THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (11), 1875–1908

Psychology Press
Taylor & Francis Group

Review article

Scalar properties in animal timing: Conformity and violations

Helga Lejeune
University of Liège, Liège, Belgium

J. H. Wearden
Keele University, Keele, UK

The article reviews data from animal subjects on a range of timing tasks (including fixed-interval and temporal differentiation schedules, stimulus timing, aversive conditioning, and Pavlovian methods) with respect to conformity to the two scalar properties of timing behaviour: mean accuracy and scalar (Weberian) variance. Systematic deviations were found in data from temporal differentiation schedules, timing of very short (<100 ms) or very long (>100 s) durations, effects of "task difficulty", and some special cases where circadian and interval timing seemed to interact, or where some specific durations seemed to be timed more precisely than others. Theoretical reconciliation of some of these deviations with underlying scalar timing can be achieved, but a number of problematical cases remain unexplained.

The purpose of the present article is to review data obtained from a range of experimental procedures used with animal subjects with respect to the conformity of behaviour to the *scalar properties* of timing. Conformity of behaviour to the scalar properties, defined later, is a requirement of Gibbon, Church, and Meck's (1984) scalar expectancy theory (SET), but can also be consistent with other theoretical positions such as the behavioral theory of timing, BeT (Killeen & Fetterman, 1988), Machado's (1997) learning to time model, LeT, and the packet theory of Kirkpatrick and Church (Kirkpatrick, 2002; Kirkpatrick & Church, 2003, 2004).

Scalar timing involves two properties, *mean accuracy* and *scalar variance*, which we define in detail in a subsequent section, but it is important to distinguish two types of conformity to scalar timing. The first is *empirical scalar timing*: that is, the question of whether or not observed data exhibit the scalar properties. The second is *theoretical scalar timing*: the question of whether the behaviour observed is consistent with underlying time representations having the two scalar properties, even if empirical scalar timing is not found.

The two main sections of our article are concerned, respectively, with empirical scalar timing (i.e., we review a large body of data and discuss

---

Correspondence should be addressed to J. H. Wearden, School of Psychology, Dorothy Hodgkin Building, Keele University, Keele, Staffordshire, ST5 5 BG, UK. E-mail: j.h.wearden@psy.keele.ac.uk

The preparation of this article was supported by a Joint Project Grant from the Royal Society and by a grant from the Belgian National Foundation for Scientific Research.

http://www.psypress.com/qjep
DOI:10.1080/17470210600784649

conformity or violations of the scalar properties in the data set) and theoretical scalar timing (i.e., we discuss specifically the question of whether or not deviations from empirical scalar timing can be reconciled with underlying scalar representations). The first main section thus provides information that is useful with respect to the evaluation of any theory of animal timing. The inventor of some new theory, for example, could compare the performance of their model with the data reviewed to ascertain strengths and weaknesses in their own work. The second main section, which discusses principal violations of empirical scalar timing, is not restricted to the framework of SET, although it borrows some ideas from this theory, such as a specification of how time representations are generated. However, in general, this section might be useful for theorists developing models incompatible with SET, as it shows some general ways in which additional processes may reconcile apparently deviant data with an underlying scalar timing process.

After this preamble, our article has three principal sections. We begin with some definitions as to what constitutes conformity to the properties of scalar timing and what might be violations. A second, larger, part of the article then reviews data from a number of procedures commonly used to test animal timing. To anticipate conclusions to be drawn later, it becomes apparent from a review of data obtained that violations of scalar timing are not random and are much more likely in some situations than others. The third section discusses these violations and, in particular, attempts some theoretical explanation of why they occur. A brief conclusion then terminates the article.

THE SCALAR PROPERTIES OF TIMING BEHAVIOUR: DEFINITIONS

Following from a series of influential articles published in the early 1980s by John Gibbon and his colleagues (Church & Gibbon, 1982; Gibbon & Church, 1981; Gibbon et al., 1984), a two-part definition of conformity of behaviour to the scalar properties developed. The first part was mean accuracy: the requirement that measures of timed behaviour should vary linearly and near-accurately with imposed time constraints. So, for example, an animal's "estimate" of some real time, t, should on average be close to t as this is varied. If this internal estimate is directly reflected in behaviour, then measured behaviour should also vary linearly with t and in some conditions also accurately. In other cases, the mean accuracy property may result in a proportional relation between behaviour and the time requirement.

However, as is seen later, the property of mean accuracy is not always found. An obvious violation might be that the mean of some measure of timed behaviour varies, not linearly or accurately with an imposed time requirement, but as some nonlinear function of it: for example, a power function with a fractional exponent, which produces substantial deviations between the behaviour measure and the duration requirement.

The upper panel of Figure 1 schematizes some possible relations between mean measures of behaviour and time. For all the panels in Figures 1, 2, and 3, the results are invented and are generally less noisy than data obtained in experiments, but are representative of effects found. Three cases are shown in the upper panel of Figure 1: In one (accurate) the behaviour measure tracks time perfectly; in another (proportional) the behaviour measure is some constant proportion of time. The third case (nonlinear) shows a situation where the behaviour measure and time are related nonlinearly. The function is in fact a power function, $m = kt^n$, where m is the mean measure, and t is the time unit, with $k = 1.5$, and $n = 0.5$. With these parameter values, the mean behaviour measure is close to those obtained with linear and proportional relations when the time value is small, but deviates markedly from the other two as time grows.

The second critical property of scalar timing is scalar variance. This is a requirement that concerns not the mean of measures of timed behaviour, but the variability of behaviour around the mean. Scalar variance is the requirement that the standard deviation of measures of behaviour varies linearly with the mean, and this can be tested in

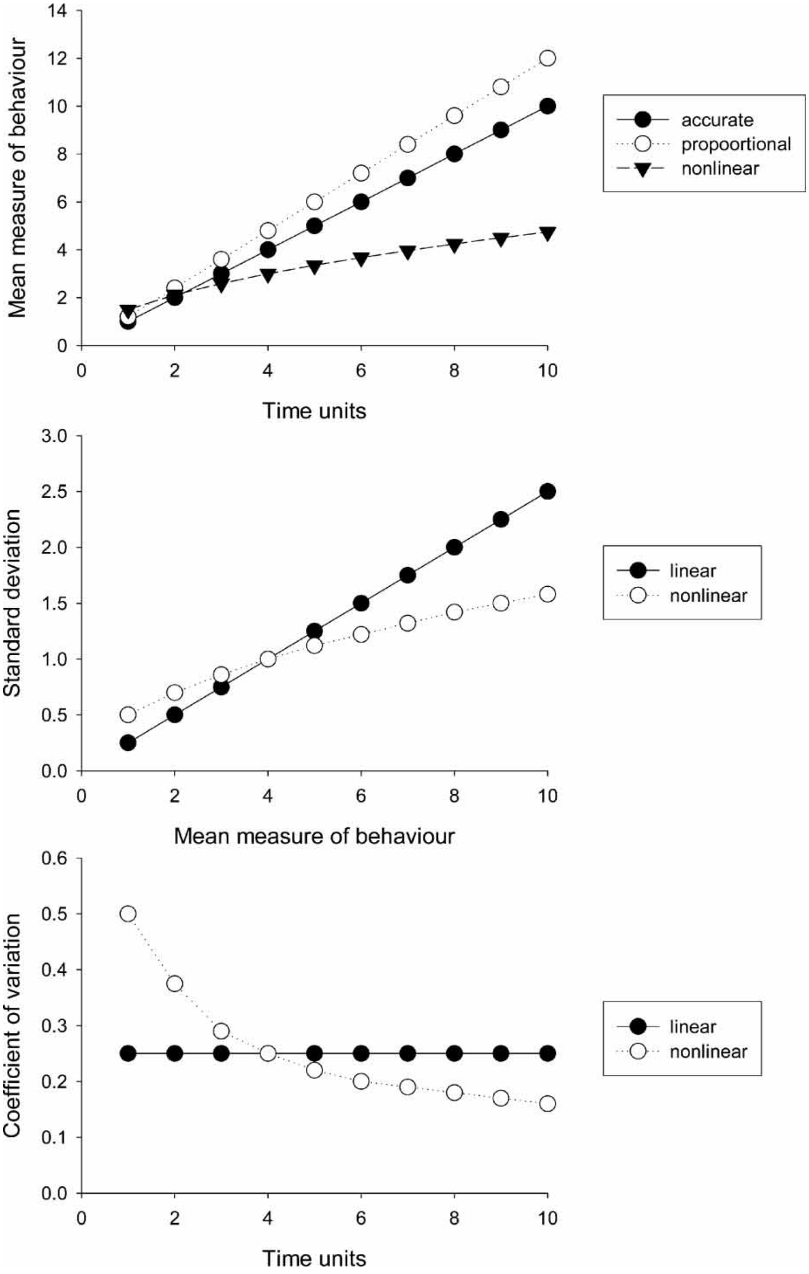

Figure 1. *Upper panel: Potential relations between mean measures of behaviour obtained from a timing experiment and the time units used. Relations shown are accurate (filled circles), proportional (open circles), and nonlinear (triangles), with the nonlinear represented by a power function. Centre panel: Potential relations between standard deviation and mean behaviour measure from a timing experiment. Relations shown are linear (filled circles), and nonlinear (open circles). Bottom panel: Coefficient of variation (standard deviation/mean) plotted against time units, using the values from the centre panel (filled and open circles, respectively).*

various ways. The simplest is to take the standard deviation (s) and the mean (m) of behaviour measures when some time constraint or requirement, t, is varied. Regression of s against m should yield a linear function with a high r^2, to exhibit conformity to scalar variance. Another method of testing scalar variance is to use each s and m to construct a *coefficient of variation* ($CV = s/m$), a Weber-fraction-like measure. The property of scalar variance requires CV to remain constant as t is varied. Plotting CV against t can produce some striking examples of conformity to the scalar property, as in Zeiler (1985) where the coefficient of variation on a temporal differentiation task remained at about 0.38 for t values from 5 to 80 s, a 16-fold range of durations.

The middle and bottom panels of Figure 1 show examples of tests of the property of scalar variance. The middle panel shows a case where the relation between standard deviation and mean is linear (the standard deviation is always 0.25 of the mean) and another case where the standard deviation grows as the square root of the mean behaviour measure, generating the nonlinear relation shown. The former would indicate conformity to the property of scalar variance, the latter a violation of it. The bottom panel shows the same values as those in the centre one, but this time converted into coefficients of variation, by dividing the standard deviation by the mean behaviour measure. Now the linear case in the centre panel generates a constant coefficient of variation as the time unit changes, whereas the nonlinear relation shows a declining value. The former is an example of conformity to scalar timing, the latter a violation of it.

Figures 2 and 3 show illustrations of another popular way of measuring the timing behaviour of animals. The upper panel of Figure 2 shows a response measure (for example, rate or probability of an operant response, probability or amplitude of a Pavlovian conditioned response) plotted against a time measure, such as elapsed time since some time marker (like previous food, stimulus duration, or interstimulus interval in Pavlovian conditioning). The invented results come from two conditions, one where the "critical" time (e.g., time of reinforcement) was 5 time units, and another where it was 10. The response measures shown have two features used for evaluating conformity to scalar timing. One is the peak location, the time value at which responding peaks, and this can be determined in various ways. In the upper panel of Figures 2 and 3, Gaussian curves are fitted to the results (a common procedure, see Lejeune & Wearden, 1991, or Whitaker, Lowe, & Wearden, 2003, although not the only one, see Church, Meck, & Gibbon, 1994). The location of the peak of the curve is determined by a curve-fitting procedure, and peak location can be examined with respect to the "critical" time: If the peak location varies accurately or proportionally with the critical time, then this shows conformity to the scalar property of mean accuracy, whereas nonlinear relations violate it.

The Gaussian curves also yield a measure of the width of the curve fitted to the results, and this can be used to test the property of scalar variance. In the upper panel of Figure 2, the invented results exhibit the scalar property: The standard deviation of the curve fitted to the "10 unit" curve is twice that of the "5 unit" curve. In the upper panel of Figure 3, imaginary results are shown, which conform to mean accuracy (i.e., the peak is at the "critical" times, 5 and 10 units), but violate scalar variance: Here the widths (standard deviations) of the two curves are the same, indicating relatively more sensitive timing of the "10 unit" condition.

Another way of testing whether scalar variance is exhibited in data is to use the method of *superimposition*. This involves plotting measures of timed behaviour from different conditions on the same relative scale. For example, in Church and Gibbon's (1982) study temporal generalization gradients were obtained from a range of standard values. When the comparison durations used were expressed as a fraction of the standard in force for the condition, then plotted on the same graph, the temporal generalization functions obtained with different standard values all superimposed. The scale used here is a relative one, because all durations used in a particular condition are "normalized" by the standard value in force for that condition (see also Church, Lacourse, & Crystal, 1998).

