BRILL

Timing & Time Perception 1 (*2013*) *159–188*

TIME

brill.com/time

Timescale Invariance in the Pacemaker-Accumulator Family of Timing Models

Patrick Simen [1,*], **Francois Rivest** [2,3], **Elliot A. Ludvig** [4], **Fuat Balci** [5] **and Peter Killeen** [6]

[1] Oberlin College, Department of Neuroscience, 119 Woodland St., Oberlin, OH 44074, USA

[2] Royal Military College of Canada, Department of Mathematics & Computer Science PO Box 17000, Station Forces, Kingston, ON, K7K 7B4, Canada

[3] Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada

[4] Princeton University, Princeton Neuroscience Institute, Green Hall, Washington Rd., Princeton, NJ 08540, USA

[5] Koç University, College of Social Science & Humanities, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey

[6] Arizona State University, Department of Psychology, P.O. Box 871104, Tempe, AZ 85287-1104, USA

Received 14 April 2013; accepted 5 September 2013

Abstract

Pacemaker-accumulator (PA) systems have been the most popular kind of timing model in the half-century since their introduction by Treisman (1963). Many alternative timing models have been designed predicated on different assumptions, though the dominant PA model during this period — Gibbon and Church's Scalar Expectancy Theory (SET) — invokes most of them. As in Treisman, SET's implementation assumes a fixed-rate clock-pulse generator and encodes durations by storing average pulse counts; unlike Treisman's model, SET's decision process invokes Weber's law of magnitude-comparison to account for timescale-invariant temporal precision in animal behavior. This is one way to deal with the 'Poisson timing' issue, in which relative temporal precision increases for longer durations, contrafactually, in a simplified version of Treisman's model. First, we review the fact that this problem does not afflict Treisman's model itself due to a key assumption not shared by SET. Second, we develop a contrasting PA model, an extension of Killeen and Fetterman's Behavioral Theory of Timing that accumulates Poisson pulses up to a *fixed* criterion level, with *pulse rates* adapting to time different intervals. Like Treisman's model, this time-adaptive, opponent Poisson, drift–diffusion model accounts for timescale invariance without first assuming Weber's law. It also makes new predictions about response times and learning speed and connects interval timing to the popular drift–diffusion model of perceptual decision making. With at least three different routes to timescale invariance, the PA model family can provide a more compelling account of timed behavior than may be generally appreciated.

* To whom correspondence should be addressed. E-mail: psimen@oberlin.edu

DOI:10.1163/22134468-00002018

Keywords
Diffusion model, BeT, scale invariance, interval timing, Weber's law

1. Introduction

Perhaps the most intuitive model of an animal's internal clock is a simple pacemaker-accumulator (PA) system: Discrete units of some physical quantity accumulate at a constant rate over the course of an interval. When the total sum reaches a critical level, the animal behaves as if the interval is over. The PA approach seems intuitive because we have lived for centuries with clocks that count oscillations of pendula, or rotations of mainsprings, or reverberations of electrons.

Treisman's (1963) PA model (hereafter denoted *TPA*) is the prototype PA model of timing. It uses a pacemaker, whose pulses are accumulated by a counter and sent to a store to encode durations. Critically, in TPA, the inter-pulse durations within trials are correlated, with shorter-than-average durations in some trials, and longer-than-average durations in others (Postulate 2, Treisman, 1963). In other words, the pace of the pacemaker varies randomly across trials around a fixed average. As shown in Treisman, this property of the TPA accounts for the strict form of 'Weber's law for timing', a temporal analogue of Weber's classic law of perception.

The classic form of this law purports to govern behavior in two-choice tasks requiring subjects to decide which of two non-temporal stimuli has greater intensity (e.g., heavier, brighter, etc.). Although Weber investigated perceptual representations by finding the *just noticeable difference* between very similar stimuli, the law can be restated as holding that accuracy is constant whenever the two comparison stimuli are proportionally strengthened or weakened in intensity. This relationship suggests a level of perceptual imprecision that is intensity-scale-invariant: specifically, the intensity estimates across repeated trials of a task are distributed so that the standard deviation S of the estimates is a constant proportion of the average estimate M. The *coefficient of variation* (CV) of the estimates, S divided by M, is therefore constant.

In Treisman (1963), the CV of human behavioral response times in timing tasks was indeed found to be roughly constant across different durations in temporal production, reproduction, decision and estimation tasks, although a correction factor a was required in a generalized form of Weber's law:

$$S = k \cdot (M + a).$$

For durations ranging from 0.25 to 9 s, Treisman found k in the range 0.05–0.1, a around 0.5, and M accurate but subject to some biases toward shorter or longer estimates, depending on the procedure. One of the main empirical goals of Treisman (1963) was to address the diversity of previous findings for which it was not clear whether even this generalized form of the classic law held. His conclusion

was that it did, and that there was little evidence for any local minimum of the CV at any particular duration, as had been previously hypothesized.

