ACTA NEUROBIOLOGIAE EXPERIMENTALIS

Acta Neurobiol Exp 2004, 64: xxx-xxx

Decision processes in models of timing

John H. Wearden

Department of Psychology, Manchester University, Manchester, M13 9PL, UK, Email: wearden@psy.man.ac.uk

Review

Abstract. The article discusses the role played by decision mechanisms in the leading model of timing, scalar expectancy theory. Examples of the roles played by decision mechanisms in explanations of behaviour on temporal generalization and bisection are presented. Decision mechanisms for different timing tasks often have a common form (thresholded normalized difference, TND), where differences between durations are "normalized" (i.e., divided) by another duration value, then compared with a threshold. The TND principle provides a rule for both similarity and identity judgements of duration. The role of threshold mechanisms in timing is discussed, and it is shown that some procedural manipulations appear to specifically alter threshold values. Finally, problems in modelling the decision processes involved in verbal estimation are discussed.

Key words: time perception, internal clocks, decision processes, temporal generalization, bisection, verbal estimation

INTRODUCTION

A reader unfamiliar with the details of recent developments in the study of the perception of time by animals and humans might think that the principal problem of time psychology was the identification of some basic internal timing process, for example a type of internal clock or oscillator-based mechanism, which translates physical time into a code that a behaving organism can use. This view is at least partly correct, in that a search for an "internal clock", or some other equivalent mechanism, occupies an important place in contemporary research on time (for just a few recent studies, see Droit-Volet and Wearden 2002,Gibbon et al. 1984, Penton-Voak et al. 1996, Treisman et al. 1990, Wearden and Penton-Voak 1995).

The present article will briefly discuss some sorts of clocks that have been proposed by recent work, as well as the involvement of memory mechanisms in timing, but its main focus is not on clock processes necessary for timing, but decision processes. Why are decision processes necessary at all? Part of the answer is general: we can only observe the behaviour emitted by our experimental subjects (be they rats pressing levers, or people making verbal responses), and it thus seems likely that the operation of the basic time-keeping mechanism (the putative clock) will only be indirectly revealed in behaviour, with other sorts of processes intervening between the appreciation of subjective time inside the organism, and the response that emerges. Another way of answering this question is to consider how people might be performing on some simple timing tasks, of which temporal generalization (Wearden 1992) will serve as an example.

In the simplest form of the temporal generalization procedure, a person is initially exposed to a standard duration (e.g., a tone 400 ms long), identified as such. After a few exposures to the standard, comparison durations are presented (e.g., tones ranging from 100 to 700 ms in duration) and the participant must simply judge whether each of them is (a "yes" response) or is not (a "no"response) the standard, with feedback as to performance accuracy usually being given.

Temporal generalization rapidly produces orderly data from humans ranging from 3-year-old children (Droit-Volet et al. 2001), through student age participants (Wearden 1992), to the elderly (McCormack et al. 1999, Wearden et al. 1997a). The details of how people might perform on this task will occupy us later, but for the moment just consider the psychological mechanisms that might be needed. Firstly, the standard is presented alone, and needs to be timed (by the putative clock), then stored in some sort of "reference" memory, at least for the duration of the experiment. The comparison durations presented as test stimuli likewise need to be timed, and retained, but this time only for the duration of the trial. A person thus has a "standard" and a "comparison" (although only the latter has been recently presented), and has to judge their equality. Even if the comparison and standard durations were physically the same (e.g., both 400 ms in our example) it seems unlikely that they will be exactly subjectively equal (as clock and memory processes may add trial-by-trial variance), so the participant must have some decision mechanism to decide whether the two durations are "close enough" to warrant a "yes" response (i.e., an identification of the comparison as being the standard).

So, even this very simple task seems to involve memory (possibly different types of memory, one type for the standard and the other for the comparison) and decision processes, and other commonly-used timing tasks likewise appear to involve much more than just the clock process itself, although the clock is necessary to produce the "raw material" that enters into the subsequent decision process. For example, in one form of a bisection task (Wearden 1991b) people initially receive "short" and "long" standards (e.g., tones 200 and 800 ms long), then have to classify other tones (comparisons 200-800 ms long) in terms of their similarity to the "short" and "long" standards. Here, once again, the standards need to be stored for the duration of the experiment, the comparisons for the duration of the trial, and a decision process is needed to generate the similarity decision. In verbal estimation, participants assign verbal labels in conventional time units (seconds or milliseconds) to presented stimuli, and once again the choice of an appropriate verbal label seems to require a complex decision process, over an above any mechanism needed to produce the "raw" time representation.

CLOCKS, MEMORIES, AND DECISIONS: SCALAR TIMING THEORY (SET)

Although the principal focus of this article will be on decision processes in timing, it is important to provide a psychological context in which these decision processes take place. The simplest way of doing this is to give a brief introduction to scalar timing theory (SET), a multi-process theory of animal and human timing, developed in its "classical" form by Gibbon et al. (1984) on the basis of earlier work by Church and Gibbon (1982), and Meck (1983). Figure 1 shows the tripartite structure of the SET system, which consists of operations at 3 levels, usually considered to operate sequentially.

The highest level of the SET system is the clock level consisting of a pacemaker-accumulator clock: that is, a pacemaker generates "ticks" or "pulses" which are gated *via* a switch to an accumulator which accumulates them. For example, to time a stimulus of some duration, the switch is closed by stimulus onset, allowing pulses to flow to the accumulator, and the switch opens again, cutting the pacemaker/accumulator connection. Thus, at the end of the stimulus period, the accumulator contains the number of pulses corresponding to the duration of the stimulus. Many aspects of the operation of the pacemaker-accumulator clock have been investigated in the literature, including attempts to change pacemaker speed (Burle and Bonnet 1999, Droit-Volet and Wearden 2002, Meck 1983, Penton-Voak et al. 1996), and alter switch processes (Droit-Volet 2003, Wearden et al. 1998).

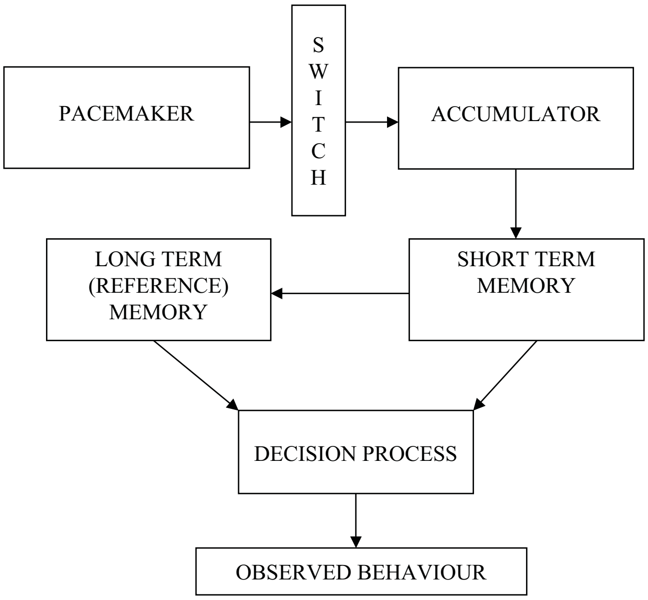

Fig. 1. Outline of the scalar expectancy theory (SET) system. The upper level shows the pacemaker-accumulator clock, the middle level the long-term reference memory and the short-term working memory, and the lowest level the decision mechanism.