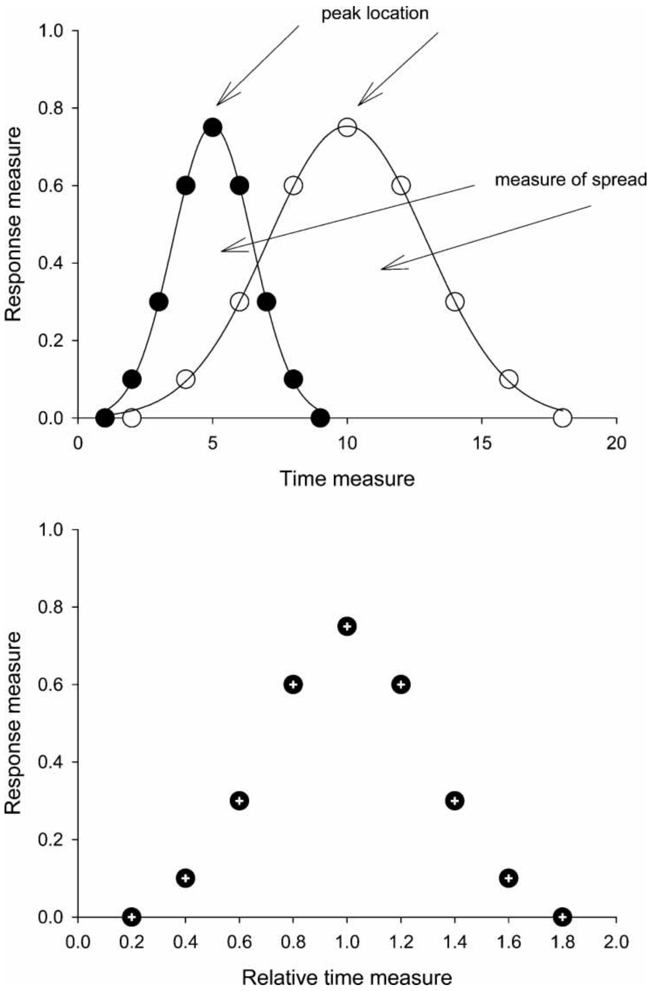

Figure 2. *Upper panel: Response measures plotted against time units for two conditions where the "critical" times were either 5 or 10 time units. Gaussian curves were fitted to the results, and two measures commonly obtained (peak location and some measure of spread) are indicated. The standard deviation of the 10-unit curve was twice that of the 5-unit curve, in conformity to scalar timing. Bottom panel: The results from the upper panel plotted against relative time (where each time measure is divided by the "critical" time for that condition, 5 or 10 time units). The results superimpose perfectly, so only one set of points is shown.*

The lower panels of Figures 2 and 3 show examples of superimposition derived from the results shown in their respective upper panels. Response measures from the "5 unit" and "10 unit" conditions are plotted against relative time (the time measure divided by the "critical" time, 5 or 10). In the lower panel of Figure 2, the results superimpose perfectly, whereas in the lower panel of Figure 3 superimposition is violated, with the relatively more precise timing in the "10 unit" condition being clearly shown.

A general consideration when assessing conformity or violation of empirical scalar timing is that both the mean accuracy and scalar variance properties may be more clearly exhibited, or more clearly violated, in experiments with a wide range of absolute time values, which provide better tests of both these properties, and we generally

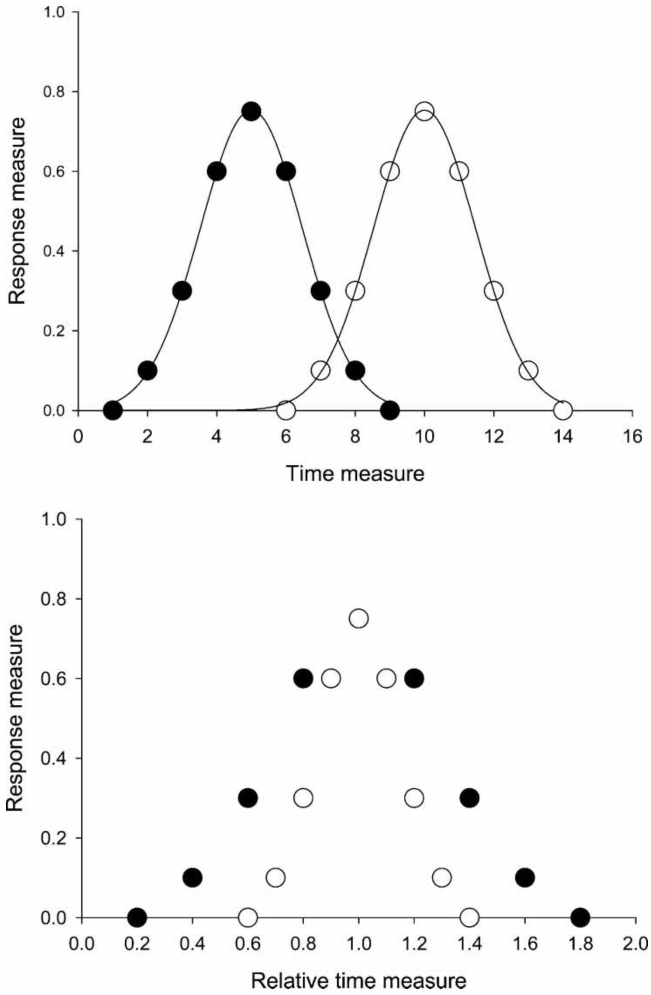

Figure 3. *Upper panel: Response measures plotted against time measures for conditions similar to those shown in Figure 2; Gaussian curves fitted to the results are shown. In this case the standard deviations of the fitted curves were the same. Bottom panel: Results from the upper panel plotted against relative time. In this case the failure of superimposition (relatively more precise timing from the 10-unit condition) is observable.*

concentrate on these in our review. When absolute time values are close together, data are necessarily ambiguous, as deviations from the scalar properties would have to be gross to be detectable.

A further issue is what kind of deviations from the scalar properties are regarded as "real" violations. For example, the lack of constancy in CV with changes in timed duration that is nevertheless still considered as reflecting underlying scalar timing may differ from one report to another, although "random" deviation from constancy is, presumably, always less serious than systematic deviations. A related problem occurs for deviations from proportional timing. Deviations from simple proportionality, if very small, may be regarded as consistent with underlying scalar timing, whereas large or systematic deviations, particularly if these are manifested in a number of studies using similar methods, will usually be considered to be deviations.

In the section that follows, we review conformity of behaviour to the mean accuracy and scalar variance properties, where data come from commonly used methods employed with animals. We exhibit a distinct preference for discussion of data collected over the last 30 years rather than earlier work (although this is sometimes mentioned), as early work often lacked measures that would enable the scalar property to be identified even if present (e.g., absence of measures of variability of behaviour).

EXPERIMENTAL STUDIES

In the next section, we review data obtained from studies of laboratory animals (usually rats or pigeons, but occasionally other species as well) tested on five sorts of procedure: fixed interval (FI), variants of FI, and related operant methods; temporal differentiation schedules; stimulus timing; aversive conditioning; and Pavlovian procedures. The articles whose data we discuss represent only a fraction of those that could potentially be included, but it seems to us that our selection represents the main trends present in the field.

FI schedules, FI variants, and related operant procedures

We discuss here data obtained from simple FI or fixed-time (FT) schedules and their variants like mixed-FI schedules and the peak-interval procedure as well as some other operant methods related to FI.

FI and FT schedules

On a simple FI schedule, the first response occurring *t* s or more after the start of the interval (which is usually timed from the previous reinforcer delivery, or from the onset of a tone or light signal) is itself reinforced. After a period of training on simple FI, the response pattern within the interval becomes temporally differentiated, with little or no responding occurring early in the interval, followed by either a gradual acceleration in responding to a peak at or near the interval value (Dews, 1978), or an abrupt transition from a low rate of responding to a high steady rate continuing until the end of the interval (Schneider, 1969). When responding from a number of intervals is aggregated together, the response pattern is well fitted by the left-hand side of a Gaussian distribution centred on the FI value (Lejeune & Wearden, 1991): for example, the left-hand parts of the curves shown in the upper panels of Figures 2 and 3.

Among the measures of timed behaviour that have become standard are (a) the postreinforcement pause (the time from the start of the interval to the first response occurring in it), (b) the breakpoint (the time from the start of the interval to the period of high-rate responding, when this can be identified), and (c) measures derived from fitting Gaussian or similar curves to averaged response rate versus elapsed time functions. The averaged response rate functions can also be superimposed, by plotting the response rate against elapsed time divided by the proportion of the interval elapsed (yet other measures are described in Richelle & Lejeune, 1980).

Virtually all studies find that the postreinforcement pause increases with the FI value, although the increase is usually better described by a power function with a fractional exponent than by a linear relation. This can be seen, for example, from cyclic-FI data where the duration of the interval increased and decreased from $2t$ to $8t$, with t values of 2, 4, 10, 20, and 40 s (Innis & Staddon, 1971; Schneider, 1969). Less than proportional pausing can also be seen in Ludvig and Staddon (2004, 2005). Other examples can be found in Lowe, Harzem, and Spencer (1979), Hanson and Killeen (1981), and Zeiler (1999, for pause data from pigeons on FI values ranging from 1 to 100,000 s).

Although postreinforcement pause means are not linear functions of the FI value, postreinforcement pause variability does, however, exhibit scalar variance, by showing a standard deviation that is a linear function of the mean pause length (Lowe & Wearden, 1981). The finding of scalar variance was also present in earlier data (e.g., McKearney, 1969). Scalar variance was also found in a fish species: The relationship between the standard deviation and the mean of wait time, a kind of breakpoint measure on FI schedules, could be described by a linear fit (Higa & Simm, 2004).

Breakpoints and related measures, on the other hand, tend to increase more linearly with FI value. Gibbon (1977) replotted data from rats exposed to progressive FI schedules ranging from 1 to 50 min on double log coordinates (Harzem, 1969). Mean time to the second response after reinforcement could be described by a linear function with a slope of 1.016 (see also Hanson & Killeen, 1981). Similar data were also found using goldfish (Talton, Higa, & Staddon, 1999) and Siamese fighting fish (Higa & Simm, 2004).

Superimposition properties in data from FI schedules can also be found when response rates (as proportions of terminal rate) are plotted against *t* on a relative time scale (e.g., proportions of the FI value), as shown by Dews (1970), for *t* values of 30, 300, and 3,000 s. Similar data came from an FI variant (10,000, 30,000, and 10,000 s), where the duration of the interval was broken into segments signalled by alternating S^D and S^{Δ} stimuli—stimuli in the presence of which responding either is followed by reinforcer delivery or is extinguished (Dews, 1965; see also Hienz & Eckerman, 1974). However, timing sensitivity seemed to be higher for the 100,000 s duration (which is close to 24 hours) than for both other durations. More recent data showing superimposition can be found in Crystal (2001a) for long FIs (3 and 7 hr; see his Figure 3, p. 71), but not for still longer FIs (14, 22, 26, and 34 hr) where, as in Dews (1965), timing sensitivity was much higher for durations close to 24 hr (22 and 26 hr) than for both other durations (see also, Crystal, 2003).

Using the left-hand part of Gaussian curves to fit response rate versus elapsed time in the interval functions (e.g., Figures 2 and 3), Lejeune and Wearden (1991) determined coefficients of variation from a range of animal species. For all data sets that had more than two FI values, a range of values could be found for which the coefficient of variation was approximately constant (for example, for pigeons over FI values ranging from 240 to 600 s), although in most data sets the coefficient of variation increased at long FI values.

The Gaussian curve-fitting procedure used by Lejeune and Wearden (1991) forced the peak of the Gaussian curve to be at the FI value, thus effectively enforcing the property of mean accuracy. However, such curves always fitted data well and thus indirectly confirmed the mean accuracy property for FI. Other curve-fitting analyses of response functions from FI have also effectively assumed mean accuracy (e.g., Lowe & Harzem, 1977), and once again fitted curves described the data well. The property of scalar variance has also been found using "general activity" or other nonoperant activities of pigeons and rats on FT schedules (Gibbon, 1977; Silva & Timberlake, 1998).

Mixed-FI schedules

On two-valued mixed-FI schedules, reinforcers are potentially available for responses at two different times after the start of the interval, although only one time of reinforcement is effective in any individual interval, and nothing signals to the animal which is in force currently. In some studies (e.g., Whitaker et al., 2003), the two potential times of reinforcement are equally likely, but in some others they are not (e.g., Catania & Reynolds, 1968). Whitaker et al. (2003) analysed data from mixed-FI schedules where the two potential times of reinforcement were close together (e.g., mixed FI 60 s, FI 80 s), or further apart (e.g., mixed FI 30 s, FI 240 s). In the latter case, responding during the long (240 s) intervals increased from low levels early in the interval to a peak at or near the earlier potential time of reinforcement (30 s) and then declined, only to rise again to a peak at the later time of reinforcement (240 s). Data from the longer intervals of all mixed-FI schedules were fitted as the sum of two Gaussian curves, one corresponding to the contribution of the shorter interval and the other to that of the longer one.

Whitaker et al.'s (2003) data conformed closely to the mean accuracy property: When the peak of the response function associated with the lower FI was calculated it was usually very close to the FI value, as this varied between 15 and 120 s in different conditions. The curve-fitting procedure could also test for scalar variance by comparing the coefficient of variation of the curve fitted to the lower FI with that for the higher FI. If scalar timing holds, then these two values should be the same. Data from 92 cases were examined, and overall there was no clear difference in coefficient of variation for the short and long durations (see also Leak & Gibbon, 1995).

Fetterman and Killeen (1995) developed a variant of mixed FI, involving reinforcer delivery at three potential times of reinforcement, in their *categorical timing* procedure. Pigeons responded in a three-key operant chamber, each key being associated to a different time from the start of the trial. When data from the different conditions were separately plotted on the same relative scale (Fetterman and Killeen, p. 50, Fig. 4) they had nearly identical forms, suggesting good superimposition, thus confirming the presence of scalar variance.

The peak interval (PI) procedure

The peak interval procedure (Catania, 1970; Roberts, 1981) is again a mixed schedule, but this time consisting of a mixture of simple FI trials and longer trials involving extinction (the *peak trials* from which data are obtained). The normal averaged pattern of behaviour is that responding increases from near zero at the start of the peak interval, to a peak somewhere near *t*, and then declines again (i.e., the form shown in the upper panel of our Figure 2). The location of the peak and measures of its variability (e.g., standard deviation, which can be divided by peak location to yield a coefficient of variation) can be determined by curve fitting (as in our Figure 2).

Data from the peak trials of the PI procedure, like those from mixed-FI schedules, can test both scalar properties. Mean accuracy is tested by examining the peak location and comparing this with *t*, the time of reinforcement on FI trials: For mean accuracy, these should be the same. Most data indicate that if the peak is not exactly at the FI value, it is certainly close (Church et al., 1994; Lejeune, Ferrara, Simons, & Wearden, 1997; Lejeune, Ferrara, Soffié, Bronchart, & Wearden, 1998b; Roberts, 1981; Roberts, Cheng, & Cohen, 1989).

Scalar variance can be tested by calculating a standard deviation or coefficient of variation for the response function and plotting this against the FI value, or by superimposing the response functions on the same relative scale (e.g., elapsed time in the interval scaled by the FI value in force). For example, Church et al. (1994) used rats and measures derived from individual trials on PI and found that both the means and standard deviations of these measures increased linearly with FI values ranging from 15 to 60 s. This was confirmed later on with pigeons and FI values of 5, 10, and 20 s (Cheng & Miceli, 1996; see also Cheng & Roberts, 1991). Superimposition was found using rats (Church et al., 1998), pigeons (Gibbon, Fairhurst, & Goldberg, 1997a), and starlings (Rodriguez-Girones & Kacelnik, 1999).