Since Treisman (1963), experiments with non-human animals, for which verbal and cognitive strategies such as counting would likely be minimized, suggested even stronger support for the strict form of the law, $S = kA$ (e.g., Gibbon, 1977, among many subsequent replications, but see also Bizo et al., 2006, for some exceptions). Furthermore, Gibbon and colleagues frequently observed that the entire distribution of response times, when divided by the mean response time, typically superimposes with any other similarly normalized response time distribution from the same experiment, regardless of the duration being timed (Gibbon et al., 1997). This superimposition property is sometimes dubbed *scalar invariance*; for consistency with the general use of the similar phrase *scale invariance* across disciplines, we will use *timescale invariance* to refer to the same property.

1.1. The TPA Route to Timescale Invariance

In the TPA model, a duration T is timed by counting the pulses emitted by a pacemaker. Treisman (1963) does not specify precisely what kind of pulse train is emitted from the pacemaker, other than to state that the inter-pulse durations are highly regular within trials. The TPA pacemaker has a fixed average rate, which is constant *within* trials, but variable *between* trials.

The TPA account of Weber's law for time presented in Treisman (1963; Equation [11]), assumes that the inter-pulse times are essentially identical and the pacemaker is essentially periodic (Postulate 1), but with a different period in each trial (Postulate 2). All inter-pulse durations within any given trial are almost the same. Under this assumption, the standard deviation, across trials, of the sum of n inter-pulse times, all of which equal the same random duration X, equals n times the standard deviation of X. This follows from the variance formula for a constant, n, and a random variable X: $\mathrm{Var}(nX) = n^2\,\mathrm{Var}(X)$. Taking the square root gives $\mathrm{Std}(nX) = n\,\mathrm{Std}(X)$. Note the contrast from the case in which *independent* pulse durations are added: in that case, the variance of the sum is the sum of the variances. If all inter-pulse durations have identical variances $\mathrm{Var}(X)$ across trials, then the variance of the sum of n independent pulses is $n\,\mathrm{Var}(X)$; the n in this case is not squared.

For the nearly-periodic TPA, different durations T are timed by counting different numbers, n, of the fixed-rate pacemaker's pulses, which occur with average inter-pulse duration equal to $\mathrm{Mean}(X)$. This process yields an estimate, T', of T, as follows: $T' = \sum_{i=1}^{n} X_i$. For this estimate, $\mathrm{CV}(T') = \mathrm{Std}(T')/\mathrm{Mean}(T') = n\,\mathrm{Std}(X)/(n\,\mathrm{Mean}(X))$. The n cancels out of the numerator and denominator, and the CV is thus constant for all T. A similar argument underlies Treisman's (1964) explanation of Weber's law in its traditional, non-temporal context.

Later theories of timing, such as Scalar Expectancy Theory (SET; Gibbon, 1977) and its information processing implementation (IPI; Gibbon & Church, 1984; Gib-

bon, Church, & Meck, 1984), emphasized pacemakers with Poisson characteristics, which emit highly irregular pulse trains. Inter-pulse durations in such models are independent and exponentially distributed. For such models with independent inter-pulse durations, the variance of the pulse counts is $n \cdot \mathrm{Var}(X)$, which implies:

$$\begin{aligned}\mathrm{CV}(T') &= \frac{\sqrt{n}\,\mathrm{Std}(X)}{(n\,\mathrm{Mean}(X))} = \frac{\mathrm{Std}(X)}{\sqrt{n}\,\mathrm{Mean}(X)} = \cdots \\ &= \mathrm{Std}(X) \Big/ \left(\sqrt{\mathrm{Mean}(X)} \cdot \sqrt{n \cdot \mathrm{Mean}(X)}\right) \\ &= \left[\mathrm{Std}(X)/\sqrt{\mathrm{Mean}(X)}\right] \cdot 1/\sqrt{T}.\end{aligned}$$

This behavioral pattern (Killeen & Weiss, 1987) is commonly known as *Poisson timing* (Gibbon & Church, 1984).

This decrease of the CV in proportion the square root of time is inconsistent with the data, and requires an additional model component to account for the strict form of Weber's law for timing. In SET, therefore, Poisson PA estimates of new durations are compared to memorized durations in a noisy way governed by Weber's law for two-alternative choice; specifically, the comparison of pulse counts to memory was taken to be performed in terms of the ratio of one count to the other. Incorporating this discrimination pattern into SET's memory comparison extends Weber's law for intensity discrimination into a law of timing. As Gibbon (1992) points out, however, it must also be the case that the noise in the ratio comparison is large relative to the Poisson noise in the accumulator; otherwise, the Poisson timing pattern will emerge.

It may therefore surprise some readers to discover that Treisman's original model works very well in accounting for strict forms of both the classic version of Weber's law for two-choice comparisons of stimulus intensity (Treisman, 1964), and for Weber's law for timing (Treisman, 1963), even when its high-correlation assumption is weakened almost out of existence. As Treisman (1966) demonstrates with computer simulations, correlation values within trials can be allowed to shrink from 1 down to 0.001, and the model's behavior still conforms to Weber's law. We have also conducted our own computer simulations of a Poisson PA timer that, like TPA and SET, uses a simple pulse-count criterion for deciding the end of an interval.[1] Remarkably low correlation levels of 0.001 are sufficient to give a nearly constant CV across different durations. For this model, SET's memory comparison noise is not needed to account for Weber's law for timing; its ratio comparison rule is therefore not the only way to account for timescale invariance.

