rarely found in the laboratory (Lejeune and Wearden, 1991), or specially designed timing tasks such as temporal generalization (Church and Gibbon, 1982), bisection

---

E-mail address: wearden@fs4.psy.man.ac.uk (J.H. Wearden)

0376-6357/99/$ - see front matter
PII: S0376-6357(99)00006-6

(Church and Deluty, 1977), time-left (Gibbon and Church, 1981), and the peak procedure (Roberts, 1981; Church et al., 1994). Other work with animals, carried out within the SET framework, has looked at effects on timing of drugs (Maricq et al., 1981; Meck, 1983, 1996), brain lesions (Meck et al., 1987), and ageing (Meck et al., 1986; Lejeune et al., 1998). Studies with humans report procedures based on those used in animal studies (Allan and Gibbon, 1991; Wearden, 1991a,b, 1992), which have also been employed to look for effects of ageing on timing (Wearden et al., 1997a) and to examine timing in children (McCormack et al., 1999). Methods and analyses related to SET have also recently been used in studies of patients with Parkinson's disease (Malapani et al., 1998), and cerebellar degeneration (Nichelli et al., 1996). Another recent article has used an SET-compatible clock model to address a 100-year-old problem in human time psychology, why auditory stimuli appear to last longer than visual ones of the same real duration (Wearden et al., 1998), thus opening up the possibility of a redevelopment of much of classical time psychology using modern ideas and methods derived from, or at least compatible with, SET.

The influence of SET extends even to its competitors, such as the *Behavioural Theory of Timing* (Killeen and Fetterman, 1988) or the Zakay/Block model of retrospective and prospective time judgements (Zakay and Block, 1995), which in many cases effectively incorporate the same processes as SET proposes (even if differently named), and usually make nearly identical predictions.

Although the successes of SET are numerous, the present article is almost exclusively concerned with issues that seem, at least to me, unjustly neglected or completely ignored, or cases in which the treatment offered by SET appears in one way or another unsatisfactory. However, I do not offer a conventional critique, as I have no desire to advance, or support, any alternative theory; rather, my aim is to develop and enrich SET by attempting to stimulate interest in a number of what I see as important unresolved problems. The present article differs from a normal critique in two other ways: firstly, when I complain about obscurities or incoherence in the treatment of some problem, it is more than likely that I myself have been guilty of committing, in published work, the very offences complained about; secondly, I intend wherever possible not only to identify problems, but also to suggest some potential solutions. These mostly take the form of proposed experiments, sometimes based on unpublished work conducted by myself and associates, sometimes designs for experiments of a new and as yet unperformed type, but all are methods which seem to me promising lines of attack on a number of problems, sometimes very difficult ones. Some suggested experiments are conventional variations on procedures that are commonplace in the literature; others take a novel approach, perhaps disconcertingly novel and speculative for some readers, by mixing ideas derived from areas of cognitive psychology, not usually linked with SET, with those from normal timing theory. However, to find out new things, we may need new methods.

The experiments all use, or are intended for, human participants, where speed of learning and the ease by which behaviour can be manipulated by instructions enables very rapid and efficient data collection. However, the issues treated apply to SET in general, and procedures to address some of the questions raised using animal subjects could probably be devised, albeit with considerable difficulty in some cases. All the real or imaginary studies with humans use judgements of short durations (less than, or around, 1 s), where non-counting-based timing is usually guaranteed, and SET is known to fit data well (Wearden, 1991a). However, some recent results suggest that judgements about longer durations by humans also conform to scalar timing theory providing that counting is discouraged or prevented (Wearden et al., 1997b; Rakitin et al., 1998), so some of the procedures outlined could probably be duplicated, using more complicated methods, with longer durations.

The article is arranged in four main sections, followed by some concluding remarks. I first provide an introduction to SET-type models, which serves as an aide-memoire to readers but is no means comprehensive. In fact its main function is to highlight the issues addressed in the three sections that follow, and these are: *Where does variance come from? Properties of memories for time* and *how do decision processes work?*

2. Introduction to SET-type models

The term SET-type models is intended to emphasize that almost all of what follows subsequently in this article applies more generally than just to SET itself, in fact, the issues are ones which need to be addressed in any model proposing an internal clock process, whether or not clock operations follow the prescriptions of SET, where a clock stage is supplemented by various types of memory representations of duration, and where behaviour is the result of comparison processes, whether or not the memory and comparison proposed follow those suggested by SET.

Fig. 1 shows an outline of the SET model. It is divided into three stages, which are usually considered sequentially. The first stage involves a pacemaker connected to an accumulator which stores pacemaker pulses via a switch. The contents of the accumulator, which represent the 'raw material' for time judgements, can be downloaded to a short-term memory (STM) containing a more or less direct copy of accumulator contents or, in other versions of SET, the accumulator and STM can be combined. Another longer-term memory for duration (LTM) contains a memory representation of 'important' times, such as those associated with reinforcement in animal experiments, or various sorts of standard durations used in studies with humans. The STM/LTM system thus represents a memory stage in the SET model. Finally STM and LTM contents are compared in a decision stage, and overt behaviour of various sorts is produced depending on the results of the comparison.

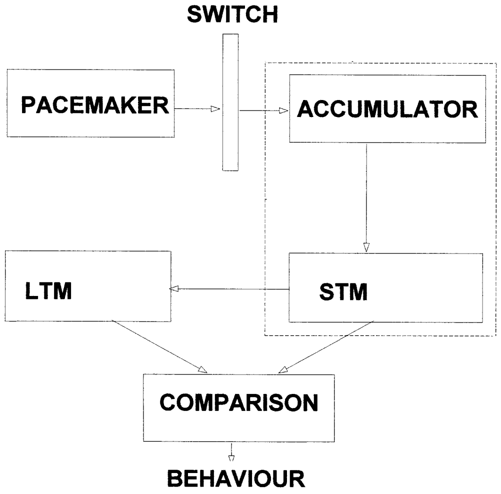

Fig. 1. Schematic diagram of SET-type systems discussed in the text. LTM/STM are long- and short-term memories for duration, respectively. The accumulator and STM are sometimes considered as a single system, and are boxed accordingly with dashed lines. Arrows show direction of information flow in the model.

The account of data obtained from the method of *temporal generalization* (Church and Gibbon, 1982; Wearden, 1992; Wearden et al., 1997b) can simply illustrate the operation of the whole system. In this procedure, subjects are initially presented with examples of a standard duration, e.g. a tone 400 ms long (Wearden, 1992), then subsequently receive presentations of stimuli, longer, shorter or equal in duration to the standard, and have to respond as to whether each stimulus presented had the standard length (usually by responding Yes or No; i.e. standard or not). The proportion of Yes responses plotted against stimulus duration usually yields, when human adults participate in the experiment, a temporal generalization gradient which (i) peaks at the standard duration (i.e. the standard produces the maximum proportion of Yes responses), and (ii) is skewed to the right, such that durations longer than the standard tend to be more confused with it than durations shorter by the same amount (i.e. 500 ms results in more Yes responses than 300 ms in the example given).

The standard treatment of such results, derived by Wearden (1992) from Church and Gibbon (1982), illustrates the operation of all levels of the SET system. When a stimulus is presented, the switch connecting the pacemaker to the STM/accumulator closes, allowing pulses to flow. The switch operation need be neither instantaneous with stimulus onset (or offset) nor constant in latency from trial to trial. When the stimulus terminates, the switch opens and the pacemaker/accumulator connection is severed. The accumulator now contains the 'raw material' for judgements about stimulus duration. The value in the STM/accumulator is compared with a sample drawn from the LTM of the standard. This LTM is proposed to contain a distribution of values which has (i) an accurate mean (so that, for example, the 400 ms standard is neither systematically under- or over-estimated) and (ii) scalar variance, in that the standard deviation of the distribution in memory is a constant fraction, c, of the mean. The STM contents and the LTM sample are compared in the decision stage, and if they are 'close enough', the participant responds Yes, otherwise he or she responds No. The comparison rule used by Wearden (1992), which is compatible with the asymmetrical temporal generalization gradients found in experiments with humans is respond Yes when

$$|s^* - t|/t < b^*$$

where s^* is the sample drawn from LTM, t is the just-presented duration, and b^* is a threshold value which is variable from trial to trial.