Zeiler and Powell (1994) used pigeons with a larger range of FI values in their PI studies than those used by other workers (7.5 s to 480 s), and they obtained a more complex pattern of results. Overall, mean accuracy as well as the scalar variance was not found, the only exception being the near constancy of the coefficient of variation of breakpoint-like acceleration points.

Complex procedures involving FI or PI

A number of studies have used more complex operant procedures that contain FI-like components, but other schedules as well, and they also involve the choice between two alternatives.

The first of these is a concurrent chain procedure, which is a conceptually simple, although procedurally rather complex, variant of the PI procedure (Grace & Nevin, 1999). Pigeons received initial training on a “normal” PI procedure, with FI values of 10 and 20 s, or 20 and 40 s, delivered on a centre key. Then, the intertrial intervals of the PI procedure were replaced by a choice phase, where two white side keys were available, and pecks on one or the other eventually resulted in centre key illumination and a PI trial, thus embedding PI trials in a concurrent-chain procedure.

Inspection of data suggests that the response versus time in the peak trial functions were almost identical in the "normal" and "concurrent-chain" PI variants, with responding peaking at the FI value and spreads of the response rate function being higher at the longer FI value, suggesting superimposition. Peak location was a linear function of the FI value (supporting mean accuracy), and the interquartile range of the response function, a measure of variability, also increased linearly with the FI value, providing another confirmation of scalar variance.

The second example is the "time-left" procedure (Gibbon & Church, 1981), which also is a type of concurrent schedule. In the most commonly used form of the time-left procedure (following Gibbon & Church, 1981, Exp. 2) animals are initially confronted with concurrent alternatives (e.g., a left white key and a right red one), and responding on these initial links leads to mutually exclusive terminal links, with the transition usually being governed by variable-interval schedules. One of the links (the standard link), when entered, produces a change of key colour (e.g., from red to green), and the reinforcer is arranged according to an FI schedule some fixed time after entry into the link. For the other alternative, the "time-left" key, there is no stimulus change when the link is entered, and the time to reinforcer availability from entry into the link depends on the time elapsed from the start of the trial when the link was entered. Suppose that the total trial time (C) was 60 s, and the time to reinforcement after entry to the standard terminal link (S) was 30 s. If T is the elapsed time in the trial when the initial links are available, then at short T values responding on the key leading to the standard terminal link should be preferred, as if this link is entered, the reinforcer is available after a short delay. For example, with $T = 5$ s, the time to reinforcement on the standard link (if activated) is 30 s, whereas the time to reinforcement on the time-left link is 55 s ($C - T$). Later in the trial, however, the time to reinforcement on the time-left link is shorter than that available on the standard (e.g., at T values >30 s in our example), so preference should shift to the initial link alternative leading to the time-left terminal link, and it does (e.g., Gibbon & Church, 1981).

The time-left procedure is complicated and has had a number of interpretations (e.g., see Cerutti & Staddon, 2004; Jozefowiez, Cerutti, & Staddon, 2005; McCarthy & Davison, 1986; Preston, 1994; Staddon & Higa, 1999, for just some examples). Our focus here is simply in whether behaviour on this schedule demonstrates scalar timing. One way to examine this is to look at the indifference point: the elapsed time in the interval when responding on the keys leading to the different terminal links was equal. Gibbon and Church (1981) reported data from a number of conditions where $S = C/2$ (that is, the time to reinforcement on the standard link was always half the trial length), and S was varied over values from 15 to 90 s. Indifference points varied linearly with S (Gibbon & Church, 1981). Gibbon, Church, Fairhurst, and Kacelnik (1988) presented data from pigeons with S values ranging from 7.5 to 60 s (S was always $C/2$). When the preference functions were normalized by the indifference point, good superimposition was found, indicating conformity to the property of scalar variance.

Interval time–place learning and foraging

Two lines of research involving operant response timing procedures aim to simulate more naturalistic conditions in which hungry animals seek food. The first is time–place learning, the second foraging simulations.

Interval time–place learning, where food is available for fixed intervals of time (not exceeding several minutes) at different locations within the experimental chamber, is similar to the categorical timing procedure of Fetterman and Killeen (1995) described above. For example, Carr and Wilkie (1998) trained different groups of rats on a procedure where food was sequentially available at each of four locations. For different groups of rats, these temporal windows lasted for 4, 6, or 8 minutes, during which responding on the appropriate lever was intermittently reinforced. Percentage of maximum response rate on each lever, plotted as a function of relative time bins, superimposed, and the peak of activity was located at the middle of each availability period (a "mean accuracy" index). Furthermore, the relation between the width of response rate functions and the duration of the temporal windows could be fitted with a linear regression line with zero intercept, in agreement with Weber's law. However, on an unequal time–place task—that is, a within-session design where different time intervals are associated with successive response levers (e.g., 6, 4, 2, and 8 min in Thorpe & Wilkie, 2002, or 60, 30, 30, and 60 s in Crystal & Miller, 2002)—response rate distributions failed to superimpose, and the scalar property was violated. Thorpe and Wilkie related their findings to arguments developed by Grondin (2001), according to which Weber's law fails when a sequence of intervals has to be learned or after extensive training, conditions that are both present in the experiment (see also Thorpe, Bates, & Wilkie, 2003).

Foraging is another important line of work using operant methods and relating timing of food delivery to the foraging decisions that animals make. For example, Brunner, Kacelnik, and Gibbon (1992; see also Kacelnik & Brunner, 2002) studied the behaviour of starlings in a foraging simulation. The starlings were initially required to make 20 perching responses alternating between two perches, to simulate the travel between two patches. When this requirement was completed, a "patch key" was illuminated, and the birds could peck this to obtain between 1 and 4 reinforcers delivered according to an FI schedule. A return to either of the perches reset the perch requirement. Two measures of behaviour on the patch key were of interest: "giving-in time" (i.e., the time of the last peck on the patch key), and the rate and distribution of responses throughout the FI. Median and interquartile range of giving-in times increased linearly with the FI schedule in force on the patch key, and the response rate versus elapsed time functions of the FI probe trials had a peak and a standard deviation that also increased linearly with the FI value, suggesting both mean accuracy and scalar variance.

Summary

Overall, some consistent trends emerge from data on FI schedules and their variants. First, postreinforcement pauses vary as power functions of FI values, thus violating the mean accuracy property, although the scalar property of variance is often found in pauses. Breakpoint-type measures generally seem to increase more linearly with FI value and may also exhibit the property of scalar variance. Second, measures derived from response rate versus elapsed time functions generally show evidence of near-perfect mean accuracy and scalar variance. The second of these conclusions seems to be generally true at "short" FI values (e.g., up to a few hundred seconds). At longer values, most studies find increasing coefficients of variation (i.e., greater than proportional increase in variability), derived from various measures of performance (see, e.g., Gibbon, Malapani, Dale, & Gallistel, 1997b).

Temporal differentiation schedules

On a temporal differentiation schedule, a time constraint for reinforcement is imposed on a response or group of responses. An early procedure of this type was differential reinforcement of low rate (DRL), where the time between two consecutive responses (the interresponse time, IRT) had to exceed some minimum value for the second response to be reinforced. Other examples are those where a response (e.g., lever holding) must have a certain duration to be reinforced on its termination (differential reinforcement of response duration, DRRD), or must be delayed for *t* seconds from a signal to be reinforced (differential reinforcement of response latency, DRRL). More complex procedures involve imposing a temporal constraint on a sequence of responses (ratio time differentiation: e.g., 30 responses have to be made in at least *t* s for the ratio to be reinforced), or nonresponding for *t* s (differential reinforcement of other behaviour, DRO). Finally, duration reproduction in animals, where the duration of a signal has to be reproduced, for example, by holding down a response device for at least the same duration, has also been used.

Temporal differentiation studies have two features that complicate interpretation. First, the imposed time constraint is usually a minimum. This means that the times at which reinforcement occurs that are actually experienced by the animal must necessarily be longer than *t*, and these are often not reported in articles (for a discussion of this issue, see Zeiler, 1983). Second, the rate and patterning of reinforcers delivered on temporal differentiation schedules is critically dependent on the performance of the animal, unlike FI, where marked variations in the rate and patterning of responding may have little effect on the interval between reinforcer deliveries.

In temporal differentiation, the mean accuracy property can be tested by plotting the mean time that the animal takes to make the critical response or responses, against the time requirement, *t*. The scalar property of variance has mainly been tested in temporal differentiation schedules by examination of the standard deviation of response measures with respect to their mean: Obviously, this relation should be linear for the scalar property to be confirmed. Several early studies using temporal differentiation methods report means (e.g., mean IRT or other measure), but not any measure of variability. Therefore, to aid exposition here, we discuss the mean accuracy property first, then the scalar property of variance.

Temporal differentiation and mean accuracy

The relation between mean times produced and time requirements in temporal differentiation is illustrated by data samples from DRL, response latency differentiation, DRRD, ratio duration differentiation, DRO, and duration reproduction.

Many studies using conventional operants like key pecking and lever pressing report that the relation between the mean and *t* is a power function rather than a linear relation, and this is particularly the case if longer time requirements (effectively greater than about 10–20 s with key-peck operants and pigeon subjects) are used (e.g., see Richardson & Loughead, 1974). However, an early study by Wilson and Keller (1953) reported a linear relation between median IRT and DRL value from rats over the range 10–30 s, and Staddon (1965) also described linear relations for pigeons on DRL values less than about 20 s, but other studies have usually reported power relations, particularly when pigeons are used (Lejeune & Jasselette, 1986; Lejeune & Richelle, 1982).

Catania (1970) reported a study of temporal differentiation of response latency in pigeons, using time requirements from 0.60 to 48 s, and found that power functions with fractional exponents fitted data well (although data from the longest and shortest time requirements were not used). Similarly, power function relations between mean or median response time produced and time requirements were found for temporal differentiation of the duration of a lever-holding responses by Platt, Kuch, and Bitgood (1973) and Kuch (1974), for key-peck duration by Zeiler, Davis, and DeCasper (1980), and for ratio duration by Zeiler (1983).

In contrast, when "unconventional" operants are employed (such as perching for pigeons or platform residence for small rodents), the mean times produced on temporal differentiation (DRL or DRRD) may be much closer to the imposed time requirement (Jasselette, Lejeune, & Wearden, 1990; Lejeune, Cornet, Ferreira, & Wearden, 1998a; Lejeune, Huynen, & Ferrara, 2000).

Zeiler (1985) used a key-peck operant with pigeons, but employed it in a novel way. Pigeons essentially switched between a schedule that reinforced "nonresponding" (DRO) to one on which responding was reinforced, and different minimum periods of "nonresponding" were imposed for reinforcement. In this case, linear relations between the duration of the "nonresponding" period emitted and the time requirement were obtained, even when time requirements were long (up to 80 s), so this experiment provided consistent evidence for proportional timing.

Finally, on a duration reproduction task using pigeons, where durations (ranging from 4 to 29.8 s) had to be reproduced by the time taken to complete a series of 15 key pecks plus an upper bound, the average duration reproduced could be equally well described by power or linear functions of the standard durations.

The best performance was obtained under the most demanding conditions, where more than four duration standards had to be reproduced in the same session (Zeiler & Hoyert, 1989).

Temporal differentiation and scalar variance

As well as relations between mean times produced and time requirements, some studies have examined relations between measures of variability and mean (e.g., standard deviation and mean). In an early study on response latency differentiation, Saslow (1968) reported semi-interquartile ranges proportional to median latencies in monkeys trained to release a key after a light or a click signal. However, Weber's law was violated in the ratio duration differentiation experiment of Zeiler (1983), in that the standard deviation of the completion times was reported to grow as a power function of the mean time, although it is unclear how well a simple linear relation between the standard deviation and mean would have fitted. Further data on ratio time differentiation can be found in DeCasper and Zeiler (1974, 1977). Platt (1979) summarized many early results from temporal differentiation studies by regressing standard deviation against mean time produced (where the mean was often a power function of the time requirement, as mentioned above). Linear relations between standard deviation and mean were often found, indicating that the property of scalar variance was present in the data, even though mean accuracy was not.

Some more recent work also confirms the finding of scalar variance in temporal differentiation. For example, Jasselette et al. (1990) fitted Gaussian curves to the relative frequency of times produced and found that the standard deviation of the curve grew proportionally with the mean, indicating a constant coefficient of variation (see also Lejeune & Jasselette, 1986; Zeiler, 1985) However, a constant CV was not obtained in the first attempt at training duration reproduction in rats (Mandell & Atak, 1982), whereas it was for two pigeons out of four when different durations had to be reproduced during the same session (Zeiler & Hoyert, 1989). Similarly, Zeiler (1991) found evidence of a constant CV in pigeons performing on a modified DRL schedule in a closed economy (i.e., an operant situation in which the animal gained its entire daily ration of food from responding in the experiment).

Summary

Overall, therefore, temporal differentiation schedules present a mixed picture for scalar timing. The mean accuracy property is frequently violated, particularly with pigeons, with conventional single-response operants like key pecking or lever pressing, and with long (>10-s) time requirements, although response measures such as perching or periods of "nonresponding" show that linear relations can be obtained, even when imposed time requirements are relatively long. Data indicating linear relations between mean times produced and time requirements can also be obtained from ratio duration differentiation. In contrast, most studies providing data relevant to the scalar variance property (e.g., constant CVs with changing duration or linear relations between standard deviation and mean) find evidence for it.

Stimulus timing

In this section, we discuss techniques with the common feature that whether a response is going to be reinforced or not, or which of two responses will be reinforced, depends on the duration of stimuli that have been previously presented. We discuss in turn temporal generalization, bisection, and discrimination procedures.