The "raw" representation of stimulus duration represented by accumulator contents is transferred to a "short-term or working memory" for time (in some recent versions of SET, the accumulator and working memory are combined), and this memory is one part of the next level of the system, the memory level. After a stimulus has been presented, its duration representation is considered to reside in working memory, and it may undergo changes there (a full discussion is beyond the scope of this article, but see Wearden et al. 2002). The other memory store (reference memory) is considered to contain "important times" such as standards used in temporal generalization and bisection. Once again, a full discussion of the operation of this memory system requires too much space to be done here, but Jones and Wearden (2003, 2004) provide one.

Finally, there is a decision level, where the usual comparison made (as will be discussed in detail later) is between a sample drawn from the reference memory of the standard or standards appropriate for the task, and the contents of working memory. What the comparison process is depends on the task, as will be seen later, but only after the comparison process is complete does a behavioural response occur.

It is clear from this brief introduction that SET is a rather complex multi-process theory, and it is further clear that according to SET we must regard observed behaviour as reflecting underlying clock processes only indirectly. Different conditions and different participant groups may produce different behaviours not because their "raw" timing (i.e., clock) processes differ, but because of differences in memory and decision mechanisms. Likewise, attempts to unravel the underlying physiology of timing must not only consider how internal clocks might be instantiated in the brain, but also how memory and decision processes might be mediated by neural systems, a problem possibly more difficult than imagining how "neural pacemakers" might work. A corollary of this approach is that measurements of brain electrical activity during timing tasks are likely to pick up activation not only resulting from clock processes, but also from memory and decision processes too.

DECISION PROCESSES IN ACTION: TEMPORAL GENERALIZATION AND THE MCG MODEL

The temporal generalization task (developed for humans by Wearden (1991a, 1992) from a procedure employed with animals by Church and Gibbon (1982)) serves to illustrate how decision processes work in the SET system. The basic procedure was described earlier, but how does SET explain the behaviour obtained? Wearden (1992) developed a "modified Church and Gibbon" (MCG) model which, as its name suggests, is a variant of a model used to model temporal generalization in rats in Church and Gibbon's original article.

The task begins with a number of presentations of the standard, *s*, which is assumed to reside in reference memory. Each comparison stimulus, *t*, is assumed to be timed without error, and each is successively present in working memory, which is overwritten with each comparison presentation. The reference memory is assumed to be represented as a Gaussian distribution with accurate mean (i.e., *s*), and some coefficient of variation (standard deviation/mean), *c*. On each trial, a random sample, *s**, is drawn from the reference memory of *s*, and the participant identifies a comparison duration, *t*, as the standard when

$$\text{abs } (s^* - t)/t < b^* \qquad (1)$$

where *s** and *t* are as defined above, *b** is a threshold variable from trial to trial, and abs indicates absolute value.

Temporal generalization gradients from adults usually: (i) peak at the standard, *s*; and (ii) are asymmetrical in that durations longer than the standard are more confused with it than durations shorter by the same amount (e.g., if *s* is 400 ms, 500 ms is more confused with it than 300 ms is). Inspection of equation 1 shows that it reflects both of these properties. The left-hand side of equation 1 reflects the difference between *s** and *t*, with smaller values indicating greater similarity, and it is obvious that the difference will be smallest on average when $t = s$, so gradients should peak at the standard.

The asymmetry is easily intuitively explained if we ignore the * indications and treat *s* and *b* as constant (which makes no difference to the average prediction). For example, suppose $s = 400$ ms, and we present $t_1 = 300$ ms and $t_2 = 500$ ms as comparisons. In the former case, the left-hand side of equation 1 is abs (400 – 300)/300 or 1/3, and in the latter (400 – 500)/500 = 1/5, thus 500 ms is more "similar" to 400 ms according to this calculation than 300 ms is, i.e., it generates a smaller value and thus one more likely to be below *b*, and thus generate a "yes" response.

Figure 2 shows temporal generalization data averaged over 83 female undergraduates. The standard duration was 400 ms, and comparisons ranged from 100 to 700 ms. The empirical temporal generalization gradients peaked at the standard, and showed the aforementioned asymmetry, with comparison durations longer than the standard being more likely to be confused with it than comparisons shorter by the same amount. The line in Fig. 2 shows the best-fitting MCG model, which obviously fits the data with a high degree of precision. To fit the model, the MCG equation was embodied in a computer program, and the threshold, *b*, and *c*, the coefficient of variation of the memory of the standard were varied to find the best fit in terms of absolute deviation between the predictions of the model and the data.

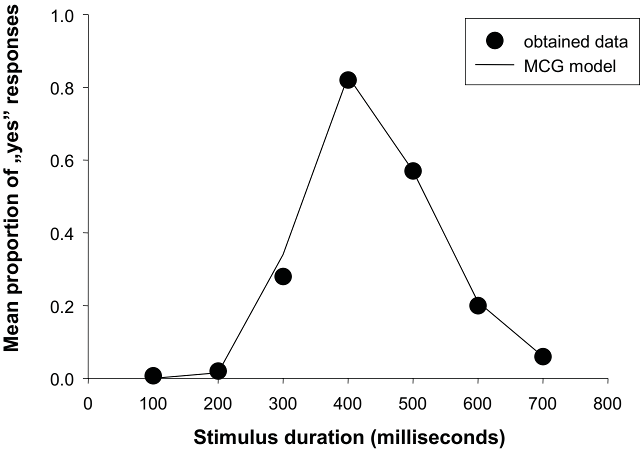

Fig. 2. Data from temporal generalization. The mean proportion of "yes" responses (identifications of a comparison stimulus as the standard) is plotted against comparison stimulus duration. Data points are shown by filled circles and the best-fitting MCG model by the line. Data are averaged over 83 female undergraduates. The standard was 400 ms and comparisons ranged from 100 to 700 ms.

A number of comments are in order. One is that the asymmetry in the temporal generalization gradient is "predicted" by the MCG model because of its decision rule. The asymmetry is not a product of internal clock processes (i.e., it does not result from the way the internal clock "scales" real time), nor of the way times are remembered. If the decision process had been different, differently-shaped gradients would be predicted. For example, Church and Gibbon's (1982) original model was to respond when

$$\text{abs } (s^* - t)/s^* < b^* \qquad (2)$$

where all terms are as Equation 1. A moment's consideration shows that Equation 2 predicts symmetrical generalization gradients, which were indeed found in rats. To fit the asymmetrical data found in humans, I modified the model merely by "normalizing" the raw difference (s^* - t) by t rather than s^*: in other words, the difference is compared with the most recently presented comparison rather than the representation of the sample, as it was in Church and Gibbon's original model.

A reader might point out, correctly, that I have chosen a model which fits the data, and "arbitrarily" rejected another one. To this I can only reply that: (i) only a fool would choose a model which obviously didn't fit the data; and (ii) the charge of choosing equations to fit data sets could be levelled equally against Newton or Einstein, but usually isn't. It is true that there is no clear justification for the normalization used in Eq. 1, although the question of how this arbitrariness in modelling can be addressed is discussed later.