Some details of the clock, memory, and decision processes will be discussed later, for present purposes the outline above illustrates a number of features. (i) Observed behaviour results from a combination of clock, memory, and decision processes, so does not reflect any of these processes in isolation in a direct way. (ii) The model has a number of potential sources of variance (discussed more fully later), such as might arise at the clock, memory, or decision stages; thus, for example, if one subject group produces more variable performance than another, the source of variance cannot be immediately identified. (iii) Decision processes play a vital role in accounting for behaviour; if different decisions are postulated, then predictions about observed behaviour will probably change, even though clock and memory processes might remain identical. Point (iii) can be illustrated by comparing temporal generalization performance in humans (discussed above) with that of rats, derived from Church and Gibbon (1982). Rats produce nearly completely symmetrical generalization gradients, with durations equal temporal distances above and below the standard producing nearly identical response levels. This animal/human difference is not, apparently, due to the different stimulus values used, as when Wearden et al. (1997b) exposed humans to the same time values as rats, their gradients were asymmetrical, much as they were with shorter durations. The animal/human difference can be modelled by just a small alteration in the decision process, in fact by proposing that animals respond when

$$|s^* - t|/s^* < b^*$$

with all terms being identical to the equation that fits data from humans.

The importance of decision processes in SET has advantages and disadvantages. One advantage is that different tasks can be treated by changing the decision process, while keeping the basic time representations produced by earlier stages of the model the same. For example, in the task of *bisection* (derived by Wearden, 1991b from Church and Deluty, 1977), people are initially presented with two standard durations, one short (S) and one long (L) (e.g. 200 and 800 ms: Wearden, 1991b). A range of durations (including the short and long standards) is presented and participants must classify each one in terms of its similarity to S and L. There are various models of bisection, but data from humans can be approximately modelled just by assuming that people respond Short when $t\text{-}S^* < L^* - t$, and Long otherwise, where S^* and L^* are samples from the LTM representations of S and L. Clock and memory operations are assumed to be identical in temporal generalization and bisection, only the decision processes differ. It should be acknowledged that other models of bisection exist (Allan and Gibbon, 1991; Wearden and Ferrara, 1995; McCormack et al., 1999), but the exposition presented above is sufficient to illustrate the point made here.

A disadvantage of the flexibility of decision processes on different tasks in SET treatments is that it is likely that some SET-compatible model can be devised by choosing appropriate decision processes (and these may be much more complicated than proposed above, incorporating biases towards one response or another, or 'random' responses, for example) *almost no matter what the data look like*. This obviously poses problems of testability, as it becomes unclear whether any particular data set is incompatible with SET, or whether the problem is that an 'inappropriate' decision rule has been selected for the task.

My solution to problems of ambiguity besetting SET is resolutely non-Popperian: if it fails or is in some way inadequate, it should be developed and strengthened, not abandoned. This attitude does not stem from a naive belief on my part that SET is literally true, or even that it represents the best model of which current psychological science is capable, although competitors will have to perform well to beat it. Rather, the general framework of clock, memory, and decision processes seems to me an excellent basis for pursuing many currently-interesting questions about how humans (and animals) perform on a wide range of timing tasks, and whether the characterisation of these processes given by SET is the best one available is an empirical question open to experimentation and theoretical development.

In the sections below, I address some issues which are intended to strengthen our understanding of how SET-type models work. The experiments and approaches suggested are tentative, although in some cases pilot data (often taken, alas, from experiments not originally intended for the use to which I have put them here, and thus not as direct as would be ideal) point the way.

3. Where does variance come from?

When a human or animal makes some judgement of a consistent duration t, the judgement will usually vary from trial to trial, that is, something in the timing system contains variance. In most experiments, this variance has the *scalar property* that gives SET its name: that is, as t varies variability of judgements around t vary such that the standard deviation of judgements is a constant fraction of the mean, so the *coefficient of variation* statistic (SD/mean) remains constant as t varies. Some experimental manipulations, such as the use of brain-damaged animals or humans, administration of drugs, or some procedural manipulation with intact subjects, may make this variance change: it may vary in absolute value compared with control conditions or subjects, or it may cease to exhibit the scalar property. Interpretation of such effects obviously depends on some prior knowledge of how variance in the system is generated, and the current section concentrates on this particular problem.

A difficulty with the SET model, where clock, memory and decision processes all interact to determined timed behaviour, is that variance can be present in any or all of these subsystems, and in many cases the resulting behavioural predictions are either identical or difficult to discriminate by almost any practicable observations because of the very small differences between them. Goodness of fit of models which incorporate variance in different parts of the system can likewise by very nearly identical. As mentioned above, Wearden (1992) modelled temporal generalization performance in humans by assuming that participants judged a just-presented duration to be the standard when

$$|s^* - t|/t < b^*$$

where s^* was a sample from the reference memory of the standard s, assumed to have an accurate mean and scalar variance, t was the just presented duration, assumed to be timed without error, and b^* was a sample from a variable threshold. Here, there are two sources of trial by trial variability, the reference memory and the threshold. However, I also showed that an equally good fit to data was obtained by keeping the threshold constant, or by introducing variance into t as well. Although such changes changed the actual parameter values of the fitted models, the different versions could not readily be distinguished by how well they fitted obtained data.

In a more general treatment of the problem of types and sources of variance, Gibbon (1992) showed that small amounts non-scalar variance from whatever sources would make little contribution to deviations from strict scalar timing provided that larger scalar sources were present.

Given that mathematical analysis and computer model construction seem unreliable as indicators of where variance lies in the system, can some experimental decomposition of variance within the scalar model be attempted? In general, a problem is that in most tasks the whole system comes into play, so methods have to be devised to dissect out possible variance sources, even though usually these cannot be studied completely on their own.

I consider first the pacemaker-accumulator clock as a source of behavioural variance, and follow the usual assumptions that the clock has three components, a pacemaker, a switch connecting the pacemaker to the accumulator, and an accumulator (which also may serve the role of a short-term memory mechanism) which stores the pulses flowing from the pacemaker. One obvious source of variance is the pacemaker itself. This is usually conceived of as a Poisson emitter, that is, one which emits pulses at an averagely constant rate, while the time between pulses is random (see Gibbon, 1977, 1992, for discussion). By itself, a Poisson emitter will produce non-scalar variance: the variance, rather than the standard deviation, grows in proportion to the mean, so the coefficient of variation declines (i.e. timing becomes relatively less variable) as the interval timed increases. Some critics of SET (e.g. Staddon and Higa, 1999) have complained that the usual finding of scalar variance in behaviour is inconsistent with the postulation of such a Poisson clock, but SET has addressed this issue in a number of ways. On the one hand, one can assume a rapidly-pulsing Poisson clock which produces output with Poisson variance properties but absolutely small variance levels. The scalar property is then derived from some other scalar source, such as memory for standard durations (e.g. Gibbon, 1992). On the other hand, if the rate of the pacemaker varies from trial to trial according to a normally distributed random variable, then the output of the clock itself will conform to scalar timing (Gibbon et al., 1984). To be fair to critics such as Staddon and Higa, SET often remains somewhat ambiguous about which of these two is proposed in any particular case.

The pacemaker is, however, not the only potential source of variance in the clock system. The switch connecting the pacemaker to the accumulator may itself contribute variance to timing. For example, when a stimulus to be timed is present, the switch may close with some variable latency, only to open with some (possibly different) variable latency when the stimulus terminates. Switch latency variance may thus contribute to variance in timed behaviour. By this argument, there is no reason to suppose that switch latency varies with stimulus duration, as it is a function of stimulus onset and offset, not stimulus duration, so the contribution of the switch, if any, should be mathematically independent of duration, that is, non-scalar. Although switch operations appear at first sight very inaccessible, there may be experimental techniques which can alter switch processes. For example, Wearden et al. (1998) used internal-clock-based ideas compatible with SET to examine the old question of why ‘sounds are judged longer than lights’, i.e. the fact, known for 100 or more years, that, in general, auditory stimuli are judged as lasting longer than visual stimuli of the same real duration.