Temporal generalization

The simplest technique, although one that has not been commonly used with animals, is temporal generalization. Church and Gibbon (1982) tested rats in a situation where periods of darkness of various durations were presented, immediately followed by availability of a retractable lever. A response on the lever was reinforced if the period of darkness had lasted a certain standard duration, but not if durations longer or shorter than the standard were presented. The response measure, effectively a temporal generalization gradient, was the probability of the response plotted against stimulus duration. Such generalization gradients peaked at the standard as this was varied between conditions (e.g., over values of 2, 4, and 8 s), and the spread of the gradient increased proportionally to the standard, so that gradients from different absolute durations superimposed when plotted on the same relative scale. Thus, data from temporal generalization demonstrated both mean accuracy and scalar variance. A more recent study by Weisman et al. (1999) reported temporal generalization gradients from zebra finches and humans using shorter standard durations (211, 374, and 663 ms), and reasonable empirical superimposition was obtained from both species.

Bisection

A method much more frequently used with animals is bisection. There are two main methods employed, the more common one being derived from Church and DeLuty (1977). In their study, rats were initially trained to discriminate two different standard durations (e.g., visual stimuli 2 and 8 s long; 2 s was the short standard, *S*, 8 s the long standard, *L*). A two-lever operant chamber was used, and if *S* had been presented a response to one lever was reinforced; if *L* was presented a response to the other lever was reinforced. The rats mastered this discrimination to a high degree of accuracy after a few sessions of training, and then bisection testing began. Essentially, this involved presenting the rat with a series of stimulus durations, including *S* and *L*, but also intermediate durations (after which responses were not reinforced), and noting which lever was pressed after each one. If we define a response on the lever that was appropriate to *L* during training as a LONG response, then Church and Deluty found that behaviour took the form of a psychophysical function of ogival shape, going from near-zero LONG responses after *S* to nearly 100% after *L*.

The psychophysical function can be analysed to yield a number of measures, but most attention has focused on the *bisection point* or *point of subjective equality*, PSE (the duration giving rise to 50% LONG responses) and measures of temporal sensitivity—that is, either the slope of the psychophysical function (which can be computed from a straight line fitted to the function or by curve fitting, yielding, respectively, the duration values that give rise to 25 and 75% of LONG choices), or the Weber ratio—that is, the ratio between the standard deviation (half the difference between durations yielding 75 and 25% of LONG choices) and the PSE. Since psychophysical functions usually have their minimum values after *S* and maximum after *L*, mean accuracy is implied in all of them, but Weber ratios can be used to test the property of scalar variance.

Church and Deluty (1977) used *L*/*S* values of 4/1, 8/2, 12/3, and 16/4 s, and they found that the psychophysical functions superimposed well and that Weber ratios from the four conditions did not differ significantly. Some of these *L*/*S* values were also used in Maricq, Roberts, and Church (1981), with the same result of constant Weber ratios across the different conditions. Note that in the Church and Deluty study, and in others that obtained constant Weber ratios, although the absolute values of the durations used varied, the *L*/*S* ratio was kept constant at 4:1.

In a recent study, Miki and Santi (2005) compared bisection performance with filled and empty stimulus durations (based on visual stimuli) in pigeons with a within-subject design, and *L*/*S* pairs of 4/1, 8/2, or 16/4 s were used. Empty intervals were judged as longer than filled intervals of the same duration. However, Weber fractions obtained from bisection of filled and unfilled intervals did not systematically differ, and, when the *L*/*S* values were the same, data from filled and unfilled intervals superimposed.

A new variant of the procedure involved a long operant chamber with choice alternatives located far apart and a pressure-sensitive floor recording the pigeon's movement during the sample stimuli (Machado & Keen, 2003). The stimuli to be judged were presented in the centre of a long box, but the response keys, corresponding to the "short" and "long" responses, were located at different ends. This arrangement meant that not only could the classification responses be measured, but also that the movement of the pigeon could be used to investigate when the animal "decided" that the stimulus was "short" or "long". In general, the pigeons gradually developed the behaviour of spending more time close to the "short" key than the "long" one early in the stimulus presentation, but then shifted towards the "long" key as the duration of the stimulus to be classified increased (Machado & Keen, 2003). Machado and Keen's study also used some trials where stimuli longer than the "long" standard were presented, and, indeed, many of their results come from these rather unusual "generalization" trials. One striking finding was that the time that elapsed when the animals moved from being on the "short" side to the "long" side seemed better predicted by the absolute duration of the "short" stimulus than by the relation between the "short" stimulus and the "long" one. The bisection point (or PSE in these cases) remained constant as long as the "short" duration was constant, regardless of the value of the "long" standard. Such data seem to violate the scalar property of proportionality, as behaviour appears related to absolute rather than relative duration values. On the other hand, some other measures of behaviour showed superimposition. A rather different procedure developed by Machado and Keen (1999; see also Machado & Pata, 2005) produced data from 1- and 4-s, and 4- and 16-s bisection pairs that did not superimpose, although different individual animals showed different sensitivity patterns with most showing more sensitive timing on the longer duration pair.

The other, less popular, bisection method derives from Stubbs (1968). With this technique, the centre key of a three-key operant chamber is illuminated white for a period of time then, with offset of the white key, two side keys are illuminated. A response to one of the keys ("short") was reinforced if the centre key illumination period had been one of a number of "short" values, and a response to the other was reinforced if the centre key illumination had been in the "long" range. For example, in Experiment 1 of Stubbs (1968), "short" key responses were reinforced after white-key periods of 1–5 s, "long" key responses after 6–10 s. In Experiment 2, the absolute durations of the white-key periods were varied (1–10 s, 2–20 s, 4–40 s), with responses reinforced on the "short" key after white-key periods 50% or less of the longest value shown, and "long" key responses after white-key periods >50% of the longest value. Good superimposition of response proportions on the "long" key over the duration ranges used was observed.

Stubbs (1979, 1980) also designed a free-operant version of his discrete-trial bisection procedure and used longer durations ranging from 7.5 to 200 s overall. Both PSE and difference thresholds ("intervals of uncertainty") increased as a power function of the duration range. However, the slopes of these power functions were close to 1.0, indicating that a linear function may have described data almost equally well. With an analysis suggested by Gibbon (1977), the slopes of the PSE and the difference thresholds were even closer to 1, and the slope of the CVs closer to 0, in agreement with scalar timing theory (Stubbs, 1980).

In a study that does not strictly involve stimulus timing, although it is usually described as a bisection method, Platt and Davis (1983, Exp. 1) presented pigeons with two keys of different colours. One was programmed to deliver reinforcers according to FI x and the other FI y, with one value being shorter than the other but, on each trial, responses on only one of the keys would deliver food. The two FI values were varied over a range from 40 to 200 s. Inspection of data from responding on the two keys suggests that the mean accuracy property held. Response rate versus elapsed time in the interval functions were generally flatter on the "long" key than on the "short" one, although it is unclear whether superimposition would have occurred if tested. A similar result was obtained in an earlier study by Stubbs (1976), and here the curves describing the probability of a LONG response as a function of elapsed time were similar for stimulus periods of 45 and 60 s, but data from the longest stimulus period (150 s) matched the others less well.

Bizo and White (1997) also used a bisection-like procedure similar to that of Stubbs (1968). Two response keys were illuminated, either both red or both green. During training, both red and green trials lasted 65 s, and responses on the left key were reinforced on a variable-interval 30-s schedule for the first half (32.5 s) of the trial, whereas responses on the other key were reinforced in the second half of the trial. After this initial training, the trial durations associated with red and green keys were varied over a range from 20 to 110 s, but the reinforcement contingency remained the same (reinforcers for responding on one key for the first half of the trial and on the other key for the second half). When the proportion of total responses emitted on the "second-half" key were plotted as a function of elapsed time in the trial divided by the trial length, data superimposed almost perfectly for trials with reinforcement, but less well for unreinforced probe trials. Means and standard deviations of the "second-half" key responses increased with trial durations, and Weber fractions remained roughly constant.

Variants of discrimination tasks

A number of other studies have used discrimination methods similar to those employed in bisection or generalization, where the assignment of reinforcement to responses was determined either by the relation between presented stimuli and some learned standard, or by the relation between two presented stimuli.

An early example from the first type of procedure, sometimes labelled an "adjustment" or "titration" schedule, comes from Church, Getty, and Lerner (1976) who used a discrimination method in which rats were initially reinforced for responding on one lever after a short (0.5-s) signal, and on another one after a long (8-s) signal. Following this training, the rats then received testing in which the "short" standard was 0.5, 1, 2, 4, and 8 s, and the "long" duration was longer than the standard by some initial amount. If the animal made 75% or more correct responses with a particular "short"/"long" pair, then the difference between the two was reduced, until the 75% correct response criterion was met. This method essentially determines the "difference limen", the smallest duration difference between the presented stimuli that the animal could detect. Such a difference limen divided by the standard yields a Weber-fraction-like measure of temporal sensitivity. Church et al. observed that while this measure was constant at standard values of 2 s and above, it was considerably greater at 1 and 0.5 s. Thus, scalar variance held at values above 2 s, but shorter durations resulted in less sensitive timing.

Fetterman and Killeen (1992) conducted an extensive study rather similar to that of Church et al. (1976) with both pigeons and humans. For both species, a Weber-fraction-type measure of sensitivity remained constant at time values above about 100 ms, but increased markedly at shorter durations.

More recently, Crystal (1999) used rats in a discrimination procedure similar to those described above. In general, Weber fractions were highest at durations less than 1 s and greater than 10 s than at values in between. However, a more striking result was the possible existence of peaks and troughs in the sensitivity of timing. Sensitivity was analysed by methods related to signal detection theory, and sensitivity peaks were found for time values of 0.3, 1.2, 10, 24, and 36 s (see also Crystal, 2001b).

The second category of procedures, sometimes called "relational timing", has involved comparison of two stimulus durations or of duration ratios presented on a trial where, for example, one response was reinforced if the second duration was longer than the first and another response in the other case. In Dreyfus, Fetterman, Smith, and Stubbs (1988), timing sensitivity decreased as durations increased, but some evidence for scalar timing might be found in that some response functions plotted on the same relative scale seem to be superimposable. In Fetterman, Dreyfus, and Stubbs (1989), sensitivity remained constant with changes in duration ratio, and there was superimposition. In a later study, Stubbs, Dreyfus, Fetterman, Boynton, Locklin, and Smith (1994) used multiples of a range of base durations running from 2 to 22 s and, like Dreyfus et al. (1988), found that timing sensitivity decreased as durations increased. In all these articles, a very large number of different absolute durations was used, so it seems unlikely that the animals were using any kind of reference memory, making these studies rare examples of the use of animals in what has become known as "episodic" or "roving standard" timing procedures, which have recently become popular in work with humans (e.g., Allan & Gerhardt, 2001; Wearden & Bray, 2001).

Summary

Taken as a whole, stimulus-timing procedures with animal subjects yield many examples of conformity to both mean accuracy (in the sense that the stimuli discriminated usually seem to be represented accurately on average) and scalar variance. The main exceptions are relatively higher variance at very short durations (e.g., less than 200 ms) and, less reliably, at durations of several tens of seconds, although most studies have concentrated on stimulus durations from a narrower range, where both scalar properties are often demonstrated.

Avoidance procedures and aversive conditioning

Avoidance procedures provided much early evidence for scalar timing. Two procedures were frequently used: free-operant avoidance, also labelled "Sidman" avoidance (Sidman, 1953), and discrete trial avoidance. Mastering these procedures requires the estimation of the interval between successive shocks when no response occurs (the shock–shock interval, S–S) and the time between a response and the occurrence of the next shock (the response–shock interval, R–S).

Work by Gibbon (1971) on free-operant avoidance was perhaps the first to suggest that the mechanism underlying performance was scalar timing. Gibbon (1971) noted examples of apparent superimposition of response frequencies when expressed as a function of relative time (e.g., when scaled by the R–S interval in force). Gibbon (1979) provided an extensive discussion of the scalar properties of avoidance responding, and in a reappraisal of a data set from the first paper on free-operant avoidance (Sidman, 1953), Gibbon (1991) described three indices of proportional timing. First, response rates were highest when the (R–S) and (S–S) intervals had the same duration. Second, when the R–S parameter was shorter than the S–S interval, responding stopped as the R–S interval was less than about 50% of the S–S interval. Third, in the case of an R–S parameter longer than S–S, response rates declined proportionally with the duration of the R–S interval. Further, mean IRTs were equal to half of the duration of the R–S interval. Gibbon (1991) related all these features to an underlying scalar perception of time.

Libby and Church (1974) studied shuttle-box avoidance in rats, where, instead of pressing a lever, subjects had to cross a barrier located in the middle of the experimental chamber to avoid or escape the shock. They found that probability of responding increased as a function of time since the last response for R–S intervals of 10, 20, and 40 s. The procedure generated an FI-like pattern of responding, with response probability increasing with time since the last response in a form similar to the left half of a Gaussian curve. When Libby and Church plotted relative response rate as a function of proportion of the interval elapsed, the data showed reasonable superimposition.

Cotton and Wood (1982; see also Fairhurst & Gibbon, 1983) also studied shuttle-box avoidance in rats and found reasonable superimposition over R–S = S–S intervals of 10, 15, and 20 s. Also, mean IRTs were linear functions of the times when shock was due, indicating mean accuracy.

Summary

Overall, therefore, aversive conditioning procedures including avoidance usually support both scalar properties very clearly and provided some of the earliest evidence that timing in animals might be controlled by an underlying scalar-consistent mechanism.

Pavlovian methods

From a timing perspective, two issues in the study of Pavlovian conditioning are central. The first is the timing of the conditioned response (CR) with respect to the interval between conditioned and unconditioned stimuli (CS–US interval or interstimulus interval, ISI). Early work by Pavlov himself (1927/1960; see also Ellison, 1964) indicated that long ISIs could give rise to temporal control of responding, what he called *inhibition of delay*. The other issue derives from Gibbon and Balsam's (1981) claim that the rate of acquisition of a Pavlovian CR depends on the ratio of the interreinforcer interval, or cycle time, to the CS–US interval, or trial time. This acquisition speed issue has been discussed elsewhere (e.g., Kakade & Dayan, 2002) and is not our focus here (for some recent reviews discussing the role of temporal information and Pavlovian conditioning, see Cooper, 1991; Gallistel & Gibbon, 2000; Kirkpatrick & Church, 1998, 2000; Savastano & Miller, 1998).