Although the "normalization" factor needs to be t rather than s^* to fit data from humans, other aspects of the model are more fluid. The original MCG model assumed two sources of trial by trial variance: one in s^* and the other in the threshold b, but exactly where the sources of variance are located is somewhat arbitrary. In my 1992 paper I showed that variants with a fixed threshold (Equation 3)

$$\text{abs}\,(s^* - t)/t < b \tag{3}$$

and one with variance also in the comparison t (which becomes t^* on each trial), Equation 4,

$$\text{abs}\,(s^* - t^*)/t^* < b^* \tag{4}$$

fitted data nearly as well as the MCG model. So, the modelling has some freedom as to the placement of sources of variance, but is constrained to fit asymmetrical gradients.

The MCG model can be developed in various ways. Droit-Volet and colleagues (2001) used a variant to model temporal generalization data from 3, 5, and 8 year-old children. An added factor in their model was "random responding", a tendency to respond "yes" or "no" at random without regard to the comparison duration. Such putative random responding was needed to model data from the two youngest groups which were both sensitive to comparison duration but involved high levels of "yes" responses at the longest and shortest comparisons (which adults hardly ever confuse with the standard). The level of random responding that needed to be added declined systematically with increasing age, reaching near-zero levels at 8 years, thus suggesting that this factor played no role in responding in adults. A further factor needed in modelling data from the youngest children was a shift in the mean of reference memory. Data were fitted best if it was assumed that the youngest children remembered s as being on average slightly shorter than it really was, but this tendency, like random responding, declined with increasing age.

JUSTIFYING AND TESTING PARAMETERS: FROM MODELS OF BEHAVIOUR TO THE BEHAVIOUR OF MODELS

Droit-Volet et al.'s work illustrates a persistent difficulty in any sort of modelling, namely that factors additional to the basic model can always be added (or things subtracted from the basic model) in order to fit data. This makes a model almost impossible to disprove as additional factors can be added without limit (see further discussion in Wearden 1999). The main defence against this proliferation of parameters is commonsense plausibility, linked to previous psychological knowledge, and the "behaviour" of the parameters used. For example, the random responding needed to model generalization data from 3 and 5 year-old children could be justified in terms of the well-known difficulties that such young children have in maintaining attention to a rather tedious laboratory task. This "psychological" justification for random responding implies that attention improves with increasing age, so predicts orderly behaviour in the random responding parameter: the amount of random responding needed should decrease with increasing age, and it does (Droit-Volet et al. 2001) If, for example, more random responding had been needed at 5 years than 3, and less at 8 than 5, then the parameter would not behave in a plausible way and its use would seem arbitrary, and questionable.

Another method of addressing the problem of arbitrariness of models is to examine whether the parameters have testable psychological meaning. For example, is b in our Eqs. 1 to 4 really a threshold independent of other factors in the equation? Wearden and Grindrod (2003) tested this by using a variant of temporal generalization in which participants were given points for correct responses (hits and correct rejections) with points deducted for errors (misses and false alarms). In their study, different groups had different numbers of points added for the different correct responses: in one group hits obtained more points than correct rejections and in another group the opposite was true. The logic here was than differentially rewarding hits would tend to make participants say "yes" more often, whereas differentially rewarding correct rejections would make them say "yes" less, even when the durations presented in the different conditions were physically the same.

Wearden and Grindrod's manipulation had the desired behavioural effect but, more importantly, modelling suggested that the behavioural effect was mediated by changes in *b*. Differentially rewarding hits increased *b*, whereas differentially rewarding correct rejections decreased it: exactly the result predicted if *b* is a response threshold.

The *c* parameter of the MCG model, essentially the variability of memory of the standard, likewise might be expected to show orderly changes across certain comparisons. For example, Droit-Volet et al. (2001) found that *c* decreased (i.e., reference memory became less variable) with increasing age (for a similar result see McCormack et al. 1999), and further evidence for orderliness in *c* comes from Wearden et al.'s (1997a) study of timing in the elderly. Here, normal, community-resident elderly were tested on a range of timing tasks one of which was temporal generalization with a 400-ms standard and 100-700 ms comparisons.

Wearden et al.'s subject population was chosen from a larger sample so that the normal confound between age and IQ (i.e., IQ decreasing on average with increasing age) was avoided, so data could be analyzed separately in terms of age (with IQ constant between groups) or IQ (with age constant). Temporal generalization gradients had the normal form (i.e., they peaked at the standard with a rightward skew) in all groups, but fitting the model revealed that *c* increased with both increasing age, and decreasing IQ, so older participants and those with lower IQs had more variable temporal memories according to the model than those who were younger and with higher IQ. Thus, once again, the *c* parameter behaved in an orderly way across comparisons between subject groups.

SOURCES OF VARIANCE: EPISODIC TEMPORAL GENERALIZATION

SET has its origins in animal psychology and was developed initially as an explanation of animal performance on reinforcement schedules involving temporal requirements or temporal periodicities, such as the fixed-interval (FI) schedule (for an application of ideas related to SET to FI see Lejeune and Wearden 1991). As discussed extensively in Jones and Wearden (2003, 2004), the legacy of animal psychology has been a conception of reference memory which is problematic when applied to humans, and perhaps also even when applied to animals.

Suppose that an animal performed under FI *t* s (e.g., FI 30 s). In this condition, the animal would be reinforced for the first response occurring more than 30 s from the previous reinforcer, and the temporal constraint of the schedule can be learned by a range of animal species (Lejeune and Wearden 1991) all of whom eventually produce a pattern of behaviour involving pausing after reinforcer delivery with an increasing tendency to respond as the time of reinforcement approaches.

Classical SET supposes that the reference memory is built up of instances of slight "misrememberings" of the time of reinforcement (30 s in our case). For example, on one trial 30 s is stored as 28 s in reference memory, on another trial 34 s, and so on. Animals may receive hundreds of instances of a certain time of reinforcement when learning a schedule like FI, so it seemed reasonable to suppose that gradually a distribution would build up in reference memory so that the distribution of reinforcement times in it would be have a Gaussian form with a mean of 30 s, and some variability. This sort of "variance by accumulation" is the source of classical SET's argument that an important source of variability in timing resides in the memory of the standard. In fact, classical SET regards the internal clock as timing almost perfectly on every trial, so 30 s is always timed as exactly 30 s, but is misremembered (i.e., in being transferred from the accumulator to the working memory it is transformed by a Gaussian function with a mean of 1.0 and some coefficient of variation, *c*). Put more casually, performance is variable not because the subjects cannot accurately time the interval, but because they cannot remember the standard duration for the task with complete accuracy.

This conception of reference memory needs some comment. One thing that it does is generate the well-known scalar property of time. This is a form of conformity to Weber's law, and can be tested in various ways. One test is to plot the standard deviation of time judgements against the mean, with a result that the standard deviation is a linear function of the mean.

Alternatively, a coefficient of variation (standard deviation/mean) statistic can be calculated from behaviour at different absolute times, and this quantity, a kind of Weber fraction, should remain constant as the absolute timed judged varies. Finally, data can be tested for the property of superimposition, the requirement that timing measures should superimpose when plotted on the same relative scale. Wearden et al. (1997b) provided a good example of superimposition in their temporal generalization experiment where humans performed with standards of 2, 4, 6, and 8 s (counting was prevented by a concurrent task). When the resulting temporal generalization gradients (which were all peaked at the standard, rightward skewed, and well-fitted by the MCG model) were plotted on the same relative scale (with comparison durations plotted as a fraction of the standard in force for the condition, 2 s, or whatever it was) the gradients superimposed almost perfectly.