Wearden et al.’s results replicated the venerable auditory/visual difference in average perceived duration, and showed that the difference between mean estimates of auditory and visual stimuli was compatible with the pacemaker for the different modalities running at different speeds. In addition we demonstrated something sometimes reported, although rarely emphasized, in the traditional timing literature, the fact that the visual stimuli appeared not only shorter than the auditory ones, but were also more variable in duration. Where does the auditory/difference in variability come from? One possibility is that it arises from pacemaker processes themselves as, in general, most proposed pacemakers of a Poisson or similar type will produce relatively less variable output as pacemaker rate increases. So, if auditory stimuli appear to last on average longer than visual ones, perhaps this is because the pacemaker speed difference which accounts for the mean difference results in the variability difference too. However, Wearden et al. (1998) showed that this was probably not the case: using manipulations derived from Penton-Voak et al., (1996), originally based on work by Treisman et al., (1990), we significantly increased pacemaker speed without any consequent effect on variance. 'Speeded up' visual stimuli were perceived as on average longer, but not as relatively less variable. Thus, the variability in pacemaker output was obviously not the source of the auditory/visual variance difference, or at least not the main source, and Wearden et al. (1998) suggested that switch latency variance was the likely cause. For example, both the onset and offset of a visual stimulus were probably more variably represented than the onset and offset of auditory stimuli, resulting in a variance difference between the modalities which did not depend on pacemaker speed.

Although it is clear that certain manipulations can potentially detect sources of variance due to pacemaker and other parts of the internal clock, in general the fact that for most timing tasks the clock, memory, and decision processes act in concert makes attributing variance to one part of the system difficult. A possible experimental strategy might therefore be to devise methods in which the whole system is not engaged. For example, can the question of whether scalar variance is derived from the pacemaker itself or long-term memory of standard durations be resolved? A method of attack on this question might be to develop tasks in which long-term memory for standard durations is not engaged, and compare performance on this type of task (particularly questions relating to variance properties of behaviour) with that on another procedure where the same stimuli are judged in a case where long-term memory is involved.

For example, a task involving pair-comparison of durations can, and probably will, be performed without developing a long-term representation of any standard, as this does not help performance in any way. Specifically, suppose that on each trial a participant receives two stimulus durations S_1 and S_2, separated by a short random gap, and has just to judge them as same or different. On each trial, S_1 and S_2 differ, and from the participant's point of view appear so variable from one trial to the next as to be subjectively random. In fact, trial sequences are structured overall so that people are making judgements about the identity of various stimuli compared with a number of repeated standards, such as 200, 400, 600, 800, and 1000 ms, although trials with the different standards are interspersed so that the underlying organization of trials is not evident. Such a method, closely resembling the 'episodic' temporal generalization reported by Wearden and Towse (1994) could yield temporal generalization gradients around each standard stimulus, but the comparisons on each trial are, presumably, carried out solely on the basis of short-term memory of the stimuli involved on the trial. This procedure thus appears to involve pacemaker, switch, accumulator/STM, and decision processes, but not LTM. If the range of intervals proposed above was used, would the scalar property be evident in judgements? If it was, then it cannot possibly come from LTM, as this is not used. In contrast, exactly the same stimuli could be presented as standards and comparisons in conventional temporal generalization, where LTM is, putatively, involved. In both types of experiments, the same comparison stimuli are compared with the same standards: would the behaviour identical or different, and if different, in what way? One possibility is that the scalar property of variance would be observed in conventional temporal generalization, using LTM, but not in pair comparison, thus verifying the idea that LTM representations of standard durations is the source of scalar variance in the system.

Rodriguez-Girones and Kacelnik (1998) have in fact performed a study somewhat similar to that proposed above, using a kind of episodic bisection method (which they call 'roving bisection'). The authors favour the view that the principal source of variance in performance on bisection (at least in humans) comes from short-term representations of durations, produced more or less directly from the internal clock, rather than longer-term memories of important durations. Their method, like that proposed above, is virtually certain to assure that bisection performance depends on short-term representations of time alone. However, unfortunately, the organisation of trials on their task does not allow the conclusion that clock/short-term memory is the main source of variance to be unequivocally supported, although their data are certainly consistent with this view.

Another technique that does not obviously involve LTM for duration, or even comparison

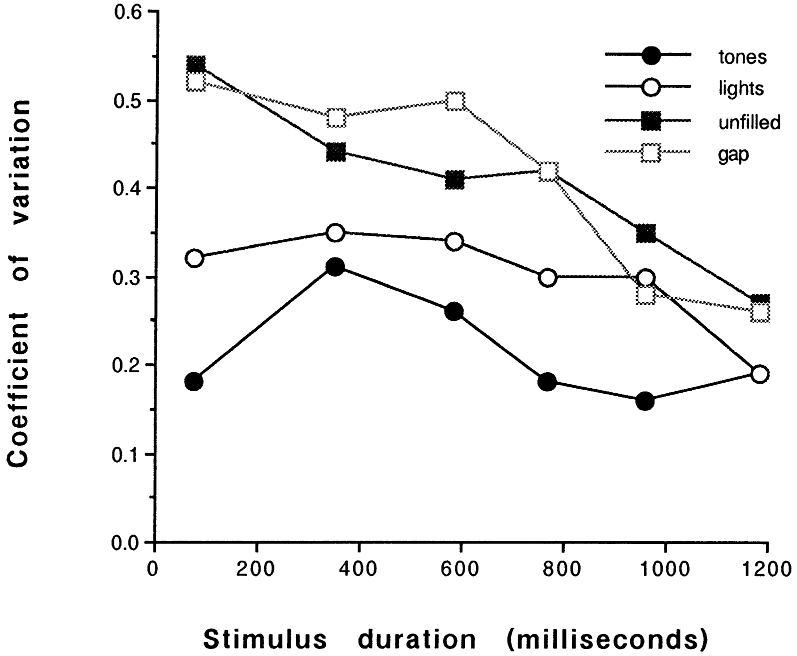

Fig. 2. Coefficients of variation (standard deviation/mean) obtained from studies of verbal estimation of stimulus durations ranging between 77 and 1183 ms, plotted against stimulus duration. The stimuli used were tones (filled circles); lights (squares presented on a computer screen: open circles), unfilled (intervals started and stopped by a brief click: filled squares); gaps (gaps in tones: unfilled squares).

processes are they are usually conceived (i.e. between two or more different durations) is verbal estimation. In this method, presented durations have some conventional verbal label attached to them, such as their estimated duration in milliseconds. Verbal estimation has been recently used in my laboratory, mainly to test hypotheses about whether pacemaker speed differs between various conditions, see Wearden et al. (1998) for the logic of this, but is interesting in its own right. One possibility is that verbal estimation reflects the ‘rawest’ type of time judgement possible in the SET system, and represents a kind of direct scaling of accumulator/STM contents, not involving LTM or comparisons. Although participants routinely express misgivings about their ability to perform verbal estimation, in fact, average estimates from groups are extremely orderly, and highly sensitive to changes in presented duration, as shown in some recent articles (Penton-Voak et al., 1996; Wearden et al., 1998). Mean estimates increase linearly, if not always perfectly accurately, with real-time stimulus duration, supporting the view that verbal estimates reflect the output of some Poisson-like pacemaker process in a reasonably direct way. However, verbal estimates sometimes violate the scalar property of time, as shown in Fig. 2, where data (taken from Wearden et al., 1998, and from unpublished work) come from estimates of continuous tones, visual stimuli (squares presented on a computer screen), unfilled auditory intervals defined by clicks, and gaps in otherwise continuous tones. In most cases, coefficient of variation decreased (i.e. estimates were relatively less variable) as the duration timed increased, with results from unfilled intervals and gaps showing the effect most markedly.

Perhaps data like those shown in Fig. 2 are telling us that the internal clock mechanism of SET produces Poisson-like timing when it operates alone, and that the scalar property of time noted in so many other circumstances comes from elsewhere in the system, as conjectured by Gibbon

(1992) among others. However, this conclusion depends on various assumptions, one of which is that verbal estimation reflects pacemaker/accumulator processes in a transparent way. There are reasons to doubt that this transparency is complete. One of these is that coefficient of variation often dips at the shortest and longest stimulus durations used, consistent with people deciding on a particular verbal label for the shortest and longest durations presented, and consistently using these verbal labels when they believe these particular stimuli have been presented. Another problem is that verbal estimates of durations given by participants are 'quantized' usually into 50 or 100 ms steps in the stimulus ranges used in my experiments (from about 75 to 1200 ms). Thus, people will say that a stimulus has lasted for 200, 250, 300, or 350 ms, but much more rarely for 217 or 323 ms. Such quantization has implications for variance of verbal estimates, but without a detailed model of how verbal estimation is performed it is difficult to say exactly what these implications are.