In the present article, we confine ourselves mostly to discussing simple timing issues with respect to data from two categories of Pavlovian procedures: aversive conditioning (such as eyelid conditioning or fear conditioning), and appetitive conditioning (such as autoshaping). Our specific focus is whether CRs have scalar properties (e.g., CR onset or the "peak" of the CR amplitude that varied linearly with CS–US interval and response measures with scalar variability). Several early studies used methods that were precursors of the PI procedure, with crucial data being recorded on probe trials where the CS lasted much longer than usual and where the US was not delivered.

Aversive conditioning

Some early studies indicated that good temporal control, and possibly also scalar variance, could be observed in data obtained from rabbit nictitating membrane and eyelid conditioning. For example, using CR latency as the dependent variable, Leonard and Theios (1967) trained the nictitating membrane of rabbits with a 250-ms ISI, or 1,000-ms ISI (their control subjects). The peak of response latency in both cases was found at about half the ISI, and the spread of the response latency distribution was greater in the longer ISI case than the shorter one. Furthermore, latency distributions at 250 and 1,000-ms CS–US intervals were close to superimposable, but this was not tested. Frey and Ross (1968), using eyelid conditioning in rabbits, reported similar results.

White, Kehoe, Choi, and Moore (2000) focused particularly on the question of variability of CR timing as a function of ISI (overall range: 200 to 1,300 ms) in nictitating membrane and eyelid conditioning. Their measures were the location of the peak of responding (mean peak latency) and the coefficient of variation of this peak location (a measure slightly different from the spread of the response rate function, the more usual measure derived in operant conditioning procedures). Peak locations tracked ISIs sensitively, but peak latencies were longer than the ISI when this was short (up to 800 ms), but shorter than it at long ISI values. Coefficients of variation varied unsystematically with ISI, although the shortest ISI used (200 ms) produced the highest coefficient of variation.

In conditioned suppression, a CS followed by an unavoidable shock (US) is superimposed upon food-reinforced operant responding. During the CS, the putative suppression of responding (the conditioned emotional response, CER) is measured with respect to baseline data preceding the CS (resulting in a "suppression ratio" measure). Rosas and Alonso (1996, Exp. 1) used four CS durations ranging between 50 and 200 s and obtained suppression ratio data that superimposed when plotted on the same relative scale, in agreement with scalar timing. A related procedure, similar to Pavlov's "conditioning to time" procedure, was used by LaBarbera and Church (1974). Data from fixed shock intervals of 60 and 120 s superimposed well (see also Gibbon, 1979).

In the one-trial contextual fear paradigm, rats are placed in an experimental cubicle and experience a single unavoidable shock. The variable of interest is freezing behaviour measured immediately after the shock (Fanselow, 1986) or during a subsequent shock-free exposure to the context (Bevins & Ayres, 1995). Using shock delays ranging between 1 and 81 s, Fanselow (1986) observed a linear increase of freezing over delays ranging from 9 to 81 s. The apparent absence of freezing at shorter delays, the so-called "immediate shock-freezing deficit", was interpreted as an associative failure between the context and the shock. However, reconsidering such data, Gallistel and Gibbon (2000) suggested that the early "absence" of freezing and its subsequent increase as a function of shock delay reflected the scalar variability of freezing around the expected time of shock. At very short delays, the temporal extent of freezing may have been too small to be detected.

Recently, Drew, Zupan, Cooke, Couvillon, and Balsam (2005) studied the behaviour of goldfish in an aversive delay conditioning paradigm in which light–shock intervals were normally either 5 s or 15 s. These conditioning trials were interspersed with PI-like trials where the light was presented for 45 s without shock. The measure of performance was general activity, and on peak trials this increased to a peak close to the time of the appropriate CS–US interval (5 or 15 s), only to decline at longer times. The distribution of activity not only exhibited this mean accuracy, but also showed scalar variability, with the width of the activity distribution in time increasing with CS–US interval.

Appetitive procedures

An experiment by Holland (1998) examined temporal control in Pavlovian occasion setting. Here, two occasion setters (X and Y) signalled that target stimuli (A and B) would be followed by food. For one occasion setter/target pair the interval between the two stimuli was 10 s, and for the other pair it was 30 s. In test conditions it was found that the occasion setter was maximally effective at the time used in training, rather than at shorter or longer times, indicating mean accuracy in the judgement of the occasion setter/target interval.

Using another appetitive procedure (delay conditioning), Holland (2000) examined rate of food-cup entry in rats when food delivery was signalled by a noise CS and presented the temporal distribution of responding during the CS for different CS durations (20, 40, 80, 160 s). When total food-cup responding was expressed on a relative scale (quarters of the total time), data superimposed well, indicating scalar variance for this response.

Kirkpatrick and Church (2000) also studied rats in a delay conditioning procedure. A white-noise CS (range: 15 to 120 s) was paired with food, and head entry into the food cup constituted the measured response. During PI-like probe trials, responding peaked at the normal CS duration, and when response versus time functions were expressed on a relative scale, they superimposed reasonably well. Superimposition was also obtained on PI-like probe trials with ring doves exposed to a similar procedure, where CS durations lasted for 4, 8, or 16 s (Ohyama, Gibbon, Deich, & Balsam, 1999).

Brown, Hemmes, and Cabeza de Vaca (1997) examined proximity of the pigeon to a response key signalling food in both trace and delay autoshaping with ISIs of 18, 24, and 60 s. For the trace group, the CS was always 12 s long. When proximity measures from the different ISIs were plotted on the same relative scale data from both trace and delay autoshaping superimposed well, although response patterns were different. For delay conditioning, responding generally increased throughout the CS, whereas in trace conditioning responding peaked at about 50% of the ISI.

Summary

Overall, therefore, Pavlovian methods yield convincing evidence for scalar timing. Most Pavlovian procedures provide good evidence that the CR is temporally sensitive and that its latency usually varies proportionally with the CS–US interval. Evidence for scalar variance is less clear, but this may be mainly because many studies of eyelid or nictitating membrane conditioning have not focused on this issue. However, some evidence consistent with scalar timing comes from these preparations, from fear conditioning, and from appetitive procedures such as magazine approach and autoshaping.

PRINCIPAL VIOLATIONS AND THEIR POSSIBLE CAUSES

In this section, we try to both (a) identify the main situations in which violations of the scalar properties of time are found and (b) explain them theoretically. The main way of doing this is by attempting to reconcile violations of empirical scalar timing with underlying theoretical scalar timing. By theoretical scalar timing we mean timing with underlying representations of time (or the equivalents of time representations in "behavioural models") that have the scalar properties of mean accuracy and scalar variance. In addition to a short discussion of SET, we also briefly indicate how some other competitor models of SET produce underlying scalar properties. For more details about how different models of animal timing operate, see Wearden (1994), Droit-Volet and Wearden (2003), and Lejeune, Richelle, and Wearden (2006).

We begin with a brief exposition of the operation of the SET system. The highest level of the SET model 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 or other event of some duration, the switch is closed by event onset, allowing pulses to flow to the accumulator, and the switch opens again, cutting the pacemaker/accumulator connection, when the event terminates. Thus, at the end of the event period, the accumulator contains the number of pulses corresponding to the duration of the event. A pacemaker model like that used by SET very naturally generates the mean accuracy property. For example, if the pacemaker either (a) was completely periodic or (b) emitted pulses at random with some constant average rate (to name but two extreme cases), on average over a number of trials, n times as many ticks would accumulate in time nt than in time t, so a linear and accurate relation between the number of clock ticks on average and real time would naturally emerge.

After an event to be timed has occurred, its duration representation is considered to reside in working memory, and it may undergo changes there. The other memory store (reference memory) is considered to contain "important times" such as times at which responses have been reinforced. Finally, there is a *decision* level, where the usual comparison made is between a sample drawn from the reference memory and the contents of working memory. What the comparison process is depends on the task, but only after the comparison process is complete does behavioural output (e.g., the emission of a key peck or lever press) occur.

How does SET generate scalar variance? Gibbon et al. (1984) provide a detailed discussion, but there are a number of possible sources. One is variability in the rate of the pacemaker from one trial to another, but the other (which has historically been the more popular one, see Jones & Wearden, 2004, for discussion), is that the scalar property is present in the reference memory: That is, the animal's representation of the "critical" time for the particular procedure (e.g., the time of reinforcement on FI) has scalar variability, so on each trial, the animal uses a slightly different reference value to control current behaviour.

"Behavioural" models like BeT (Killeen & Fetterman, 1988) and LeT (Machado, 1997) operate rather differently from SET and assume that timed behaviour in animals is mediated by a series of behavioural states, which either serve as cues for measured operants (BeT) or are associatively linked to operants (LeT). The transition from one state to another is governed by a pacemaker in BeT, but this pacemaker is not directly read by the animal: Rather, its function is to move the animal from one behavioural state to the next. The existence of an underlying pacemaker process leads naturally to linear timing, and scalar variance is generated in BeT by requiring that the pacemaker rate varies inversely with the rate of reinforcement in the experimental situation, so the rate effectively changes when the duration to be timed changes (see Lejeune et al., 2006, for a more detailed discussion and evaluation of this suggestion). Machado's LeT likewise proposes a series of successive behavioural states as the basis for operant timing, but it is less clear what governs transition from one state to another. In LeT, the mathematical properties of successive states are chosen so that they generate, on average, linear relations between the ordinal number of the state currently active and real time, and the variability of the duration of successive states generates the scalar property.

Church and Broadbent (1990; see also Wearden & Doherty, 1995) suggest that duration is represented in terms of the states of members of a bank of oscillators, with each oscillator having a different period. This complex model cannot be discussed in any detail here, but oscillator values can be chosen that lead to approximately linear relations between elapsed time and measured behaviour. Scalar timing is built in to the oscillator set, so the period of each oscillator has a standard deviation that is proportional to its mean (i.e., scalar variance). Oscillator-based models have the property that not all time values are represented with equal precision, in spite of the building in of scalar variance, as we discuss later below.

Some other timing models that should be briefly mentioned are the multiple-time-scales model of Staddon and Higa (1999) and the packet theory of Kirkpatrick (2002). The former derives "raw" time representations from a memory decay process (or sometimes more than one) rather than a clock so, for example, on an FI schedule the fading memory of the reinforcer delivery that started the interval may be used as a cue to control responding. The basic memory decay function is logarithmic rather than linear, but the model can be developed to yield both proportional timing (i.e., mean accuracy), and conformity to Weber's law. In some cases, however, the model parameters can generate deviations from strict scalar variance (for example, increasing coefficients of variation with increasing duration).

Packet theory, on the other hand, assumes an underlying linear timing process as part of its basic mechanism (Kirkpatrick, 2002). The novelty of this model lies in the fact that it pays considerable attention to the local structure of behaviour, assuming that responses are emitted in "packets" separated by interpacket intervals. So, for example, when a packet is initiated the animal responds with short IRTs within the packet, then the timing mechanism is used to decide whether or not another packet should be emitted. The mechanism for generating scalar variance is not completely clear, but is linked to the idea that the number of packets emitted between reinforcers remains constant.

When discussing how violations of empirical scalar timing might be handled by timing theories, we usually just assume that some underlying linear timing process exists, as well as some mechanism for producing scalar variance, without being specific about details.

Violation 1: Temporal differentiation schedules

It is clear from our review that data from temporal differentiation schedules, where some time requirement for reinforcement is imposed on the temporal features (e.g., spacing or time to complete) of single operant responses or ratios of responses, often violate empirical scalar timing. The form of the violation most commonly involves power relations between mean measures of behaviour and imposed time requirements. In almost all cases, the form of the power function means that average measures of behaviour (e.g., mean IRT, or other response measure) may track or overshoot imposed time requirements when these are short, but progressively undershoot them when they are longer, often very considerably (see Catania, 1970, for data sets collected between 1964 and 1970). One consequence of such behaviour is that performance becomes increasingly "inefficient" at longer time requirements, in the sense that the proportion of emitted responses that is reinforced declines.

Although the focus of our review is conformity or violation of scalar properties, in the case of temporal differentiation schedules this issue is probably linked to another one, that of why temporal control of behaviour is often so "poor" on temporal differentiation schedules compared with other procedures. Lejeune (1978) pointed out differences between performances under conditions like FI, on the one hand, and temporal differentiation schedules like DRL, on the other. This "FI/DRL paradox" is illustrated by the fact that responding in pigeons with a key-peck operant shows poor temporal control on DRL values exceeding a few tens of seconds, in some case showing complete insensitivity to increases in the time requirement beyond certain values. On the other hand, FI schedules yield examples of clear temporal regulation of key pecking at time values of many minutes (e.g., see Ferster & Skinner, 1957). Given that pigeons can exhibit good temporal regulation of behaviour in some situations, their poor performance on schedules like DRL seems to severely underestimate their real capacity.

One possible explanation of this paradox is that performance is determined by a mixture of temporally regulated responses and responses not controlled by the experimental contingency, the latter type of response continually resetting the time requirement for reinforcement. Some theorists have developed models along these lines. Zeiler (1981) was the first to develop such a model involving two states: a "timing state", where timing occurs, and a "free response state" representing "base duration"—that is, the tendency to respond in the absence of any duration criterion. Computer simulation over a wide range of duration requirements (from 3 to 300 s) was successful in producing simulated output that was close to real performance on temporal differentiation experiments. Wearden (1985), following a similar line, modelled power relations between measures of behaviour on DRL, by means of an interaction between an SET-type linear timing system and "random" responses. According to this analysis, performance on key-pecking DRL is poor not because the animal's representation of the minimum reinforced duration is inaccurate but because the animal has a small probability of responding "randomly" each second, so that as the DRL requirement lengthens, more and more IRTs are terminated by the random responses. Put more casually, the animal "cannot resist" emitting the operant randomly, and this random response resets the DRL requirement. The random response probability, although constant at different DRL values, is much more likely to prematurely terminate the intended IRT when the DRL requirement is long than when it is short, so performance matches the DRL requirement well when this is short, but progressively undershoots it as the DRL value lengthens, eventually arriving at a DRL value that the animal can never reliably exceed.