The existence of the scalar property is attributed by classical SET to the properties of reference memory. If each instance of a standard in, say, temporal generalization is misremembered by being multiplied by a Gaussian function with mean of 1 and some coefficient of variation, c, then the representation in reference memory will have the scalar property, as a result of c being a relative measure.

There is some doubt, however, about whether this conception of reference memory is correct. For one thing, the idea of reference memory as an accumulation of a very large number of individual instances of an "important" time, while plausible in (some) animal studies seems much less applicable to timing in humans, where learning of standards in temporal generalization, to give just one example, may be very rapid (taking place in a single trial, according to Jones and Wearden (2003)). Jones and Wearden (2003) present an alternative to the classical conception of reference memory, but the details are too complicated for discussion here. There are even recent data from animal studies that suggest that once a particular time of reinforcement is learned, an animal may change its behaviour very rapidly when the reinforcement time is changed, even after a few experiences with the new time of reinforcement (e.g., Lejeune et al. 1997). Such rapidity of behavioural adjustment obviously suggests a reference memory that can be rapidly over-written rather than one which changes gradually as a result of accumulation of instances.

Another problem with classical SET may reside in its assertion that the source of scalar variability resides in reference memory. A more commonsense view might be that the "raw" representations of durations produced by the internal clock itself already contain the necessary scalar variance, because the clock runs at slightly different speeds on each trial, so the number of pulses accumulated in some constant time, t, varies from one trial to another. Gibbon et al. (1984) even presented a potential clock mechanism which would generate scalar properties directly.

If the reference memory is the source of scalar variance, then what happens when tasks are devised where reference memory is not used? The last few years have seen the development of a number of timing tasks which appear to make the use of reference memory difficult or impossible ("roving standard" or "episodic" tasks, see Allan and Gerhardt 2001, Rodriguez-Girones and Kacelnik 2001), and here I will concentrate on only one of these, the "episodic temporal generalization" method of Wearden and Bray (2001).

A condition from their Experiment 3 illustrates the basic method. Each trial consisted of the presentation of 2 tones, separated by a brief random-length gap, and the participant's task was to judge whether the two tones had the same duration, although no feedback was given. One of the tones (the sample) was drawn from one of three distributions, 350-500 ms, 450-750 ms, or 600-1000 ms, and the other (the comparison) was one of the following multiples of the sample, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75. The order of the sample and the comparison was randomized from trial to trial. In this task, duration pairs were never repeated except by chance, and durations presented could be very variable from one trial to another. There was, obviously, no standard duration presented and even constructing a standard (e.g., by averaging all the durations presented together, resulting in 600 ms) seems useless to the participant. Superimposition could be tested by plotting data obtained from the different duration ranges, and was found to hold (see Wearden and Bray 2001 (Fig.4, p. 305)).

This experiment, like others, showed the scalar property of time even in the likely absence of reference memory, so questioned whether reference memory could be the sole, or principal, source of variance in timing models. How might the episodic temporal generalization task of Wearden and Bray be modelled?

The MCG model (Eq. 1) makes two assumptions which are inappropriate for modelling episodic data: first, it assumes that reference memory of s is used, and second, by implication it assumes that s and t are treated differently (i.e., one has variance and the other does not, and one serves as the denominator – the "normalizing factor" for the absolute difference between s^* and t).

Neither assumption can be used when modelling episodic data. If the duration pairs are never repeated, then it seems impossible for the participant to develop any sort of useful reference memory, so no s can be formed. Furthermore, the absence of any distinction between s and t means that neither of the stimuli on the trial can be a privileged normalizing factor for the judgement. I therefore sought a model in which the two durations presented on the trial (t_1 and t_2) were treated identically.

The starting point, like that of the MCG model, was to assume that the basis of the decision to classify a presented duration as the standard was a difference, in the present case the absolute difference between the two durations presented on the trial. Previous data from episodic generalization (Wearden and Bray 2001) showed that response functions from different duration ranges showed approximate superimposition, suggesting that absolute differences, by themselves, could not be the basis of the decision to classify a presented duration as the standard, and that some normalizing factor was needed. Given that neither of the stimuli on the trial had any privileged status I decided to use the mean of the two as the normalizing factor, and thus derived a quantitative model of the episodic task.

In the episodic model, a "yes" response was produced when

$$\text{abs}\ (t_1^* - t_2^*)/m < b^* \qquad (5)$$

where t_1^* and t_2^* are transformations of t_1 and t_2 (the to-be-compared pair of stimuli), m is the mean of the pair of transformed stimuli, and b^* is a variable threshold. t_1 and t_2 are transformed by converting each time value into a Gaussian distribution with mean accuracy and a coefficient of variation c, and drawing a random sample from this distribution. b^* is assumed to be drawn from a Gaussian distribution with mean b and a standard deviation of $0.5b$. In this model, c is assumed to reflect perceptual (rather than mnemonic) variance, since there are no repeated stimuli in the episodic task that could be stored in reference memory.

Figure 3 shows data from an unpublished experiment like that of Wearden and Bray (2001), conducted in collaboration with Steven Grice. One point of interest is that the empirical generalization gradients (upper panel) show superimposition. For the "short" range condition the sample was drawn from a uniform distribution running from 300-500 ms, and for the "long" range condition the sample was drawn from a distribution running from 600-1000 ms. Data from both ranges superimposed, confirming the contention of Wearden and Bray (2001) that scalar timing could be found in the absence of reference memory.

However, in addition, inspection of the empirical generalization gradients shows that they are asymmetrical (with more "yes" responses occurring when the comparison/standard multiples are above 1.0). At first sight this seems inexplicable, as according to the arguments made here there is no psychological distinction between

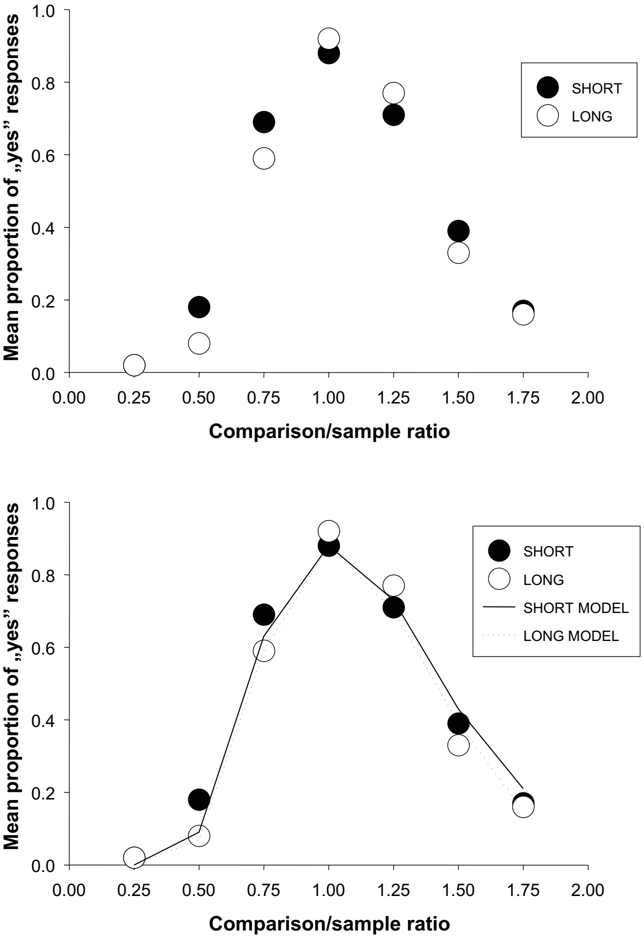

Fig. 3. Data and modelling of episodic temporal generalization. Upper panel: proportion of "yes" responses (judgements that the two presented durations are equal) plotted against comparison/sample ratio (see text for details). Data are shown separately for the "short" range and "long" range conditions. Lower panel: data points as upper panel (unconnected circles), and best-fitting episodic model (solid and dotted lines). See text for other details.