Overall, however, a combination of different methods, verbal estimation, pair-comparison, and temporal generalization, applied to the same stimulus sets, may help us to begin to identify the nature and extent of variance in the different parts of the SET system. More sophisticated experiments might build on conjectures about how different stimulus types are represented by internal clock mechanisms (e.g. pacemaker speed and switch variance may differ in predictable ways between visual and auditory stimuli). Perhaps questions about where variance in the SET system comes from, although by no means easy to address, are less intractable than sometimes suggested (e.g. McCormack et al., 1999).

4. Properties of memories for time

Although, as discussed above, SET proposes two types of memory for time, a short-term memory representing accumulator contents more or less veridically and a longer-term representation of important times, and although the theoretical treatment of performance on many types of timing tasks routinely proposes that these memories enter into comparison processes, memories for duration per se have been little studied, at least within the framework of SET itself. This is particularly surprising as some evidence suggests that memory for the duration of a stimulus has unusual properties, in that it usually manifests a peculiar form of forgetting called *subjective shortening*. The phenomenon of subjective shortening can be simply described: if a stimulus of duration t is represented in a short term store, increasing the length of its retention leads to the representation of t becoming progressively subjectively shorter, rather than, for example, degrading randomly, so that changes in the direction of representations $<t$ are much more likely than a change towards $>t$. Spetch and Wilkie (1983) introduced the notion of subjective shortening, on the basis of data from experiments with pigeons, although the phenomenon can be noted in earlier work with animals (e.g. Church, 1980). It should be acknowledged that not all forms of memory for duration exhibit subjective shortening, and the reader is referred to Grant et al. (1997) for a thorough review of when the phenomenon occurs.

Wearden and Ferrara (1993) used a procedure similar to the *roving standard* method of classical time psychophysics (Allan, 1979) to demonstrate subjective shortening in humans. On each trial, participants received two brief tones separated by a gap (offset-onset), which was usually either 1, 2, 5, or 10 s. After the second tone, participants had to judge whether the second tone was longer, shorter, or equal in duration to the first. In fact, all three possibilities were equally likely, and when durations did differ, they differed by 100 ms around a mean of 400 ms. Wearden and Ferrara's method produces a complex set of results, but the clearest demonstration of subjective shortening came from what they called LONG trials, those where the second tone really was 100 ms longer than the first (i.e. the response 'long' was in fact correct). On these trials, increasing the gap between the tones led to progressive improvement in performance, i.e. people were much more accurate at making this kind of judgement when the gap between the stimuli was 10 s than when it was 1 s, contrary both to the participants' own intuitions and to apparent commonsense, as performance improved with increasing retention interval. However, the data are consistent with subjective shortening if we assume that people are performing the task by comparing the relatively fresh representation of the second tone with the degraded representation of the first. As the gap between the tones increases, the first tone becomes increasingly subjectively shortened, thus easier to discriminate from the (actually longer) second tone, with consequent improvements in performance as the retention interval lengthens.

The discovery of subjective shortening in humans has led to some presently-unpublished experiments in my laboratory which have addressed a number of questions. One of these is whether this 'directional forgetting' manifested in subjective shortening is unique to duration memories, or occurs for other types of stimuli. More relevant for present purposes is the question of what types of underlying psychological mechanisms might be responsible for duration representation.

Wearden and Culpin (1998) were interested in linking memory for duration with more conventional studies of memory, which in effect means verbal memory, as memory for words or verbalizable material is employed in the overwhelming majority of studies of human memory. They proposed that in a task like Wearden and Ferrara's people actively tried to rehearse a copy of the first tone throughout the inter-stimulus (i.e. retention) interval, using the *phonological loop* of the Baddeley (1986) working memory model. The phonological loop is the mechanism by which phonologically-encoded material, almost always words in other studies, is maintained over the short term, and involves the rehearsal of the retained material in a limited capacity store. Wearden and Culpin (1998) reported the effects of two methods taken from working memory methodology: *articulatory suppression*, where people repeated a single word or some conventional verbal formula during the retention interval, and *preloading*, where the phonological loop is preloaded with some additional material before the item of interest (in our case the first tone to be remembered on the trial). Both articulatory suppression and preloading the phonological loop with a word, so that a person had to remember a word as well as the tone, produced performance deficits compared to conditions with suppression or preloading absent.

Almost all studies of subjective shortening in animals have employed visual stimuli, and studies from our laboratory first demonstrated that visual stimulus durations (the stimuli being squares of colour presented on a computer screen) were also prone to subjective shortening. Given that this occurred, the obvious question arises of how people are representing these visual durations and another part of the working memory model, the relatively-neglected *visuo-spatial sketchpad*, was an obvious candidate. This subsystem of the working memory model parallels the articulatory loop, and involves the short-term retention and rehearsal of material which cannot be encoded phonologically, such as complex visually-presented figures.

A study, conducted in my laboratory in collaboration with Alison Cosgrove, parallels that of Wearden and Culpin (1998) in that it sought to investigate whether visual stimulus durations could be interfered with, much in the same way that articulatory suppression and preloading interfered with the processing of the duration of tones. The experiment was, however, somewhat different from that of Wearden and Ferrara (1993) in that the interval between the stimuli whose durations had to be compared was kept constant at 6 s. So, for example, on a particular trial, a participant would receive a briefly presented white square (mean presentation time 400 ms, although this differed from trial to trial), followed 6 s later by another white square which was either had the same duration as the first, or was 100 ms longer or shorter. The judgement to be made was whether the second duration was longer, shorter, or equal in duration to the first. On a quarter of trials, these were the only events; on other trials, a yellow square, with a duration randomly chosen from 400–600 ms, was presented between the two white squares, and three temporal locations were equally likely, 1, 3, and 5 s after the first white square. Participants had previously been told that the duration of the yellow square was irrelevant and that they could