Although modelling with this basic idea fits data well (see Wearden, 1985), a difficulty is that the model does not tell us how to distinguish between the putatively temporally controlled responses and those emitted at random. However, one implication of this idea is that using operants that are much more difficult to emit (e.g., those that require more energy) might decrease the likelihood of random responses and thus reveal the animal's "real" timing capacity even on DRL. Indeed, as previously described, mean accuracy and scalar variance were obtained using a perching response with pigeons (Jasselette et al., 1990; Lejeune & Richelle, 1982) or a platform response in small rodents (Lejeune et al., 1998a, 2000).

Another way of reconciling the power relation between mean measures of timing behaviour and the imposed criterion usually found in temporal differentiation schedules with linear timing has been suggested by Gibbon (1977). He proposed that the nominal imposed time criterion, *t*, should be replaced by the "functional duration"—that is, the average of reinforced durations. Taking data from response duration differentiation in rats (Platt et al., 1973) or from reinforcement of long key-peck latencies in pigeons (Catania, 1970) as examples, Gibbon showed that when mean behaviour measures were regressed against averaged reinforcement time, slopes ranged from 0.92 to 1.02 for rats, and 0.88 to 0.95 for all birds except one, whereas values were 0.82 to 0.98 (Platt, Kuch, & Bitgood, 1973) or 0.73 to 0.88 (Catania, 1970) when the regression involved mean response measure and the criterion *t* (Gibbon, 1977). However, Zeiler (1983) questioned whether apparent deviations from scalar timing obtained under temporal differentiation conditions could always be explained in this way.

Some aspects of the effects explained by "random responses" might also be related to another idea, that of responses occurring in "bouts", which are groups of responses separated by short IRTs and distinguished by interbout pauses from other bouts. Although the idea that animals might respond in bouts has a long history, this idea has taken on renewed significance with the development of packet theory (e.g., Kirkpatrick, 2002). The short IRTs occurring within bouts might be controlled by different variables from those controlling the interbout IRTs. Indeed, some early studies (e.g., Blough, 1963; Malott & Cumming, 1964, 1966; Millenson, 1966) considered that short IRTs might be insensitive to reinforcement contingencies, or the result of stereotyped response patterns, whereas longer IRTs might be more malleable by reinforcement contingencies. The existence of "uncontrollable" short IRTs obviously has some effect on performance on schedules like DRL. On the other hand, as Wearden (1990) showed by computer modelling, the existence of such short IRTs may make little difference to the overall reinforcement rate obtained on DRL: In fact, if the animal is forced by some species-specific behavioural process to emit "uncontrollable" responses, the shorter the IRTs of these responses, the fewer reinforcers per hour are lost on a schedule like DRL, where the most maladaptive strategy is to emit IRTs just shorter than the DRL requirement, rather than very short IRTs.

Another possible difference between performance on temporal differentiation schedules and FI might relate to the stimulus control during conditioning and the temporal "structure" of the schedule or the "periodicity" of occurrence of reinforcer-bound stimuli (the noise of the feeder, the sight of the reinforcer being dropped, some changes of visual stimuli, etc.). According to this view, the difference between temporal differentiation and FI behaviour might be related to differences in the periodicity of the delivery of the reinforcing event (primary reinforcer and related external stimuli). In principle, FI and temporal differentiation schedules are both "periodic" schedules, as on temporal differentiation schedules a reinforcer can be obtained every *t* seconds provided that perfectly timed responding occurs. In practice, however, periodicity in the delivery of the reinforcing event is never reached on temporal differentiation schedules, therefore, the "synchronizing" power of the reinforcing event, present in FI, is usually absent during temporal differentiation.

A strong argument in favour of the synchronizing effect of the periodic reinforcing event comes from experiments using FI where reinforcer-associated stimuli were suppressed (Deliège, 1975). In this study, rats were exposed to FI schedules in enclosures preventing the occurrence of any acoustic or visual stimulus associated with the delivery of the reinforcer, which was a food pellet silently dropped into a velvet-lined recess located behind an opaque flap. In spite of successful previous magazine training and lever-press shaping, no rat exhibited the typical FI response pattern. Performance was highly irregular and was eventually extinguished. This result shows that the mere FI contingency stripped of its visual and acoustic correlates cannot sustain behaviour, which can become variable and "temporal differentiation-like" when external synchronizing factors are absent.

Does the absence of regular temporal structure mean that performance on temporal differentiation schedules never fits the requirements of scalar timing? The answer is clearly "no", as several examples have been described in our previous sections. It is, however, perhaps notable that none of these experiments used single standard responses (e.g., key peck or leverpress).

Violation 2: Very short and very long durations

Almost all studies that have used durations less than 100 ms have reported increased relative variance (i.e., increased CV or Weber-fraction-like measures) at very short durations relative to those obtained from longer durations. Such an effect is exactly what would be expected from a pacemaker–switch–accumulator clock mechanism like that proposed by SET, and it is also generally consistent with any model where the putative timing process needs some variable time to start and stop. Suppose, for example, that the timing of some interval or stimulus duration is affected by two sources of variance: One is the duration-independent "switch" or, more generally, "start–stop" variance, and the other source of variance is related to the duration of the stimulus or event timed ("clock variance" or, more neutrally, "timing variance"). At long durations, timing variance dominates, and the (putatively small and constant) start–stop variance makes only a small contribution to the total variance of performance on the task. On the other hand, at very short durations, start–stop variance makes a large relative contribution, so inflates relative measures of variance such as CV, because a nearly constant variance source is being divided by smaller and smaller means as the duration timed decreases. According to this argument, the increase of CV seen at very short durations does not represent a serious violation of theoretical scalar timing, as this property refers to the putative timing variance, rather than the start–stop variance. Data from both animals and humans showing a marked increase in a Weber-fraction-like measure at a duration shorter than 100 ms were provided by Fetterman and Killeen (1992). Such data could be described in terms of the generalized Weber law, according to which time perception reflects not only Weber's law but also the influence of an additive (constant) variance factor, the relative importance of which grows as the durations to be estimated decrease (see also Clarke, Ivry, Grinband, Roberts, & Shimizu, 1997, for similar data from rats).

Another possible factor contributing to duration-independent variance is the motor variance due to performing the response itself. This source of variance might well act in the same way as the start–stop variance and consequently will increase relative measures of variance when short durations are timed. Motor variance in animal timing is likely to depend on the topography of the operant. For example, when Lejeune et al. (2000) studied temporal differentiation of the time of residence on a small platform in OF1 albino mice, wood mice, and gerbils, the CV was highest at the shortest time requirements (less than 3 s), whilst showing scalar timing at longer time requirements.

Some experiments with animals have also found increases of CV at very long durations (usually several hundreds of seconds). Lejeune and Wearden (1991) reported such effects from a range of animal species on FI, and Zeiler and Powell (1994) found similar results with PI schedules with pigeons. Lejeune and Wearden (1991) simulated such effects using a model like that employed by Wearden (1985) to account for power law deviations from linear timing on temporal differentiation schedules—namely, the addition of "random" responses (i.e., those not controlled by elapsed time in the interval on FI) to a timed response process. It seems, therefore, that deviations from empirical scalar timing at long durations, like those at very short durations, may be explicable theoretically in terms of the operation of additional factors (start–stop or motor variance in the case of very short durations, random responses in the case of very long durations) over and above a basic scalar-consistent system. However, scalar properties are not invariably violated with long durations. As described above, Crystal (2001a) found good superimposition at values of 3 and 7 hours on a modified FI procedure, where a whole "meal" served as the reinforcer, although time values close to 24 hours produced more sensitive timing, possibly because of the operation of circadian timing.

As suggested in the previous section, and as discussed by Platt (1979), effects of absolute durations timed may interact with the procedure used, with violations of scalar timing, or even loss of any evidence of temporal control at all, occurring for temporal differentiation schedules with long time requirements. Schedules where variations in the pattern of responding hardly affect periodicity of reinforcer delivery (such as FI), or schedules involving stimulus discrimination, may be much less affected by changes in the absolute values of the durations timed, keeping in mind, however, that durations studied using the latter procedures are usually much shorter than those explored using FI.

Although increase in the coefficient of variation at very short and very long durations can be modelled accurately by adding nonscalar sources of variance ("start–stop" variance for very short durations and random responses at very long durations), it should be acknowledged that some recently developed theories allow such deviations to occur, unlike unmodified SET. For example, the adaptive timer model of Dragoi, Staddon, Palmer, and Buhusi (2003) actually predicts increases in coefficient of variation at very long durations, and Staddon and Higa's (1999) MTS model can encompass a range of possibilities, including inflation of coefficient of variation at short durations and increasing coefficient of variation at long durations. However, these different effects apparently depend on different parameter values, so it is unclear whether the same setting of the MTS model can predict inflated relative variability at both short and long durations.

Violation 3: "Task difficulty" effects

Ferrara, Lejeune, and Wearden (1997) drew attention to the fact that manipulations that appeared to change the difficulty of the time discrimination that had to be performed produced data that systematically violated superimposition. In their experimental study, humans received a temporal generalization task where the nonstandard durations were spaced around the standard in "easy" steps (150 ms) or "difficult" steps (75 ms). They reported that temporal generalization gradients were steeper in the latter case than in the former, so people confused a particular comparison less frequently with the standard when it came from a "difficult" comparison set than when it came from an "easy" one. Ferrara et al.'s explanation of some of the effects obtained (an explanation later elaborated by Wearden, 2004) was that changes in "task difficulty" essentially affected the decision process: Decisions about the potential equality of the two durations used a stricter criterion in the "difficult" case than in the "easy" one.

A similar effect was obtained on another duration discrimination task: bisection. Here the *L:S* ratio can be considered as an index of task difficulty: The lower the ratio, the more difficult the task. For example, discriminating 8 from 4 s (*L:S* ratio of 2:1) is more difficult than a 12:4 s discrimination (*L:S* ratio of 4:1), and equal *L:S* ratios are considered as reflecting tasks of equal difficulty. Using pigeons and *L:S* ratios between 3:2 and 6:1, Meck (1985) found lower Weber fractions as task difficulty increased. However, Crystal (2002) observed more sensitive timing under conditions of equal *L:S* ratios when the absolute durations of these intervals were small (3–12 s instead of 25–100 s or 50–200 s pairs).

Task difficulty effects in animals, as in humans, may depend on changes in decision processes rather than changes in the representations of the durations judged, as Ferrara et al. (1997) proposed. For example, more sensitive timing associated with a more conservative threshold for initiating responding was described in pigeons on the PI task when Cheng (1992) manipulated behaviour by punishing responses emitted during the first half of the 12.5 s FI. Compared with data from a control condition, response rate versus time functions obtained under the punishment contingency had a narrower spread, and thus showed apparently greater timing sensitivity, even when the timed interval was kept constant, thus violating superimposition. Consistent with the idea that such manipulations might affect response thresholds rather than time representations per se, Cheng (1992) explained the effects in terms of a change of the threshold used for starting and stopping responding.

Animals trained on duration differentiation tasks yielded similar results. Sensitivity increased (i.e., CVs were smaller) as the number of durations that pigeons were required to reproduce increased (Zeiler & Hoyert, 1989) and as a limited hold was added to a response duration differentiation task in rats (Kuch, 1974).

Modelling task difficulty effects with changing response thresholds has simulated data from studies with animals accurately whenever it has been used, although not all studies have either noted that "task difficulty" effects were present, or attempted to model them. It seems, therefore, that task difficulty effects may not actually theoretically violate the scalar properties of time, in the sense that changing task difficulty does not actually change the mean and variability of underlying time representations (which remain scalar), even when superimposition is violated between conditions. However, it should be noted that treating task difficulty effects with threshold changes can only be accomplished by models that have such thresholds. Such reconciliation is easy for SET, but may be much harder for "behavioural" models like BeT and LeT, where the behaviour observed arises from behavioural sequences or associations between operant behaviour and underlying "states", so is more "automatic" than in SET.

Violation 4: Unusual variance and sensitivity patterns

If we consider the "sensitivity" of timing (as shown by Weber fractions and similar measures), deviations from scalar timing (constant sensitivity) are often "normal" with respect to the duration timed. Examples are decreases in timing sensitivity at very short or very long durations, discussed above. In contrast to this, some studies suggest "unusual" patterns of sensitivity, and among these are "step-like" changes in timing variability, or they suggest that certain absolute time values are timed more precisely than others.

An example of the former steplike sensitivity has been described in data from pigeons exposed to a categorical scaling task, where ranges of durations (short, medium, long) were associated with different response keys (Fetterman & Killeen, 1995). Standard deviations increased more slowly within than between duration ranges. As a consequence, Weber fractions within a duration range were not scalar, whereas scalarity was observed between ranges.

Other bodies of results suggest that certain absolute durations are more precisely timed than others. A summary figure in Crystal (2003) shows that, in animals, sensitivity highs have so far been observed in both the short and the circadian duration ranges. The number of time values from which such effects have been obtained remain, however, rather restricted, and whole ranges of potential time intervals (e.g., between 50 and 10,000 s) remain to be explored.

Within the circadian range of time values, the existence of high-sensitivity windows had already been suggested by data from "restricted feeding" experiments in animals, where the daily meal can be accessed at a 24-hour interval. Records of general or conditioned activity showed that animals anticipated the occurrence of the daily meal (Crystal, 2001a; Terman, Gibbon, Fairhurst, & Waring, 1984). Most importantly, anticipation of food is less precise when intermeal durations are located outside the circadian range (Crystal, 2001a, 2003). High-sensitivity windows found in the circadian range suggest that similar windows at shorter durations may also depend on oscillators with periods in the seconds and milliseconds range.

The existence of high-sensitivity windows at certain absolute time values contradicts SET, according to which a constant sensitivity across duration range should be found, or at least that departures from linearity should be located at random. It is also incompatible with models like BeT, LeT, MTS, and packet theory. However, in agreement with the connectionist model of Church and Broadbent (1990), Crystal et al. (1997) suggested that replacing the single pacemaker and the distribution of remembered durations in memory, proposed by SET, with a set of oscillators with different periods may be able to predict nonlinearities in various measures of timing. Oscillator-based models usually imply that certain absolute durations will be more precisely timed than others, although the exact mechanism by which this is done differs from model to model, and the area has received little systematic exploration (except in Wearden & Doherty, 1995, who discuss the relation between the oscillator periods used in a model and the timing of specific durations).