the two stimuli presented, but the fit of equation 5 is shown in the lower panel, and this also predicts asymmetrical gradients. The reason becomes clear on inspection of equation 5: when multiples are greater than 1.0, the comparison is absolutely longer than when the multiples are less than 1.0, so the mean of the comparison and the sample is higher. This means that some given abs ($t1^* - t2^*$) difference will be smaller (and thus more likely to be below threshold and thus generate a "yes" response) when normalized by a larger *m* (i.e., one deriving from absolutely longer duration values). Note that once again, the predicted asymmetry of the gradients comes from the assumed decision processes: had these been different, then the predicted gradients may not have been asymmetrical.

It is also important to note the similarity of the MCG model (Eq. 1), and its variants, and the episodic model (Eq. 5). They are both psychologically and mathematically similar, and both have the form of a thresholded normalized difference (TND) model. That is, they both use a difference between two durations, this difference is "normalized" (i.e., divided) by something, then the result compared with a threshold. The left-hand sides of both equations 1 and 5 thus represent models of subjective difference, or conversely subjective similarity, and the judgement of identity required in both types of temporal generalization task (normal and episodic) is obtained when this similarity is sufficiently great (or the difference sufficiently small) to warrant the decision that the durations compared are the same.

MODELLING TEMPORAL BISECTION

The temporal generalization task offers simple examples of how decision processes play a central role in the explanation of timed behaviour. The task of temporal bisection, which has been even more frequently used in recent studies of human timing, offers some slightly more complicated cases. In bisection (Wearden 1991b), as outlined earlier, the participant is initially presented with two standards, one identified as "short" (*S*, e.g., 200 ms) and another as "long" (*L*, e.g., 800 ms). Following these standard presentations, comparison durations (e.g., from 200 to 800 ms in 100-ms steps) are presented, and the participant is required to make one response ("short") if the comparison is judged more similar to *S* and another response ("long") if the comparison seems more similar to *L*.

The result of this procedure is a psychophysical function usually graphed in terms of the proportion of "long" responses plotted against comparison stimulus duration. In student-age adults, such a function has an ogival shape, ranging from virtually no "long" responses when the comparison duration which is actually *S* is presented to nearly 100% "long" when *L* is.

The psychophysical function can be analyzed to yield various measures, but the two that have attracted most interest are: (i) the bisection point, the comparison duration giving rise to 50% "long" responses; and (ii) the Weber ratio, essentially a Weber-fraction-like measure of temporal sensitivity which reflects the steepness of the psychophysical function.

The location of the bisection point has provoked the most discussion. In animal studies it is almost invariably located at the geometric mean of *S* and *L*, the square root of their product (400 ms with the example stimuli given above, although most studies with animals use longer standard durations). Studies with humans yield mixed results. Most find that the bisection point is located at or near (usually just below) the arithmetic mean of *S* and *L* (500 ms in the example given above), although other results are sometimes found.

Modelling the decision processes involved in bisection appears at first sight very straightforward. The bisection point might be supposed to be duration value that is equally similar to both *S* and *L*. If time accrues on a linear scale, then the point, *t*, equidistant from *S* and *L* is their arithmetic mean, that is, the point when $L - t = t - S$, is $t = (S+L)/2$, so such a simple difference decision process is consistent with arithmetic mean bisection, which is close to what is found in many cases. Unfortunately, other results make such a simple model impossible to sustain.

Wearden and Ferrara (1995) challenged such a simple conception of bisection performance in two ways. Firstly, they showed that, for a constant *S*/*L* pair (e.g., 200 and 800 ms), the location of the bisection point depended on the spacing of comparison durations between *S* and *L*: if the spacing was logarithmic then the bisection point was lower than when it was linear. Secondly, Wearden and Ferrara (1995) compared performance on "normal" bisection of the type described above (which they called "classification bisection") with another method which did not involve explicit identification of *S* and *L* at all. In their novel bisection method ("partition bisection") people just received a set of durations, for example, from 200 to 800 ms in 100-ms steps, and were required to classify each as either "short" or "long", using any criterion they wished; that is, they were required to partition the set of durations into two. The set was repeated 20 times, and when data from the last 10 repetitions were combined into a psychophysical function, such a function was identical to that obtained with the classification method.

It seemed, therefore, that not only did stimulus spacing between *S* and *L* affect the classification of a duration, *t*, so that, for example, *t* could be judged more similar to *S* in one set and more similar to *L* in another, but also that no explicit identification of *S* and *L* was necessary at all for bisection. These results cast doubt on any models which just use *S*, *L*, and *t* as the basis for classification of *t* (as stimulus spacing effects cannot be reconciled with such a model), and also cast doubt on whether participants are even making any kind of comparisons of presented durations with *S* and *L* at all (see also Allan 2002).

Wearden and Ferrara (1995) proposed that, rather than using *S* and *L* as the basis for classifying some comparison, *t*, as "short" or "long", participants are actually using the arithmetic mean of the set of all the stimuli presented. Such an idea has two immediate advantages. Firstly, a model based on it will "automatically" predict stimulus spacing effects. For example, the bisection point is shifted to the left with logarithmic spacing because the mean of all the durations is lower with logarithmic compared to linear spacing. According to this reasoning, a "reverse logarithmic" spacing with a mean higher than the arithmetic mean should shift the bisection above the arithmetic mean, and it does (Wearden and Ferrara 1995). The second advantage of Wearden and Ferrara's model was, of course, that the similarity of behaviour under classification and partition bisection was predicted: in both cases, participants are just using the average of all the durations presented, regardless of explicit identifications of *S* and *L* in the classification case.

Specifically, Wearden and Ferrara (1995) proposed that people classify some duration, *t*, as "long" when

$$(t - M)/t < b \qquad (6)$$

and "short" when

$$(t - M)/t < -b \qquad (7)$$

and when

$$-b < (t - M)/t < b \qquad (8)$$

a condition of ambiguity, the model responds "short" or "long" at random with equal probability. In Eqs. 6, 7 and 8, *M* is the arithmetic mean of the set of presented durations (assumed to be represented as a Gaussian distribution with mean *M* and some coefficient of variation, *c*), and *b* is a fixed threshold.

Inspecting Eqs. 6, 7, and 8, the reader immediately notes that: (i) they are all of the TND type; and (ii) they closely resemble the decision rules used for temporal generalization, in particular the fixed-threshold model (Eq. 3).

It should be acknowledged that there are other models of bisection (e.g., Allan and Gibbon 1991, Killeen et al. 1997) than that mentioned above, and I have conducted a recent unpublished review of an extensive set studies of temporal bisection in humans which concluded that that the best model was a slight modification of that of Wearden and Ferrara's model, one which assumes that participants are using 95% of the arithmetic mean of the set of durations instead of 100%, although the arguments justifying this position are too lengthy to be repeated here.