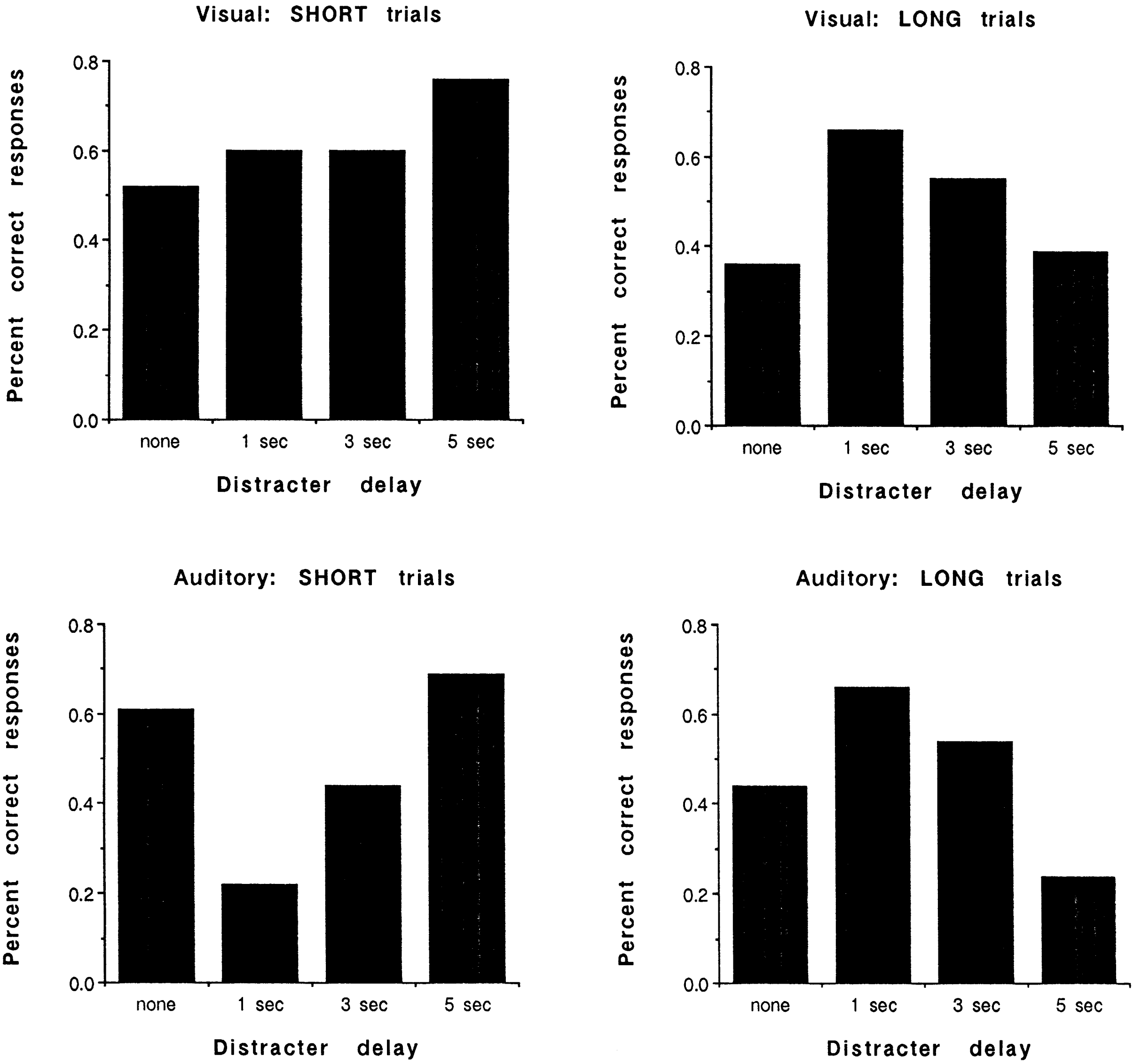

Fig. 3. Percent correct judgements for visual (upper two panels) and auditory (lower two panels) versions of the 'irrelevant distracter' experiment described in the text. Data are presented from SHORT trials (second duration actually shorter than the first), and LONG trials (second duration longer). Correct judgements are given for the control condition without the irrelevant stimulus (none), and for the three different temporal positions of the distracter within the trials (1, 3, and 5 s).

ignore it and, in fact, they were never required to make any judgement of this irrelevant duration. In a variant of this experiment, tones which differed in pitch were used as target and distracter stimuli, to produce an auditory version of the task.

We were interested in two questions. Firstly, would the irrelevant yellow square interfere with duration judgements? Secondly, if there was some interference from this yellow square, what form would in take; in particular, would the interfering stimulus produce shortening, lengthening, or random degradation of the stimulus in interfered with? Fig. 3 shows some results. Data are shown from four trial types: visual SHORT and LONG trials, and their auditory equivalents. Here, a SHORT trial is one in which the second duration really was shorter than the first, a LONG trial one in which the second duration was longer.

As usual in memory for duration experiments, results were at first sight rather complex, and the interfering stimulus had different effects depending on its temporal location in the trial and the trial type. However, the data in Fig. 3 can be simply summarized by saying (i) the putative interfering stimulus *did* perturb time judgements of the stimulus it was temporally closest to and (ii) the form of interference was always shortening. This latter result means that when the interfering stimulus followed the first duration by 1 sec it made it appear shorter, and when it *preceded* the second duration (i.e. it came 5 s after the first one) it shortened the second. So, for example, introducing the interfering stimulus at 1 sec into the trial decreased the correctness of judgements on SHORT trials, but increased it on LONG ones. The reverse effect was found at the 5-s temporal location. Thus the location of the interfering stimulus determined its effects, and interference always produced shortening.

This second result suggests that subjective shortening in memory for duration may be a more pervasive effect than previously supposed. Experiments such as Spetch and Wilkie (1983) and Wearden and Ferrara (1993) demonstrate what might be called *subjective shortening by delay* (i.e. by increasing retention intervals), the results in Fig. 3 show *subjective shortening by interference*. So, when memories for duration are degraded by the need to retain them for a substantial period, or by using an interference manipulation, the result, shortening, is the same.

In the working memory model, both the phonological loop and the sketchpad are slave systems under the control of a *central executive* process, which selects the appropriate subsystem for the tasks, switches between tasks when this is required, and generally supervises performance. Some readers may find the idea of a central executive disturbingly homuncular and mentalistic, and even its inventor expresses occasional qualms about this part of the working memory model (Baddeley, 1996). However, for present purposes the focus of interest is the effect on memory for duration on tasks which are designed to disturb different components of the working memory model, in particular central executive processes. Work carried out in collaboration with Susan O'Rourke explored the effects of a number of potential interference procedures on memory for the duration of visual stimuli, by requiring participants to perform a memory for duration task concurrent with something else.

The procedure used was a simplified version of Wearden and Ferrara (1993) memory for duration task. Each trial consisted of the presentation of two white circles, with the delay of either 0.5 or 6 s between them. The first circle was presented for a random time ranging from 400–700 ms, and the second circle either had the same duration, or was 200 ms longer or shorter. After the second circle, participants had to decide whether it was shorter, longer, or had the same duration as the first. These were the only events in a control condition, but in three other conditions, participants had to perform another task concurrently with the memory for duration procedure. The other task was either articulatory suppression (repeating of the word 'bat', heard on tape every 700 ms), repetitive tapping (pressing a key whenever 'bat' was heard), and random letter generation (avoiding consecutive letters of the alphabet, common letter combinations, or rhymes). Random letter generation is said (Baddeley, 1996) to interfere specifically with central executive processes.

Fig. 4 shows some data on performance accuracy, for the control condition and the three interference procedures. Obviously, all interference procedures decreased performance accuracy, but the random letter generation task produced worse interference than the other two. The interfering tasks in this experiment were present throughout the trial, and other data showed that they interfered about equally with all trial types (SHORT, LONG, and EQUAL), so it is possible that this sort of interference increased timing variance or interfered with decision processes rather than specifically shortening time representations. The experimental design used in this study was, however, inconclusive on this point, as the study

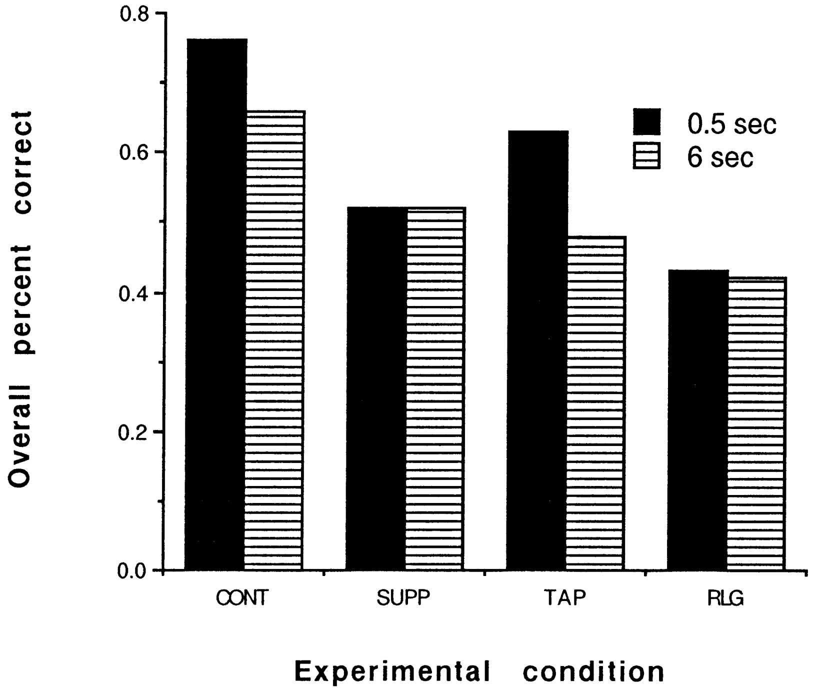

Fig. 4. Percent correct judgements of the relative duration of two visual stimuli separated by a gap which was either 0.5 s (filled columns) or 6 s (striped columns). Data are shown separately for the control (CONT) condition without a concurrent task, and for the three concurrent tasks imposed (articulatory suppression: SUPP; repetitive tapping: TAP; random letter generation: RLG). See text for experimental details.

was not designed to discriminate between shortening and other possible interfering effects, so the interfering manipulations need to be used in other procedures for their effects to be definitively determined. Nevertheless, the data shown in Figs. 3 and 4 show how the content of temporal memory can be manipulated, and these manipulations may allow us new insights into the way that temporal memory operates.