Overall, therefore, evidence that some absolute times are timed with greater precision than others remains rather scarce, although some convincing cases have been demonstrated. One problem that studies of this topic encounter is that overall mean measures of timed behaviour can be well described as linear functions of real time, and the deviations from linearity obtained are usually very small. Even in articles where significant nonlinearities were found, simple proportional timing and Weber-like variance properties are often an excellent first approximation to the data, an approximation so good that in most areas of psychology it would be regarded as a completely sufficient description. For example, in one case, assuming that sensitivity was constant as durations were varied over a 5-fold range of time values, a function accounting for 95% of data variance was produced (Crystal, 2001b). However, some results, like the effects of near-24-hour periods found in Crystal (2001a), are so large as to be unlikely to be artifactual, and, in some other cases, different experiments find deviations from simple linearity occurring at the same values (or many of the same values: see Crystal, 1999), suggesting that effects are real.

CONCLUSIONS

The review above reports extensive evidence for both properties of scalar timing (mean accuracy and scalar variance) in results obtained from a wide range of procedures, although the data presented here are considered to be representative of effects obtained, rather than an exhaustive catalogue of all results reported. They come from studies using FI schedules and variants, temporal differentiation schedules, stimulus timing procedures, aversive conditioning techniques, and Pavlovian methods. Evidence for empirical scalar timing thus comes from a range of interval values and from different animal species, although studies with rats and pigeons are overwhelmingly common.

The ubiquity of empirical scalar timing revealed in our review has been implicitly acknowledged by theorists whose models in their simple form might predict either deviations from linear timing (e.g., Staddon & Higa, 1999), or nonscalar variance (Killeen & Fetterman, 1988), as the theorists have taken considerable care to construct their models so that scalar properties emerge from them, in order to accommodate the main trends in data. Other accounts have scalar timing built in to the structure of their model, usually by using some factors that produce a multiplicative transform of an underlying timing process (like oscillator period) from one trial to another (Church & Broadbent, 1990; Matell & Meck, 2000, 2004).

As mentioned above, our main intention in this article was to identify when deviations from the empirical scalar timing occur and to try and assess their significance. We identified a number of situations in which scalar properties are often violated: temporal differentiation schedules, very short and long durations, task difficulty effects, and the suggestion of unusual variance and sensitivity patterns, possibly linked in some cases to circadian effects. Other deviations have been described in isolated reports (e.g., Kirkpatrick-Steger, Miller, Betti, & Wasserman, 1996), but are not discussed here because their implications for models of timing are at present hard to evaluate.

Of the deviations noted, effects of short durations and task difficulty effects seem easy to reconcile with the principles of SET (although they may pose problems for some other models), using explanations in terms of clock mechanisms like switch effects, or decision threshold changes. Data from temporal differentiation schedules that deviate from the mean accuracy property have been accurately modelled using combinations of timing and nontiming processes, although whether this explanation is completely satisfactory remains unclear. In response timing, species-specific factors seem related to different sensitivities to time, as reported in Lejeune and Wearden (1991), and some of the several factors shown to modulate time discrimination, such as the physical nature of the stimuli presented in the timing task (auditory vs. visual, for example), stimulus content (e.g., empty or filled), or intensity, can probably be related to the phylogeny and ontogeny of the species studied (Lejeune, 1990). A recent study by Buhusi, Perera, and Meck (2005), for example, attributed the different effects of gaps in stimuli to be timed to differences in the salience of the stimuli present during the gaps. They tested this idea by using rat strains with different degrees of visual discriminative abilities and found support for this hypothesis.

It is clear from our review that conformity to empirical scalar timing is overwhelmingly common, but this perhaps makes deviations from scalar timing more theoretically interesting than conformities. In this respect, data suggesting that some absolute time values are timed with greater precision than others may be particularly challenging. The "nonlinearities" reported in some studies using albino rats, a species far from being exotic, are perhaps the deviations from scalar timing that are the most interesting and, from the point of view of SET and most other models that assume some underlying linear timing process, most threatening, theoretically. As mentioned above, deviations are often very small, only a few percent of the absolute time values, but their importance lies not in their size but in their implication that timing is performed by some mechanism quite different from the clock, memory, decision process specified by SET, or by the processes specified in models like BeT, LeT, MTS, and packet theory. Radical replacement of what are basically linear-timing models by oscillator-based accounts on the basis of results in a handful of experiments (and data that are often not completely replicable in detail) seems a step that is at present unwarranted, particularly as oscillator-based models that have been proposed (e.g., Church & Broadbent, 1990; Matell & Meck, 2000, 2004) have difficulties accounting for performance on many timing tasks (see Droit-Volet & Wearden, 2003; Wearden & Doherty, 1995).

Original manuscript received 16 March 2005
Accepted revision received 17 October 2005

REFERENCES

Allan, L., & Gerhardt, K. (2001). Temporal bisection with trial referents. *Perception and Psychophysics, 63*, 524–540.

Bevins, R. A., & Ayres, J. J. B. (1995). One-trial context fear conditioning as a function of the interstimulus interval. *Animal Learning and Behavior, 23*, 400–410.

Bizo, L., & White, G. (1997). Timing with controlled reinforcer density: Implications for models of timing. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 44–55.

Blough, D. S. (1963). Interresponse times as a function of continuous variables: A new method and some data. *Journal of the Experimental Analysis of Behavior, 9*, 581–591.

Brown, B. L., Hemmes, N. H., & Cabeza De Vaca, S. (1997). Timing of the CS–US interval by pigeons in trace and delay autoshaping. *Quarterly Journal of Experimental Psychology, 50B*, 40–53.

Brunner, D., Kacelnik, A., & Gibbon, J. (1992). Optimal foraging and timing processes in the starling *Sturnus vulgaris*: Effects of intercapture interval. *Animal Behaviour, 44*, 597–613.

Buhusi, C. V., Perera, D., & Meck, W. H. (2005). Memory for the timing of visual and auditory signals in albino and pigmented rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 18–30.

Carr, J. A. R., & Wilkie, D. M. (1998). Characterization of the strategies used by rats in an interval time–place learning task. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 151–162.

Catania, A. C. (1970). Reinforcement schedules and psychophysical judgment: A study of some temporal properties of behavior. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 1–42). New York: Appleton-Century-Crofts.

Catania, A. C., & Reynolds, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 11*, 327–383.

Cerutti, D. T., & Staddon, J. E. R. (2004). Immediacy versus anticipated delay in the time-left experiment: A test of the cognitive hypothesis. *Experirmental Psychology: Animal Behavior Processes, 30*, 45–57.

Cheng, K. (1992). The form of the timing distribution in pigeons under penalties for responding early. *Animal Learning and Behavior, 20*, 112–120.

Cheng, K., & Miceli, P. (1996). Modelling timing performance on the peak procedure. *Behavioural Processes, 37*, 137–156.

Cheng, K., & Roberts, W. A. (1991). Three psychophysical principles of timing in pigeons. *Learning and Motivation, 22*, 112–128.

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

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

Church, R. M., Getty, D. J., & Lerner, N. D. (1976). Duration discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes, 2*, 303–312.

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

Church, R. M., Lacourse, D. M., & Crystal, J. D. (1998). Temporal search as a function of the variability of interfood intervals. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 291–315.

Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 135–155.

Clarke, S., Ivry, R., Grinband, J., Roberts, S., & Shimizu, N. (1997). Exploring the domain of the cerebellar timing system. In M. A. Pastor & J. Artieda (Eds.), *Time, internal clocks and movement* (pp. 257–280). Amsterdam: Elsevier Science.

Cooper, L. (1991). Temporal factors in classical conditioning. *Learning and Motivation, 22*, 129–152.

Cotton, M. M., & Wood, W. P. (1982). Rapidly developed scalar timing in unsignalled shuttlebox avoidance learning. *Quarterly Journal of Experimental Psychology, 34B*, 209–219.

Crystal, J. D. (1999). Systematic nonlinearities in the perception of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 3–17.

Crystal, J. D. (2001a). Circadian time perception. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 68–78.

Crystal, J. D. (2001b). Nonlinear time perception. *Behavioural Processes, 55*, 35–49.

Crystal, J. D. (2002). Timing inter-reward intervals. *Learning and Motivation, 33*, 311–326.

Crystal, J. D. (2003). Nonlinearities in sensitivity to time: Implications for oscillator-based representations of interval and circadian clocks. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 61–75). Boca Raton, FL: CRC Press.

Crystal, J. D., Church, R. M., & Broadbent, H. A. (1997). Systematic nonlinearities in the memory representation of time. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 267–282.

Crystal, J. D., & Miller, B. J. (2002). Simultaneous temporal and spatial processing. *Animal Learning and Behavior, 30*, 53–65.

DeCasper, A. J., & Zeiler, M. D. (1974). Time limits for completing fixed ratios: IV. Components of the ratio. *Journal of the Experimental Analysis of Behavior, 27*, 235–244.

DeCasper, A. J., & Zeiler, M. D. (1977). Time limits for completing fixed ratios: III. Stimulus variables. *Journal of the Experimental Analysis of Behavior, 22*, 285–300.

Deliège, M. (1975). *Le comportement de régulation temporelle spontané en programme à intervalle fixe*. Unpublished doctoral thesis, University of Liège, Belgium.

Dews, P. B. (1965). The effect of multiple S^{Δ} periods on responding on a fixed-interval schedule: III. Effects of changes in pattern of interruption, parameters and stimuli. *Journal of the Experimental Analysis of Behavior, 8*, 427–435.

Dews, P. B. (1970). The theory of fixed-interval responding. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 43–62). New York: Appleton Century Crofts.

Dews, P. B. (1978). Studies on responding under fixed-interval schedules of reinforcement: II. The scalloped pattern of the cumulative record. *Journal of the Experimental Analysis of Behavior, 29*, 67–75.

Dragoi, V., Staddon, J. E. R., Palmer, R. G., & Buhusi, C. V. (2003). Interval timing as an emergent learning property. *Psychological Review, 110*, 126–144.

Drew, M. R., Zupan, B., Cooke, A., Couvillon, P. A., & Balsam, P. D. (2005). Temporal control of conditioned responding in goldfish. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 31–39.

Dreyfus, L. R., Fetterman, J. G., Smith, L. D., & Stubbs, D. A. (1988). Discrimination of temporal relations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 349–367.

Droit-Volet, S., & Wearden, J. (2003). Les modèles d'horloge interne en psychologie du temps. *L'Année Psychologique, 104*, 617–654.

Ellison, J. G. D. (1964). Differential salivary conditioning to traces. *Journal of Comparative and Physiological Psychology, 57*, 373–380.

Fairhurst, S., & Gibbon, J. (1983). Timing the second response in two-response avoidance. *Journal of the Experimental Analysis of Behavior, 39*, 199–211.

Fanselow, M. A. (1986). Associative versus topographical accounts of the immediate shock-freezing deficit in rats: Implications for the response selection rules governing species-specific defense reactions. *Learning and Motivation, 17*, 16–39.

Ferrara, A., Lejeune, H., & Wearden, J. H. (1997). Changing sensitivity to duration in human scalar timing: An experiment, a review and some possible explanations. *Quarterly Journal of Experimental Psychology, 50B*, 217–237.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Fetterman, J. G., Dreyfus, L. R., & Stubbs, D. A. (1989). Discrimination of duration ratios. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 253–263.

Fetterman, J. G., & Killeen, P. (1992). Time discrimination in *Columba livia* and *Homo sapiens. Journal of Experimental Psychology: Animal Behavior Processes, 18*, 80–94.

Fetterman, J. G., & Killeen, P. (1995). Categorical scaling of time: Implications for clock-counter models. *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 43–63.

Frey, P. W., & Ross, L. E. (1968). Classical conditioning of the rabbit eyelid response as a function of interstimulus interval. *Journal of Comparative and Physiological Psychology, 65*, 246–250.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review, 107*, 289–344.

Gibbon, J. (1971). Scalar timing and and semi-Markov chains in free-operant avoidance. *Journal of Mathematical Psychology, 8*, 109–138.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279–324.

Gibbon, J. (1979). Timing the stimulus and the response in aversive control. In M. D. Zeiler & P. Harzem (Eds.), *Advances in analysis of behaviour: Vol. 1: Reinforcement and the organization of behaviour* (pp. 299–340). Chichester, UK: Wiley.

Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation, 22*, 3–38.

Gibbon, J., & Balsam, P. (1981). Spreading association in time. In C. M. Locurto, H. S. Terrace, & J. Gibbon (Eds.), *Autoshaping and conditioning theory* (pp. 219–253). New York: Academic Press.

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

Gibbon, J., Church, R. M., Fairhurst, S., & Kacelnik, A. (1988). Scalar expectancy theory and choice between delayed rewards. *Psychological Review, 95*, 102–114.

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

Gibbon, J., Fairhurst, S., & Goldberg, B. (1997a). Cooperation, conflict and compromise between circadian and interval clocks in pigeons. In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: psychological and neurobehavioural analyses* (pp. 329–384), Amsterdam: Elsevier Science.

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997b). Towards a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology, 7*, 170–184.

Grace, R. C., & Nevin, J. A. (1999). Timing and choice in concurrent chains. *Behavioural Processes, 45*, 115–127.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127*, 22–44.

Hanson, S. J., & Killeen, P. R. (1981). Measurement and modeling of behavior under fixed-interval schedules of reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes, 7*, 129–139.

Harzem, P. (1969). Temporal discrimination. In R. M. Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning* (pp. 299–319). London: Academic Press.

Hienz, R. D., & Eckerman, D. A. (1974). Latency and frequency of responding under discrete-trial fixed-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 21*, 341–355.

Higa, J. J., & Simm, L. A. (2004). Interval timing in Siamese fighting fish (*Betta splendens*). *Behavioural Processes, 67*, 501–509.

Holland, P. C. (1998). Temporal control in Pavlovian occasion setting. *Behavioural Processes, 44*, 225–236.

Holland, P. C. (2000). Trial and intertrial durations in appetitive conditioning in rats. *Animal Learning and Behavior, 28*, 121–135.

Innis, N. K., & Staddon, J. E. R. (1971). Temporal tracking on cyclic-interval reinforcement schedules. *Journal of the Experimental Analysis of Behavior, 16*, 411–423.

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

Jones, L. A., & Wearden, J. H. (2004). Double standards: Memory loading in temporal reference memory. *Quarterly Journal of Experimental Psychology, 57B*, 55–77.