It can be seen therefore that the basic TND structure of decision processes needed to model temporal generalization (where judgements are about the identity of durations) can be extended to temporal bisection (where the judgements appear to involve similarity of durations). In fact, the TND structure implies that identity is just thresholded similarity, and the same rules can apply to both sorts of decisions about durations.

WHAT AFFECTS DECISION PROCESSES?

Are there factors in timing experiments that affect decision processes alone? One example, from Wearden and Grindrod (2003), has already been presented above and the reader will recall that we used differential payoff manipulations to encourage or discourage participants from identifying comparisons as the standard in temporal generalization. Another similar example comes from Wearden and Culpin (1998) who used differential payoffs for errors made to different comparison stimuli. The standard was always 500 ms, with comparison durations ranging from 200 to 800 ms. Points were gained for correct responses and were deducted for errors, but in different conditions identifying one comparison stimulus as the standard was particularly heavily punished by point loss. In one case this was the comparison (600 ms) just longer than the standard, and in another the comparison (400 ms) just shorter. This manipulation shifted the shape of the generalization gradient, presumably by affecting decision processes, although no modelling was conducted to test this.

A more interesting example of a factor that is both widespread in timing experiments and which may affect decision processes directly is task difficulty. Ferrara et al. (1997) noted such an apparent effect in a temporal generalization experiment with a 600-ms standard. Two different groups had non-standard comparisons durations spaced in either 150-ms or 75-ms steps around the standard, and the two groups thus experienced some common stimuli (i.e., both experienced trials where 750 ms and 450 ms had to be compared with the 600 ms standard). Ferrara et al. found that the discrimination between the comparisons and the standard was better in the group that performed the more difficult discriminations overall (the group with the 75-ms spacing). This result was replicated by Wearden and Grindrod (2003) who again found apparently greater timing sensitivity when the discrimination to be performed was more difficult. The problem in these cases is to understand how trials which involve physically identical stimuli, and identical comparisons, can actually produce behaviour which is different.

Ferrara et al. reviewed data from bisection experiments and found that if the difficulty of the bisection task is assayed by the *L/S* ratio (with smaller ratios indicating greater task difficulty), then Weber ratios (a measure of temporal sensitivity on bisection tasks, where lower values indicate greater sensitivity) were systematically lower when the bisection task was more difficult.

The commonsense explanation of these task difficulty effects is that subjects "pay more attention" when the task is difficult than when it is easy, and thus discriminate stimuli better. However, this explanation, while intuitively plausible, does not explain the mechanism by which performance is changed, nor does it fit clearly into other "attentional" effects on timing. Almost all attentional effects on timing involve some sort of concurrent interference (for a review see Brown 1997) so, for example, timing is poorer if a concurrent non-timing task must be performed simultaneously with a timing task, or when the participant must time more than one duration at once (e.g., Brown and West 1990). Here, the normal explanation is in terms of attentional or resource "division" between the two tasks. However, in the case of the effects noted by Ferrara et al. (1997) there is no concurrent interfering task, and participants report that the are trying to perform as well as they can on the timing task in all conditions. However, their performance apparently improves when the task is harder!

Ferrara et al. (1997) discussed several possible explanations of task difficulty effects, and were able to model the effect on temporal generalization in terms of a lower (i.e., more strict) threshold (*b* in Eq. 1) in the more difficult case. This effect is illustrated (with imaginary data) in the upper panel of Figure 4. Here, temporal general-

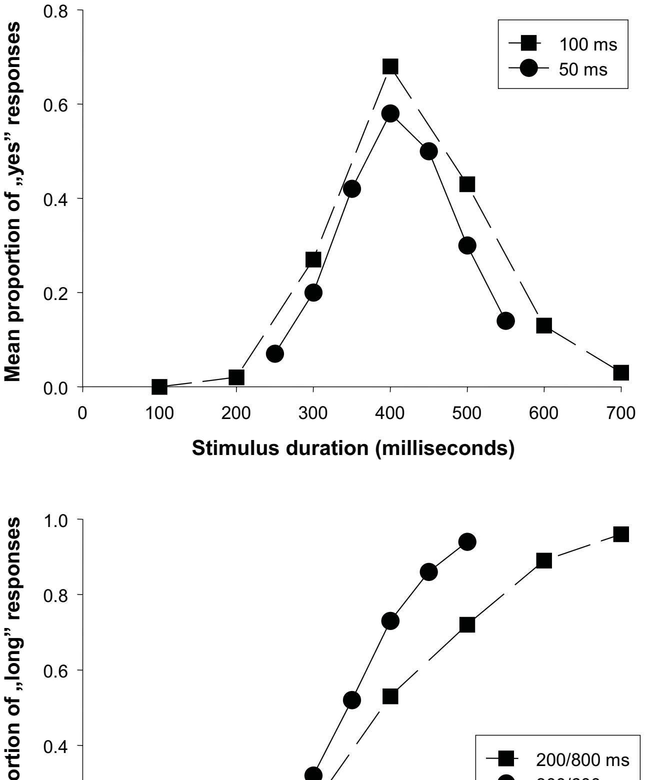

Fig. 4. Simulation of task difficulty effects. All results come from simulations. Upper panel: comparison of temporal generalization with 100-ms and 50-ms non-standard duration spacing. The MCG model produced the results, and the threshold was lower in the 50-ms case. Lower panel: simulation of bisection performance. Results come from the Wearden and Ferrara (1995) model. Conditions compared involve a 200/800 ms *S/L* pair and a 300/600 ms *S/L* pair, with the threshold being lower for the latter simulation.

ization gradients are generated by the MCG model, but the difference between the conditions shown resides in a threshold difference only. A condition with a 400-ms and 100-ms comparison spacing around the standard is contrasted with a more difficult condition with a 50-ms spacing. The only difference in the simulations was that the threshold, *b*, was lower for the 50-ms case, with everything else remaining the same.

Obviously, such a threshold change changes the generalization gradient in the direction of greater apparent timing sensitivity with the stricter threshold. Overall, "yes" responding is lower at all comparison stimulus values, but the finer discrimination in the 50-ms case is evidence by comparing the comparison stimuli common to both sets. The 300-ms comparison generates 40% of the "yes" responses that occur at the standard in the 100-ms case, but only 34% in the 50-ms case. Likewise, the 500-ms comparison generates 63% of the "yes" responses that occur at the standard in the 100-ms case, but only 51% in the 50-ms case.

Ferrara et al. (1997) had no such ready explanation of bisection performance, yet the similarity of the models used for bisection (Eqs. 6, 7, and 8) and those used for generalization (Eqs. 1-4) encourage the same sort of explanation. Fig. 4 shows that this logic is sustainable. In the (imaginary) data shown (lower panel of Fig. 4), the Wearden and Ferrara (1995) bisection model is used to simulate performance with a 200/800 ms *L/S* pair and a 300/600 ms *L/S* pair. In both cases the comparison durations are spaced in equal linear steps between *S* and *L*. The only difference between the simulations was that the threshold, b in Eqs. 6, 7, and 8, was lower for the 300/600 ms case, reflecting its putatively greater "difficulty". Obviously, the gradient of the psychophysical function is steeper (indicating higher timing sensitivity) with more the more conservative *b* value.