We thus may be able to develop tools which can be used to alter temporal representations at different stages of a more conventional timing task, such as a temporal generalization variant. For example, suppose that X is some interfering manipulation. X can be presented along with a stimulus to be timed in two ways; either when the standard in temporal generalization is presented (the comparisons being given without interference), or at the comparison phase (with the standards being presented without interference). Are the effects of perturbing representations of standards, putatively held in a long-term memory for duration, the same as perturbations of the comparisons, held in a different short-term store, or read directly from the accumulator? At the moment the answers to these questions remain unclear.

Methods derived from conventional memory studies may thus enable the contents of temporal memory (both STM and LTM) to be manipulated, possibly either just shortened (as in Fig. 3) or made more variable. The consequences of such manipulations on subsequent behaviour would thus enable memory processes to be observed more clearly than before. At present, SET often treats memory processes in a manner which to me seems particularly unsatisfactory (although I admit my own guilt, see Lejeune et al., 1998). The idea of a ‘memory storage speed’ constant, or ‘memory constant’ is sometimes used, to account for situations in which, for some reason, the subject does not appear to represent the time of reinforcement veridically (e.g. Meck et al., 1986). The idea is that LTM or STM contents are multiplied by a constant: If this equals 1.0 temporal memory veridically represents duration, if <1.0, time values are stored as shorter than they really were, if >1.0 they are stored as longer. This treatment is nakedly circular, as the only evidence for it is the data that the 'constant' explains. In addition, although values <1.0 might be accounted for by 'loss of pulses' or 'subjective shortening', cases where the constant has to be >1.0 seem particularly hard to understand (and this case is not unknown, see Lejeune et al., 1998). More work needs to be done here, in my view, but perhaps illumination as to how memory processes operate might be sought from experiments like those proposed above, remote though they may initially appear from conventional studies of timing. For example, an interfering manipulation may be able to shorten the durations encoded in LTM. Is the resulting behaviour, observable only when all the result of the SET system has operated, consistent with a 'memory constant' analysis, or not?

5. How do decision processes work?

A central explanatory tool of the scalar timing system is the use of decision processes. One might even say that the introduction of decision processes, the idea that it is not sufficient to know what type of time representations a person or animal has, but also know what they do with these representations, is one of the major advances of SET over previous more poorly specified clock models, like that of Treisman (1963). One general issue that arises in the study of decisions about time is whether consistency or diversity is desirable. On the one hand, consistency requires that temporal judgements which seem in some way similar should require similar decision processes; on the other hand, very different timing tasks might require different types of decisions, and attempting to fit behaviour from too wide a range of timing tasks with the same underlying decision process might be counter-productive and futile. It is difficult, of course, to decide in some particular case whether consistency or diversity is desirable, and too much of the latter can give SET its 'erector kit' image complained of by Staddon and Higa (1999), where explanations of different tasks can be assembled from SET's building blocks, in a way which may appear, or actually be, undesirably inconsistent. However, given that we currently know very little about the way animals and people actually make decisions during timing tasks, such general issues might best be shelved until further progress is made.

It is important to note that as usual in SET, the fact that some mathematical decision formula might fit data well (or excellently by current standards in Psychology), does not mean that the decision process used in the formula is understood. There may be many ways of obtaining the same result, particularly when the data have a simple form, as can be noted in the different explanations of performance on bisection tasks (Allan and Gibbon, 1991; Wearden, 1991b; Wearden and Ferrara, 1995, 1996). Furthermore, even when we might be convinced that the type of decision process we propose is mathematically correct, questions about how a person or animal actually performs the task are far from conclusively answered.

Decision processes, like memories, in SET are usually invented with a view to accounting for the data available; that is, they are theoretical inventions whose justification lies in the fact that the model embodying them fits data well. Little or no attempt to manipulate decision processes, to gain some insight into performance mechanisms, is ever made. However, simple methods might be used to change the way people or animals decide whether to respond in some particular case, with clock and memory operations presumably held constant, and Wearden and Culpin (1998) illustrate one of these. The task employed was, once again, a temporal generalization variant. A 500-ms tone served as a standard and stimulus durations ranging from 200 to 800 ms served as comparisons. After each comparison the participant judged whether or not the stimulus had the standard duration. The method so far described is completely conventional (e.g. see Wearden, 1991a, 1992), Wearden and Culpin's additional manipulation was to associate each decision made with a payoff. Correctly identifying the standard earned 5 points, and correctly rejecting the non-standard stimulus earned two. In a control group, each error resulted in a 1-point deduction. In two other experimental groups, one particular mistake was penalized by the loss of 5 points; in one group it was misidentifying the 400 ms duration as the standard, in the other it was misidentifying the 600 ms duration. Participants were instructed to earn as many points as possible.

In the control group, typically asymmetrical temporal generalization gradients were obtained, with stimuli longer than the standard being more likely to be confused with it than stimuli the shorter by the same amount (i.e. 600 ms was more confused with the standard than 400 ms). When misidentification of the 400 ms stimulus resulted in a 5-point loss, this asymmetry was greatly increased; misidentification of the 400 ms stimulus declined, but stimuli longer than the standard were more likely to be confused with it than in the control group. When misidentification of the 600 ms stimulus was punished, this particular error declined, and gradients became almost perfectly symmetrical.

The results of this simple study may be highly predictable, but illustrate that a very simple payoff manipulation may be used to change timing judgements, when standards and comparison stimuli remain constant. Payoff manipulations have been extensively used in studies of other sorts of perceptual judgements, and are of course commonplace within the framework of signal detection theory, but using them to understand how people (and animals too, although experiments with humans may be procedurally much easier) make decisions about time seems potentially very fruitful.

In general, it seems likely that observed behaviour depends on some kind of unspecified consideration of the payoff consequences of various actions, although how these consequences might influence behaviour has been little considered. Wearden (1990) shows how two different ways of considering performance payoffs on a differential-reinforcement-of-low-rate or other temporal differentiation task can give very different views of the adaptiveness of behaviour. As is well known, when rats or pigeons make lever-presses and key pecks and are required to space their responses by more than t s from the previous response, performance contains many non-reinforced responses that are closely spaced in time regardless of the reinforcement requirement t. Comparing the ratio of reinforced to total responses suggests that animals' adaptation to such schedules is very 'inefficient'. However, if the consequence of behaviour is measured in terms of the overall rate of reinforcement obtained, computer simulations showed that the non-reinforced responses had minimal effects. Indeed, if a certain number of non-reinforced responses are forced by some characteristic of the organism, it is adaptive in terms of overall rate of reinforcement to emit these as rapidly as possible, regardless of the time requirement for reinforcement.

There are many different types of decisions that organisms can make about duration, with human timing exhibiting particular flexibility. For example, events may be judged in terms of their relative gross durations (i.e. is A longer than B?), or may be related in a more complex way (e.g. is A twice or half as long as B?). However, many timing tasks involve one or other of two fundamental types of decisions, the decision of *identity* and that of *similarity*. Temporal generalization most neatly illustrates the former (e.g. people are asked whether stimulus A is the standard), bisection the latter (e.g. people are often instructed to say whether A is more similar to the short or long standard). The relation between identity and similarity decisions remains somewhat ambiguous in SET. For example, an identity judgement may be made if similarity exceeds some threshold, so identity might be supposed to depend on similarity, but the two are clearly not the same. The difference can be illustrated by bisection tasks; if the short and long standards are sufficiently different, people and animals may make the same response (more similar to short or to long) to stimuli that they could otherwise discriminate. For example, if the short standard is 100 ms and the long 900 ms (as in Wearden, 1991b) people will mostly make the same responses to stimuli like 100 and 300 ms, in that both are more similar to the short standard than the long one, even though they could be told apart on most trials if an identity judgement was required. When the standards on bisection are close together (as in Allan and Gibbon, 1991, for example), people may be instructed to make identity decisions even on a bisection task (although Allan and Gibbon themselves admit that the basis of the decision might often be similarity).