Jozefowiez, J., Cerutti, D. T., & Staddon, J. E. R. (2005). Timing in choice experiments. *Journal of the Experimental Analysis of Behavior, 31*, 213–225.

Kacelnik, A., & Brunner, D. (2002). Time and foraging: Gibbon's scalar expectancy theory and optimal patch exploitation. *Learning and Motivation, 33*, 177–195.

Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review, 109*, 533–544.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review, 95*, 411–422.

Kirkpatrick, K. (2002). Packet theory of conditioning and timing. *Behavioural Processes, 57*, 89–106.

Kirkpatrick, K., & Church, R. M. (1998). Are separate theories of conditioning and timing necessary? *Behavioural Processes, 44*, 163–182.

Kirkpatrick, K., & Church, R. M. (2000). Independent effects of stimulus and cycle duration in conditioning: The role of timing processes. *Animal Learning and Behavior, 28*, 373–388.

Kirkpatrick, K., & Church, R. M. (2003). Tracking of the expected time to reinforcement in temporal conditioning procedures. *Learning and Behavior, 31*, 3–21.

Kirkpatrick, K., & Church, R. M. (2004). Temporal learning in random control procedures. *Journal of Experimental Psychology: Animal Behavior Processes, 30*, 213–228.

Kirkpatrick-Steger, K., Miller, S. S., Betti, C. A., & Wasserman, E. A. (1996). Cyclic responding by pigeons on the peak timing procedure. *Journal of Experimental Psychology: Animal Behavior Processes, 22*, 447–460.

Kuch, D. O. (1974). Differentiation of press duration with upper and lower limits of reinforced values. *Journal of the Experimental Analysis of Behavior, 22*, 275–283.

LaBarbera, J. D., & Church, R. M. (1974). Magnitude of fear as a function of the expected time to an aversive event. *Animal Learning and Behavior, 2*, 199–202.

Leak, T., & Gibbon, J. (1995). Simultaneous timing of multiple intervals: Implications of the scalar property. *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 3–19.

Lejeune, H. (1978). Sur un paradoxe dans l'estimation du temps chez l'animal. Revue critique. *L'Année Psychologique, 78*, 163–181.

Lejeune, H. (1990). Timing: Differences in continuity or generality beyond differences? In D. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 53–90). Hove, UK: Lawrence Erlbaum Associates Ltd.

Lejeune, H., Cornet, S., Ferreira, M. A., & Wearden, J. H. (1998a). How do Mongolian gerbils (*Meriones unguiculatus*) pass the time? Adjunctive behavior during temporal differentiation in gerbils. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 352–368.

Lejeune, H., Ferrara, A., Simons, F., & Wearden, J. H. (1997). Adjusting to changes in the time of reinforcement: Peak-interval transitions in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 211–231.

Lejeune, H., Ferrara, A., Soffié, M., Bronchart, M., & Wearden, J. H. (1998b). Peak procedure performance in young adult and aged rats: Acquisition and adaptation to a changing temporal criterion. *Quarterly Journal of Experimental Psychology, 51B*, 193–217.

Lejeune, H., Huynen, M. C., & Ferrara, A. (2000). Temporal differentiation in two strains of small rodents: A wood mouse (*Apodemus sylvaticus*) and an albino mouse (*Mus musculus* OF1). *Behavioural Processes, 52*, 155–169.

Lejeune, H., & Jasselette, P. (1986). Accurate DRL performance in the pigeon: Comparison between perching and treadle pressing. *Animal Learning and Behavior, 14*, 205–211.

Lejeune, H., & Richelle, M. (1982). Differential reinforcement of perching duration in the pigeon: A comparison with differential-reinforcement-of-low-rate keypecking. *Behaviour Analysis Letters, 2*, 49–57.

Lejeune, H., Richelle, M., & Wearden, J. (2006). About Skinner and time: Behavior-analytic contributions to research on animal timing. *Journal of the Experimental Analysis of Behavior, 85*, 125–142.

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

Leonard, D. W., & Theios, J. (1967). Effects of CS–US interval shift on classical conditioning of the nictitating membrane of the rabbit. *Journal of Comparative and Physiological Psychology, 63*, 355–358.

Libby, M., & Church, R. M. (1974). Timing of avoidance responses by rats. *Journal of the Experimental Analysis of Behavior, 22*, 513–517.

Lowe, C. F., & Harzem, P. (1977). Species differences in temporal control of behavior. *Journal of the Experimental Analysis of Behavior, 28*, 189–201.

Lowe, C. F., Harzem, P., & Spencer, P. T. (1979). Temporal control of behavior and the Power law. *Journal of the Experimental Analysis of Behavior, 31*, 333–344.

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

Ludvig, E. A., & Staddon, J. E. R. (2004). The conditions for temporal tracking under interval schedules of reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes, 30*, 299–316.

Ludvig, E. A., & Staddon, J. E. R. (2005). The effects of interval duration on temporal tracking and alternation learning. *Journal of the Experimental Analysis of Behavior, 83*, 243–262.

Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological Review*, *104*, 241–265.

Machado, M., & Keen, R. (1999). Learning to time (LET) or scalar expectancy theory (SET)? A critical test of two models of timing. *Psychological Science*, *10*, 285–290.

Machado, A., & Keen, R. (2003). Temporal discrimination in a long operant chamber. *Behavioural Processes, 62*, 157–182.

Machado, A., & Pata, P. (2005). Testing the scalar expectancy theory (SET) and the learning-to-time model (LeT) in a double bisection task. *Learning and Behavior*, *33*, 111–122.

Malott, R. W., & Cumming, W. W. (1964). Schedules of interresponse time reinforcement. *Psychological Record, 14*, 211–252.

Malott, R. W., & Cumming, W. W. (1966). Concurrent schedules of interresponse time reinforcement: Probability of reinforcement and the lower bound of the reinforced interreponse time interval. *Journal of the Experimental Analysis of Behavior, 9*, 317–326.

Mandell, C., & Atak, J. R. (1982). Temporal reproduction in the rat. *Behaviour Analysis Letters, 2*, 141–151.

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

Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behaviour. *Bioessays, 22*, 94–103.

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: Coincidence detection and oscillatory processes. *Cognitive Brain Research, 21*, 139–170.

McCarthy, D., & Davison, M. (1986). On the discriminability of fixed from variable stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes, 12*, 48–58.

McKearney, J. W. (1969). Fixed-interval schedules of electric shock presentation: Extinction and recovery of performance under different shock intensities and fixed-interval durations. *Journal of the Experimental Analysis of Behavior, 12*, 301–313.

Meck, W. H. (1985). Postreinforcement signal processing. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 52–70.

Miki, A., & Santi, A. (2005). The perception of empty and filled time intervals by pigeons. *Quarterly Journal of Experimental Psychology, 58B*, 31–45.

Millenson, J. R. (1966). Probability of response and probability of reinforcement in a response-defined analogue of an interval schedule. *Journal of the Experimental Analysis of Behavior, 9*, 87–94.

Ohyama, T., Gibbon, J., Deich, J. D., & Balsam, P. D. (1999). Temporal control during maintenance and extinction of conditioned keypecking in ring doves. *Animal Learning and Behavior, 27*, 89–98.

Pavlov, I. P. (1960). *Conditioned reflexes*. New York: Dover Publications. (Original work published 1927.)

Platt, J. R. (1979). Temporal differentiation and the psychophysics of time. In M. D. Zeiler & P. Harzem (Eds.), *Advances in analysis of behaviour: Vol. 1. Reinforcement and the organization of behaviour* (pp. 1–30). Chichester, UK: Wiley and Sons.

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

Platt, J. R., Kuch, D. O., & Bitgood, S. C. (1973). Rat's lever-press duration as psychophysical judgment of time. *Journal of the Experimental Analysis of Behavior, 9*, 239–250.

Preston, R. A. (1994). Choice in time-left procedure and in concurrent chains with a time-left terminal link. *Journal of the Experimental Analysis of Behavior, 61*, 349–373.

Richardson, W. K., & Loughead, T. E. (1974). Behavior under large values of differential-reinforcement-of-low-rates schedules. *Journal of the Experimental Analysis of Behavior, 22*, 121–129.

Richelle, M., & Lejeune, H. (Eds.). (1980). *Time in animal behaviour*. London: Pergamon Press.

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

Roberts, W. A., Cheng, K., & Cohen, J. S. (1989). Timing light and tone signals in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 23–35.

Rodriguez-Girones, M. A., & Kacelnik, A. (1999). Behavioural adjustment to modifications in the temporal parameters of the environment. *Behavioural Processes, 45*, 173–191.

Rosas, J. M., & Alonso, G. (1996). Temporal discrimination and forgetting of CS duration in conditioned suppression. *Learning and Motivation, 27*, 43–57.

Saslow, C. A. (1968). Operant control of response latency in monkeys: Evidence for a central explanation. *Journal of the Experimental Analysis of Behavior, 11*, 89–98.

Savastano, H. I., & Miller, R. R. (1998). Time as content in Pavlovian conditioning. *Behavioural Processes, 44*, 147–162.

Schneider, B. A. (1969). A two-state analysis of fixed-interval responding in the pigeon. *Journal of the Experimental Analysis of Behavior, 12*, 677–687.

Sidman, M. (1953). Two temporal parameters in the maintenance of avoidance behavior by the white rat. *Journal of Comparative and Physiological Psychology, 46*, 253–261.

Silva, K. M., & Timberlake, W. (1998). The organization and temporal properties of appetitive behaviors in rats. *Animal Learning and Behavior, 26*, 182–195.

Staddon, J. E. R. (1965). Some properties of spaced responding in pigeons. *Journal of the Experimental Analysis of Behavior, 8*, 19–27.

Staddon, J. E. R., & Higa, J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior, 71*, 215–252.

Stubbs, D. A. (1968). The discrimination of stimulus durations by pigeons. *Journal of the Experimental Analysis of Behavior, 11*, 223–238.

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

Stubbs, D. A. (1979). Temporal discrimination and psychophysics. In M. D. Zeiler & P. Harzem (Eds.) *Advances in analysis of behaviour: Vol. 1. Reinforcement and the organization of behaviour* (pp. 341–369). New York: Wiley.

Stubbs, D. A. (1980). Temporal discrimination in a free-operant procedure. *Journal of the Experimental Analysis of Behavior, 33*, 167–185.

Stubbs, D. A., Dreyfus, L. R., Fetterman, J. G., Boynton, D. M., Locklin, N., & Smith, L. D. (1994). Duration comparison: Relative stimulus differences, stimulus age, and stimulus predictiveness. *Journal of the Experimental Analysis of Behavior, 62*, 15–32.

Talton, L. E., Higa, J. J., & Staddon, J. E. R. (1999). Interval schedule performance in the goldfish *Carassius auratus. Behavioural Processes, 45*, 193–206.

Terman, M., Gibbon, J., Fairhurst, S., & Waring, A. (1984). Interaction of circadian and interval timing. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception* (pp. 470–487).

Thorpe, C. M., Bates, M. E., & Wilkie, D. M. (2003). Rats have trouble associating all three parts of the time–place–event memory code. *Behavioural Processes*, *63*, 95–100.

Thorpe, C. M., & Wilkie, D. M. (2002). Unequal interval time–place learning. *Behavioural Processes, 58*, 157–166.

Wearden, J. H. (1985). The power law and Weber's law in fixed-interval postreinforcement pausing: A scalar timing model. *Quarterly Journal of Experimental Psychology, 37B*, 191–212.

Wearden, J. H. (1990). Maximizing reinforcement rate on spaced-responding schedules under conditions of temporal uncertainty. *Behavioural Processes, 22*, 47–60.

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

Wearden, J. H. (2004). Decision processes in models of timing. *Acta Neurobiologiae Experimentalis, 64*, 303–317.

Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory: Episodic temporal generalization and bisection in humans. *Quarterly Journal of Experimental Psychology, 54B*, 289–310.

Wearden, J. H., & Doherty, M. F. (1995). Exploring and developing a connectionist model of animal timing: Peak procedure and fixed-interval simulations. *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 99–115.

Weisman, R., Brownlie, L., Olthof, A., Njegovan, M., Sturdy, C., & Mewhort, D. (1999). Timing and classifying brief acoustic stimuli by songbirds and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 139–152.

Whitaker, J. S., Lowe, C. F., & Wearden, J. H. (2003). Multiple-interval timing in rats: Performance on two-valued mixed fixed-interval schedules. *Journal of Experimental Psychology: Animal Behavior Processes, 29*, 277–291.

White, N. E., Kehoe, E. J., Choi, J.-S., & Moore, J. W. (2000). Coefficients of variation in timing of the classically conditioned eyeblink in rabbits. *Psychobiology, 28*, 520–524.

Wilson, M. P., & Keller, F. S. (1953). On the selective reinforcement of spaced responding. *Journal of Comparative and Physiological Psychology, 46*, 190–193.

Zeiler, M. D. (1981). Model of temporal differentiation. In C. M. Bradshaw, E. Szabadi, & F. Lowe (Eds.), *Quantification of steady-state operant behavior* (pp. 205–214). Amsterdam: Elsevier/North-Holland Biomedical Press.

Zeiler, M. D. (1983). Integration in response timing: The functional time requirement. *Animal Learning and Behavior, 11*, 237–246.

Zeiler, M. D. (1985). Pure timing in temporal differentiation. *Journal of the Experimental Analysis of Behavior, 43*, 183–193.

Zeiler, M. D. (1991). Ecological influences on timing. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 13–25.

Zeiler, M. D. (1999). Reversal schedule effects in closed and open economies. *Journal of the Experimental Analysis of Behavior, 71*, 171–186.

Zeiler, M. D., Davis, E. R., & DeCasper, A. J. (1980). Psychophysics of key-peck duration in the pigeon. *Journal of the Experimental Analysis of Behavior, 34*, 23–33.

Zeiler, M. D., & Hoyert, M. S. (1989). Temporal reproduction. *Journal of the Experimental Analysis of Behavior, 52*, 81–95.

Zeiler, M. D., & Powell, D. G. (1994). Temporal control in fixed-interval schedules. *Journal of the Experimental Analysis of Behavior, 61*, 1–10.