If we assume that task difficulty effects of the type noted by Ferrara et al. (1997) are mediated by changes in threshold, then they can only be explained by models which have such a modifiable threshold, so demonstrating task difficulty effects in some particular task is tantamount to requiring a threshold-based model for the task. Episodic temporal generalization also demonstrates task difficulty effects, as unpublished work conducted in collaboration with Steven Grice demonstrates, but Eq. 6 (the episodic model) contains a threshold and can thus potentially account for this effect with threshold changes. In fact, task difficulty effects in episodic generalization can be modelled by threshold changes in the same way as Ferrara et al. achieved for "normal" generalization.

The arguments above show that there may be some effects in the timing literature which a mediated solely by their effects on decision processes, rather than by changes in the "raw" perception of time, such as those resulting from changes in the operation of the internal clock.

Overall, the temporal generalization and bisection tasks (and related procedures like categorical timing (Wearden 1995) provide clear examples of the operation of decision mechanisms in timing, and show how our understanding of performance on such tasks depends centrally on understanding decision mechanisms as well as other parts of the timing system.

NEW HORIZONS: MODELS OF OTHER TASKS

The "mechanics" of psychological models of timing like SET enable us to develop precise quantitative accounts of how people are performing of timing tasks rather than just mathematical fits to the data that people produce. That is, the models specify the psychological processes involved in some timing task in precise detail (although, naturally, the models that exist will not always be accurate or correct). What is true of temporal generalization and bisection is also true of other tasks: without a detailed psychological model of the task, differences in performance in different conditions or between different groups (children, elderly, patients of one sort or another) remain very difficult to interpret precisely. Yet, some quite commonly used tasks (such as the threshold determination procedure extensively used in drug studies by Rammsayer and colleagues (e.g., 1997) apparently do not have any specified psychological model, so remain "empirical" techniques which tell us much less about timing than they would if psychological models of them were developed and tested.

However, it should be acknowledged that developing such models is no trivial task. I have concentrated in this article on temporal generalization and bisection because they are both simple and have more than a decade of theoretical development behind them. Some other timing procedures, particularly those used in classical time psychology such as reproduction and verbal estimation not only may not have any clear theoretical model which accounts for them, but may also be resistant to simple models of the sort discussed in this article. I recently reported a proto-model of temporal reproduction (Wearden 2003) which is basically compatible with the principles of SET, and which fits some aspects of reproduction data well (although it fits others less well), but developing a model of another classical task, verbal estimation, presents some formidable challenges.

I will outline only some of these here, to illustrate possible avenues of development. In a procedure commonly used in my work (e.g., Penton-Voak et al. 1996, Wearden et al. 1998) people receive a series of auditory or visual stimuli, all of which are too short to make counting useful, and have to assign verbal labels to their duration estimates, using a scale of 1000 = 1 second. The behaviour that results has a number of properties. Firstly, mean estimates grow linearly (but not accurately) with real time. Secondly, coefficient of variation of estimates (standard deviation/mean) tends to decrease with increasing duration, although the shortest duration presented can have a small coefficient of variation also (see Wearden 1999 (Fig. 2, p. 10)). Thirdly, verbal estimates are almost invariably "quantized", so numbers ending in "00", or (less frequently) "50", are almost the only ones ever used, accounting for more than 98% of estimates in the data I collect. So, for example, a person presented with a tone which is actually 457 ms long never says "457" but, instead "400", "500", or "550".

To model verbal estimation, not only do we need some model of how "raw" time "sensations" increase with real time (a scalar timing mechanism might serve for this), but we need to tackle the tricky problem of quantization, how internal sensations are translated into the numbers that are produced. In the case of verbal estimation, it seems particularly clear that the quantization process "translates" the raw temporal sensation into the behaviour observed, so two different individuals could actually "feel" the stimulus durations in exactly the same way, yet report different verbal estimates. Likewise, measures of the mean and variability of verbal estimates are affected both by changes in internal sensation as stimulus duration changes, and by the quantization process.

Figure 5 shows an outline of a potential model of verbal estimation. In the upper part, two stimuli (S_1 and S_2) give rise to different duration sensations. However, the verbal responses are quantized to certain values, shown in the bottom part of the diagram as R_0 to R_4. One simple rule might be for the sensation, S, to be translated into the closest permitted response. This poses no problem for S_1, which produces the verbal response R_1, but S_2 is ambiguous, being equidistant from R_3 and R_4, and modelling might permit probabilistic choice, so sometimes R_3 occurs and sometimes R_4. However, the differential variability of responses to S_2 is not a result of its intrinsic variability (although this may well exist, so that an S value does not always give rise to the same sensation) but of the quantization process involving translation between internal sensations and permitted verbal responses. Another complexity is that if the sensation is systematically changed (e.g., by a "speeding up the clock" manipulation, see Penton-Voak et al. 1996), then the verbal estimates of some durations may be much more affected than others. For example, if S_2 in Fig. 5 is "lengthened", then R_4 may become much more probable than R_3, so the variability of verbal estimates of S_2 may change, but the effect is largely due to the quantization of verbal estimates. Increasing the subjective duration of S_2 would be expected, according to scalar principles, to increase its subjective variability, but pushing S_2 towards a value almost invariably quantized as R_4 might reduce variability in its verbal estimates.

Obviously, the development of a precise quantitative model of verbal estimation remains a significant challenge, and seems likely to be a much more complex task than the development of models of procedures like tem-

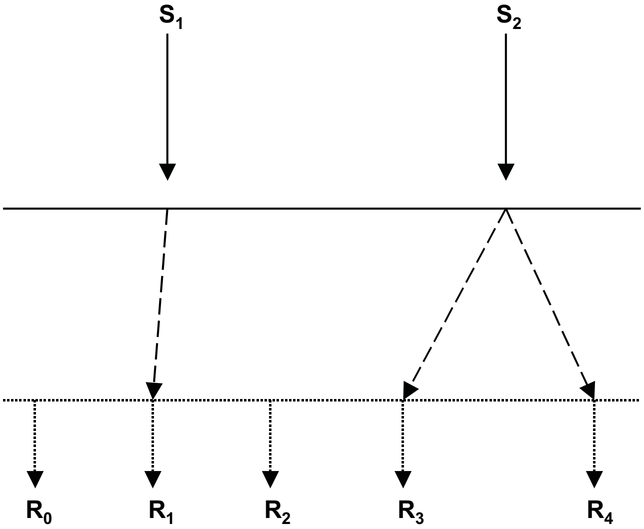

Fig. 5.Sketch of a potential model of "quantization" in verbal estimation. S_1 and S_2 are two stimulus durations, and R_0 to R_4 permissible verbal responses. Dashed arrows between the horizontal lines show potential quantization processes.

poral generalization and bisection. Nevertheless, the effort involved in developing such a model may well deepen our understanding of how people scale durations in conventional units, and enable us to understand more precisely than before just why the results of experiments on verbal estimation have the properties that they have.

CONCLUSIONS

The principal points of this article might be simply summarized.

Firstly, decision processes play an essential role in the leading contemporary model of timing, scalar expectancy theory (SET).

Secondly, the predictions of performance on tasks like temporal bisection and generalization may depend more critically on the decision processes chosen than by some other aspects of the theoretical explanation (such as where timing variance is in the system).