Some progress in understanding the way identity and similarity judgements are related might come from requiring these two different judgements of the same stimulus set. There are various ways in which methods for doing this could be developed. For example, if a single stimulus is presented on a trial, there are various things that a human could be asked to do with it. One of these is to estimate its length verbally, another is to classify it as short or long relative to all the stimuli presented (partition bisection, see Wearden and Ferrara, 1995), another is to decide whether it has a standard length (as in various forms of temporal generalization) based on post-response feedback, yet another is to decide whether it was shorter than, longer than, or equal in duration to the stimulus presented on the previous trial (making the task a variant of a Wearden and Ferrara, 1993, memory for duration task). These different methods do not vary decision processes alone: memories of different sorts may be differentially invoked by the different tasks. Nevertheless, they may give some insight into the way in which decisions about temporally-related responses are made when the stimuli to be judged are the same. A particularly interesting experimental method might be one that mixes temporal generalization and bisection decisions for the same set of stimuli: sometimes people are asked whether some time value is more similar to *S* or to *L* (a bisection decision, presumably based on similarity), whereas at other times they are asked whether a time is *S* or *L* (or neither: a temporal generalization decision, presumably based on identity). As mentioned above, intuition suggests that people will probably make identical similarity decisions to stimuli they judge non-identical, but the experiment would confirm or deny this conjecture, and provide other information about relations, if any, between the psychological processes underlying bisection and temporal generalization tasks.

A final question about decision processes concerns their psychological basis. Suppose that a person has to judge whether a just-presented stimulus has the same duration as a previously presented standard. Conventional theory, derived from Church and Gibbon (1982), with a variant introduced by Wearden (1992), proposes that a sample from the long-term memory of the standard is compared with the representation of the just-presented duration in short-term memory. But how is this comparison process done? One idea is that a person or animal uses the sample from the long-term memory as the basis for a ‘copy’ or ‘image’ of the standard, which is then compared with the comparison stimulus active in STM by virtue of its recent presentation. Thus, in fact, both the long-term memory sample and the just-presented duration are present in a short-term store where comparisons are made. Even more specific questions might be asked. For example, if a stimulus duration is represented by a certain number of ‘ticks’ from the internal clock, are the stimuli encoded into ticks, which are then compared, the ticks being a kind of ‘common currency’ of time judgements, or is the internal clock used to create analogue-type images of the stimuli compared?

These questions are probably difficult to answer absolutely conclusively, but there may be ways of launching some kind of preliminary attack on them. Fig. 5 shows a possible experimental starting point. The proposed experiment uses a simple temporal generalization variant. Standard and comparison durations are either auditory (AUD) or visual (VIS), and in Fig. 5 in each of the ten conditions shown the stimulus above the arrow is the standard, that below it is the comparison condition (e.g. top left condition: auditory standards and comparisons). In addition, methods like those used to generate the data shown in Figs. 3 and 4 might be used to develop reliable methods of interfering with the processing of auditory (AUDINT) and visual (VISINT) stimulus durations. These interfering manipulations are used in the comparison phase (e.g. AUD + VISINT in Fig. 5 means auditory comparisons with a visual interference manipulation).

The upper two rows are the most straightforward and investigate whether the putative auditory and visual interfering stimulus have a specific effect on the stimuli whose processing they are designed to perturb. So, for example, if the effect is specific, performance with auditory standards (top row of Fig. 5) should be more disturbed in the AUD + AUDINT condition compared with the AUD + VISINT one, and likewise for the visual equivalent (second row).

The third and fourth rows show the more interesting conditions. Here, previous work on the timing of auditory and visual durations (e.g. Wearden et al., 1998) is used to generate sets of stimuli for cross-modal temporal generalization. Comparison values are selected to have similar subjective durations (so, for example, comparison visual stimuli will be actually longer than auditory standards). The left column of rows three and four show the simple cross-modal comparison. The critical conditions of the experiment are shown boxed. Here, the type of interference present on the comparison trials is different from the stimuli actually presented (e.g. VIS + AUDINT). Suppose, for example, that when performing cross-modal temporal generalization with an auditory standard and visual comparisons, a person actually generates an auditory 'image' during the comparison phase. In this case, auditory interference should specifically interfere with it, even though no auditory stimuli are present in the comparison phase (i.e. VIS + AUDINT). The same is true for the comparable visual condition in row four (AUD + VISINT).

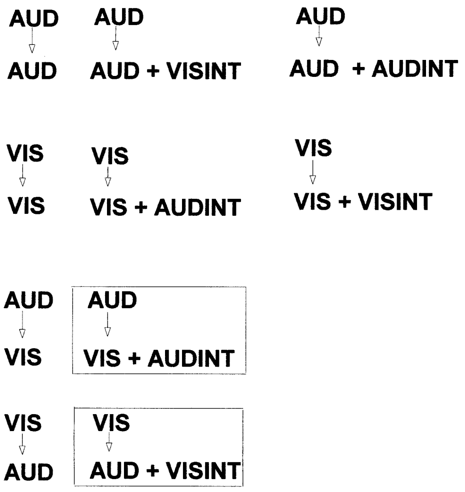

Fig. 5. Schematic outline of the proposed experiment discussed in the text. AUD and VIS stimuli are used, as well as AUDINT and VISINT events. Stimulus above the arrow shows the standard used in the temporal generalization procedure used, events on the comparison trials are shown below the arrow. The conditions of particular interest are shown in boxes on rows three and four, see text for details.

A method like that outlined in Fig. 5 might well enable new insights into just how comparison processes, which form such a central part of SET's account of behaviour, actually work psychologically. I would remind those daunted by the prospect of experiments like that suggested in Fig. 5 that the proposed study is hardly more complex than others used to investigate issues in modern associative learning research (e.g. Rescorla, 1990).

Although decision processes are routinely invoked in the explanation of timed behaviour by SET (indeed, no behaviour can be explained within the SET framework without positing some explicit or implicit decision process) I contend that they are presently very poorly understood, and that fairly simple experimental techniques (albeit techniques different from those currently employed by most studies in the SET framework) might provide considerable insight into their operation.

6. Concluding remarks

The material above is intended to take some aspects of SET, particular as it applies to human timing, 'beyond the fields we know', at least at present, to address some neglected and/or obscure questions, and to indicate some ways in which what seem to be hard problems about how human timing might operate can be attacked, at least in a preliminary way. The approaches suggested may turn out to be less useful than they seem at the time of writing, but the issues addressed, of the generation of variance, and the operation of memory and decision processes, are surely important for the development of SET. They also appear important any currently conceivable approach to timing, whether consistent with SET or not, as it is hard to understand how any future model can account for data from timing experiments without (i) some mechanism for generating time representations (i.e. a clock equivalent, although the mechanism may be structurally different from than proposed by SET), (ii) some store of temporal representations (so that the duration of more than one event can be judged relative to that of another, previous, event), and (iii) some kind of comparison process to generate behaviour (so that different types of decisions about absolute or relative event duration can be made).

A final reason why development of currently obscure parts of SET-like models is so important is that this may be essential for real progress in the search for the brain mechanisms underlying timing. Although at the time of writing the study of neural mechanisms of timing is very much in vogue (see Gibbon et al., 1997, for a review), results from current research are complex and contradictory, often in effect showing little more than the fact that lesion *X* or drug *Y* produced a change in timed behaviour which is, at present, hard to interpret. For example, Artieda et al. (1992) show, in the words of their own title, that 'temporal discrimination is abnormal in Parkinson's disease', but leave the reader none the wiser as to exactly how the timing system in these patients is impaired by their disease. A consideration of the structure of SET-type models (Fig. 1) shows just why interpretation can be so difficult; a lesion or drug may affect pacemaker speed, switch operations, STM/accumulator processes, LTM for duration, decision processes, or even additional factors such as the expression of an internal decision into overt motor action. Many of these effects may have a very similar behavioural consequences, although some pacemaker speed and memory effects can be distinguished with certain existing experimental designs (Meck, 1996). By developing more precise characterisation of how the different parts of the SET system operate some of the ambiguity of effects of lesions and drugs may be resolved, thus opening the way for significant progress in understanding the brain mechanisms of timing.

Acknowledgements

The ideas developed in this article have benefitted from discussions over several years with colleagues and students, and from attempts to teach SET to the latter. In particular, Helga Lejeune has often provided a patient audience for my attempts to wrestle with some of the problems discussed here, and in French too, although she bears no responsibility for the present contents. Talks with Sue O'Rourke and Lorraine Allan also improved the clarity of my ideas.