Thirdly, decision processes proposed for different tasks involving judgements of the similarity or identity of durations appear to have a common form: thresholded normalized difference (TND). Here, the absolute difference between a to-be-judged duration and some kind of standard is normalized (i.e., divided) by some quantity, and the result is compared with a threshold.

Fourthly, the decision processes that might be operating in tasks like verbal estimation of duration may be much more complicated than any studied up to this time: however, a precise specification of the decision processes may be critical to understanding behaviour on verbal estimation and other "classical" timing tasks.

This article will have achieved its aim if the reader (i) has been convinced of the importance of decision processes in understanding timing; and (ii) has gained some insight into the way some simple models employing decision processes have been used. More generally, the study of differences in timing between groups, or attempts to understand how the brain mediates timing, might well benefit from some consideration of decision processes. Different groups may differ in timing performance not because of fundamental differences in their "internal clock", but because of differences in memory or decision mechanisms. Likewise, a search for brain mechanisms of timing should not concentrate solely on putative mechanisms which might serve as a "timekeeper", such as oscillating neurons. While the search for a timekeeper (or timekeepers) in the brain is valuable and important, purely psychological studies which, as the present article shows, have reached high levels of quantitative precision, argue that the search for memory and decision mechanisms relating to timing is at least as important.

REFERENCES

Allan LG (2002) Are the referents remembered in temporal bisection? Learn Motiv 33: 10-31.

Allan LG, Gerhardt K (2001) Temporal bisection with trial referents. Percept Psychophys 63: 524-540.

Allan LG, Gibbon J (1991) Human bisection at the geometric mean. Learn Motiv 22: 39-58.

Brown SW (1997) Attentional resources in timing: Interference effects in concurrent temporal and non-temporal working memory tasks. Percept Psychophys 59: 1118-1140.

Brown SW, West AN (1990) Multiple timing and the allocation of attention. Acta Psychol (Amst) 75: 103-121.

Burle B, Bonnet M (1999) What's an internal clock for? From temporal information processing to the temporal processing of information. Behav Processes 45: 59-72.

Church RM, Gibbon J (1982) Temporal generalization. J Exp Psychol Anim Behav Processes 8: 165-186.

Droit-Volet S (2003) Alerting attention and time perception in children. J Exp Child Psychol 85: 372-384.

Droit-Volet S, Wearden J (2002) Speeding up an internal clock in children? Effects of visual flicker on subjective duration. Q J Exp Psychol 55B: 193-211.

Droit-Volet S, Clément A, Wearden JH (2001) Temporal generalization in 3- to 8-year-old children. J Exp Child Psychol 80: 271-288.

Ferrara A, Lejeune H, Wearden JH (1997) Changing sensitivity to duration in human scalar timing: an experiment, a review, and some possible explanations. Q J Exp Psychol 50B: 217-237.

Gibbon J, Church RM, Meck W (1984) Scalar timing in memory. In: Annals of the New York Academy of Sciences, 423: Timing and time perception (Ed. J. Gibbon and L. Allan) New York Academy of Sciences, New York, p. 52-77.

Jones LA, Wearden JH (2003) More is not necessarily better: examining the nature of the temporal reference memory component in timing. Q J Exp Psychol 56B: 321-343.

Jones LA, Wearden JH (2004) Double standards: memory loading in temporal reference memory. Q J Exp Psychol 57: 55-77.

Killeen PR, Fetterman JG, Bizo LA (1997) Time's causes. In: Time and behaviour: psychological and neurological analyses (Eds. C.M. Bradshaw and E. Szabadi) Elsevier Science, Amsterdam, p. 79-113.

Lejeune H, Wearden JH (1991) The comparative psychology of fixed-interval responding: Some quantitative analyses. Learn Motiv 22: 84-111.

Lejeune H, Ferrara A, Simons F, Wearden JH (1997) Adjusting to changes in the time of reinforcement: peak interval transitions in rats. J Exp Psychol Anim Behav Proc 23: 311-331.

McCormack T, Brown GDA, Maylor EA, Darby A, Green D (1999) Developmental changes in time estimation: comparing childhood and old age. Dev Psychol 35: 1143-1155.

Meck WH (1983) Selective adjustment of the speed of internal clock and memory processes. J Exp Psychol Anim Behav Proc 9: 171-201.

Penton-Voak IS, Edwards H, Percival A, Wearden JH (1996) Speeding up an internal clock in humans? Effects of click trains on subjective duration. J Exp Psychol Anim Behav Proc 22: 307-320.

Rammsayer TH (1997) Are there dissociable roles of the mesostriatal and mesolimbocortical dopamine systems on temporal information processing in humans? Neuropsychobiology 35: 36-45.

Rodriguez-Girones MA, Kacelnik A (2001) Relative importance of perceptual and mnemonic variance in human temporal bisection. Q J Exp Psychol 54A: 527-546.

Treisman M, Faulkner A, Naish PLN, 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 JH (1991a) Do humans possess an internal clock with scalar timing properties? Learn Motiv 22: 59-83.

Wearden JH (1991b) Human performance on an analogue of an interval bisection task. Q J Exp Psychol 43B: 59-81.

Wearden JH (1992) Temporal generalization in humans. J Exp Psychol Anim Behav Processes 18: 134-144.

Wearden JH (1995) Categorical scaling of stimulus duration by humans. J Exp Psychol Anim Behav Processes 21: 318-330.

Wearden JH (1999) "Beyond the fields we know...": exploring and developing scalar timing theory. Behav Processes 45: 3-21.

Wearden JH (2003) Applying the scalar timing model to human time psychology: progress and challenges. In: Time and Mind II: Information-processing perspectives (Ed. H. Helfrich). Kluwer, Dordrecht, p.21-39.

Wearden JH, Bray S (2001) Scalar timing without reference memory: episodic temporal generalization and bisection in humans. Q J Exp Psychol 54B: 289-310.

Wearden JH, Culpin V (1998) Exploring scalar timing theory with human subjects. In: Time and the dynamic control of behavior (Eds. V. DeKeyser, G. d'Ydewalle and A. Vandierendonck). Hogrefe and Huber, Gottingen, p. 33-49.

Wearden JH, Ferrara A (1995) Stimulus spacing effects in temporal bisection by humans. Q J Exp Psychol 48B: 289-310.

Wearden JH, Grindrod R (2003) Manipulating decision processes in the human scalar timing system. Behav Processes 61: 47-56.

Wearden JH, Penton-Voak IS (1995) Feeling the heat: body temperature and the rate of subjective time, revisited. Q J Exp Psychol 48B: 129-141.

Wearden JH, Wearden AJ, Rabbitt P (1997a) Age and IQ effects on stimulus and response timing. J Exp Psychol Hum Percept Performance 23: 962-979.

Wearden JH, Denovan L, Fakhri M, Haworth R (1997b) Scalar timing in temporal generalization in humans with longer stimulus durations. J Exp Psychol Anim Behav Processes 23: 502-511.

Wearden JH, Edwards H, Fakhri M, Percival A (1998) Why "sounds are judged longer than lights": application of a model of the internal clock in humans. Q J Exp Psychol 51B: 97-120.

Wearden JH, Parry A, Stamp L (2002) Is subjective shortening in human memory unique to time representations? Q J Exp Psychol 55B: 1-25.

Received 12 September 2003, accepted 11 December 2003