References

Allan, L.G., 1979. The perception of time. Percept. Psychophys. 26, 340–354.

Allan, L.G., Gibbon, J., 1991. Human bisection at the geometric mean. Learn. Motiv. 22, 39–58.

Artieda, J., Pastor, M.A., Lacruz, F., Obeso, J.A., 1992. Temporal discrimination is abnormal in Parkinson's disease. Brain 115, 199–210.

Baddeley, A.D., 1986. Working Memory. Oxford University Press, Oxford.

Baddeley, A.D., 1996. Exploring the central executive. Quart. J. Exp. Psychol. 49, 5–28.

Church, R.M., 1980. Short-term memory for time intervals. Learn. Motiv. 11, 208–219.

Church, R.M., DeLuty, M.Z., 1977. Bisection of temporal intervals. J. Exp. Psychol: Anim. Behav. Proc. 3, 216–228.

Church, R.M., Gibbon, J., 1982. Temporal generalization. J. Exp. Psychol.: Anim. Behav. Proc. 8, 165–186.

Church, R.M., Meck, W.H., Gibbon, J., 1994. Application of scalar timing theory to individual trials. J. Exp. Psychol.: Anim. Behav. Proc. 20, 135–155.

Gibbon, J., 1977. Scalar expectancy theory and Weber's law in animal timing. Psych. Rev. 84, 279–325.

Gibbon, J., 1992. Ubiquity of scalar timing with a Poisson clock. J. Math. Psychol. 36, 283–293.

Gibbon, J., Church, R.M., 1981. Time left: linear versus logarithmic subjective time. J. Exp. Psychol.: Anim. Behav. Proc. 7, 87–108.

Gibbon, J., Church, R.M., Meck, W., 1984. Scalar timing in memory. In: Gibbon, J., Allan, L. (Eds.), Ann. N.Y. Acad. Sci. 423: Timing and Time Perception. New York Academy of Sciences, New York, pp. 52–77.

Gibbon, J., Malapani, C., Dale, C.L., Gallistel, C.R., 1997. Toward a neurobiology of temporal cognition: advances and challenges. Curr. Opin. Neurobiol. 7, 170–184.

Grant, D.S., Spetch, M.L., Kelly, R., 1997. Pigeons' coding of event duration in delayed-matching-to-sample. In: Bradshaw, C.M., Szabadi, E. (Eds.), Time and Behaviour: Psychological and Neurobiological Analyses. Elsevier, Amsterdam, pp. 217–264.

Killeen, P.R., Fetterman, J.G., 1988. A behavioral theory of timing. Psych. Rev. 95, 274–295.

Lejeune, H., Wearden, J.H., 1991. The comparative psychology of fixed-interval responding: some quantitative analyses. Learn. Motiv. 22, 84–111.

Lejeune, H., Ferrara, A., Soffié, M., Bronchart, M., Wearden, J.H., 1998. Peak procedure performance in young adult and aged rats: acquisition and adaptation to a changing temporal criterion. Quart. J. Exp. Psychol. 51B, 193–217.

Nichelli, P., Alway, D., Grafman, J., 1996. Perceptual timing in cerebellar degeneration. Neuropsych. 34, 863–871.

Malapani, C., Rakitin, B., Meck, W.H., Deweer, B., Dubois, B., Gibbon, J., 1998. Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction. J. Cogn. Neurosci. 10, 316–333.

Maricq, A.V., Roberts, S., Church, R.M., 1981. Metamphetamine and time estimation. J. Exp. Psychol.: Anim. Behav. Proc. 7, 18–30.

McCormack, T., Brown, G.D.A., Maylor, E.A., Darby, A., Green, D., 1999. Developmental changes in time estimation: comparing childhood and old age. Dev. Psychol. (in press).

Meck, W.H., 1983. Selective adjustment of the speed of internal clock and memory processes. J. Exp. Psychol.: Anim. Behav. Proc. 9, 171–201.

Meck, W.H., 1996. Neuropharmacology of timing and time perception. Cogn. Br. Res. 3, 227–242.

Meck, W.H., Church, R.M., Wenk, G.L., 1986. Arginine vasopressin inoculates against age-related increases in sodium-dependent high affinity choline uptake and discrepancies in the content of temporal memory. Eur. J. Pharmacol. 130, 327–331.

Meck, W.H., Church, R.M., Wenk, G.L., Olton, D.S., 1987. Nucleus basalis magnocellularis and medial septal lesions differentially impair temporal memory. J. Neurosci. 7, 3505–3511.

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. J. Exp. Psychol.: Anim. Behav. Proc. 22, 307–320.

Rakitin, B.C., Gibbon, J., Penney, T.B., Malapani, C., Hinton, S.C., Meck, W.H., 1998. Scalar expectancy theory and peak-interval timing in humans. J. Exp. Psychol.: Anim. Behav. Proc. 24, 15–33.

Rescorla, R.A., 1990. Evidence for an association between the discriminative stimulus and the response-outcome association in instrumental learning. J. Exp. Psychol.: Anim. Behav. Proc. 16, 326–334.

Roberts, S., 1981. Isolation of an internal clock. J. Exp. Psychol.: Anim. Behav. Proc. 7, 242–268.

Rodriguez-Girones, M.A., Kacelnik, A., 1998. Response latencies in temporal bisection: implications for timing models. In: DeKeyser, V., d'Ydewalle, G., Vandierendonck, A. (Eds.), Time and the Dynamic Control of Behavior. Hogrefe and Huber, Gottingen, pp. 51–70.

Spetch, M.L., Wilkie, D.M., 1983. Subjective shortening: a model of pigeons' memory for event duration. J. Exp. Psychol.: Anim. Behav. Proc. 9, 14–30.

Staddon, J.E.R., Higa, J.J., 1999. Time and memory: towards a pacemaker-free theory of interval timing. J. Exp. Anal. Behav. 71.

Treisman, M., 1963. Temporal discrimination and the indifference interval: implications for a model of the 'internal clock'. Psych. Monogr., 77 (whole number 576).

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., 1990. Maximizing reinforcement rate under spaced responding schedules under conditions of temporal uncertainty. Behav. Proc. 22, 47–59.

Wearden, J.H., 1991a. Do humans possess an internal clock with scalar timing properties? Learn. Motiv. 22, 59–83.

Wearden, J.H., 1991b. Human performance on an analogue of an interval bisection task. Quart. J. Exp. Psychol. 43B, 59–81.

Wearden, J.H., 1992. Temporal generalization in humans. J. Exp. Psychol.: Anim. Behav. Proc. 18, 134–144.

Wearden, J.H., Culpin, V., 1998. Exploring scalar timing theory with human subjects. In: DeKeyser, V., d'Ydewalle, G., Vandierendonck, A. (Eds.), Time and the Dynamic Control of Behavior. Hogrefe and Huber, Gottingen, pp. 33–49.

Wearden, J.H., Ferrara, A., 1993. Subjective shortening in humans' memory for stimulus duration. Quart. J. Exp. Psychol. 46B, 163–186.

Wearden, J.H., Ferrara, A., 1995. Stimulus spacing effects in temporal bisection by humans. Quart. J. Exp. Psychol. 48B, 289–310.

Wearden, J.H., Ferrara, A., 1996. Stimulus range effects in temporal bisection by humans. Quart. J. Exp. Psychol. 49B, 24–44.

Wearden, J.H., Towse, J., 1994. Temporal generalization in humans: three further studies. Behav. Proc. 32, 247–264.

Wearden, J.H, Wearden, A.J., Rabbitt, P., 1997a. Age and IQ effects on stimulus and response timing. J. Exp. Psychol.: Hum. Percep. Perf. 23, 962–979.

Wearden, J.H., Denovan, L., Fakhri, M., Haworth, R., 1997b. Scalar timing in temporal generalization in humans with longer stimulus durations. J. Exp. Psychol.: Anim. Behav. Proc. 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. Quart. J. Exp. Psychol. 51B, 97–120.

Zakay, D., Block, R.A., 1995. An attentional-gate model of prospective time estimation. In: Richelle, M., DeKeyser, V., d'Ydewalle, G., Vandierendonck, A. (Eds.), Time and the Dynamic Control of Behavior. Psychology Faculty PAI, Liege, Belgium, pp. 167–178.