Thus at least two modifications of the basic PA approach — pacemaker variance across trials in TPA and ratio-based memory comparison in SET — do a good

[1] Matlab code for all simulations in the paper is available on the Web at: http://www.oberlin.edu/faculty/psimen/GoldenAnniversaryCode.html

job of accounting for empirically observed behavioral patterns. We will shortly propose a third modification that works just as well, that provides specific predictions about the shape of response-time distributions, and that connects PA timing models to diffusion models, which arguably constitute the leading class of perceptual decision making models in psychology and neuroscience at present.

1.2. Why Use Any Other Type of Model?

In addition to simplicity and intuitive appeal, linear PA models (i.e., PA models in which the pulse-rate is constant within a trial) clearly have substantial explanatory power. This makes the wide variety of nonlinear and/or non-accumulating alternatives striking (e.g., Ahrens & Sahani, 2011; Almeida & Ledberg, 2010; Grossberg & Schmajuk, 1989; Haß et al., 2008; Karmarkar & Buonomano, 2007; Ludvig, Sutton, & Kehoe, 2008; Machado, 1997; Matell & Meck, 2004; Miall, 1989; Shankar & Howard, 2012; Staddon & Higa, 1996; Wackermann & Ehm, 2006, to name just a few). Some of the motivation for developing alternative models may stem from perceived weaknesses of PA models, especially more recent incarnations such as SET (Gibbon, 1977; Gibbon et al., 1984).

We now consider a few possible objections to PA models, including objections based on behavioral patterns that may seem to remain unexplained, and objections based on a lack of neural evidence for the models' components. In our view, these objections either stem from misconceptions, or are at least as applicable to other models of timing as they are to PA models.

1. Objections based on inadequately explained behavioral data:
   1.1. Timescale invariance: As noted above, SET's IPI accounts for timescale invariance by deemphasizing the noisy pacemaker and adding the error into counting through a ratio-comparison decision rule (Gibbon, 1992). It thereby builds Weber's law into the memory comparison process. Some researchers find this fix unsatisfying (Staddon & Higa, 1999a), since it seems to beg the question of the provenance of the memory-comparison noise (Staddon & Higa, 1996; Wearden & Bray, 2001). Other accounts of Weber's law have famously been given — e.g., Fechner's classic account, or Treisman's (1964), or Link's (1992), among many others — so it seems legitimate to reduce an account of Weber's law for timing to a more general account of Weber's two-choice law. Yet it is striking that SET's account of timescale invariance reduces to its model of decision processes, rather than to its model of the pacemaker, and it is conceivable that this well known flaw of the Poisson timing approach has cast doubt on the entire PA family. As we show below, however, multiple types of Poisson process models can account directly for timescale invariance without any appeal to memory comparison noise or ratio comparisons.

1.2. Bisection at the geometric mean: Another salient regularity in timed behavior is that animals often bisect two different durations near their geometric mean. This has been taken to favor a logarithmic internal representation of time (Allan & Gibbon, 1991; Church & Deluty, 1977; Jozefowiez, Staddon, & Cerutti, 2009). Similarly, Weber's law itself was long taken as evidence for Fechner's logarithmic representation of subjective stimulus intensity (see, for example, the history of the Weber–Fechner law recounted in Link, 1992). PA models with a fixed clock-speed cannot easily account for bisection at the geometric mean, as the slower growth of error that they entail would place the *point of subjective equality* (PSE) above the geometric mean. Logarithmic coding is not, however, necessary to account for bisection data — the generalized form of Weber's law by itself requires that PSEs range between the harmonic and arithmetic mean, depending on the amount of constant error and the separation of the stimuli in these tasks (Killeen, Fetterman, & Bizo, 1997). Furthermore, Balci et al. (2011) showed that the PSE may depend on the task participant's level of temporal precision, with the PSE of more precise timers nearer to the arithmetic mean.

2. Objections based on the notion of neural implausibility:

2.1. Lack of evidence for brain localization: A remarkable feature of brain organization is that it is quite difficult to lesion the brain in such a way as to selectively knock out an internal clock (e.g., Gooch et al., 2011). Evidence linking timing to the basal ganglia, cerebellum, supplementary motor cortex, parietal cortex, hippocampus, and more recently, primary visual cortex and even the retina, suggests that timing capabilities are distributed throughout the brain. Indeed, it seems distributed in such a diffuse, nonmodular way as to conform to Lashley's early view of the brain as a functionally undifferentiated mass of neural tissue, once past the relatively well-defined input and output circuits (Uttal, 2008). How, then, could a simple counting model be distributed across the brain? As demonstrated in Simen et al. (2011b), the simple machinery of the new model we propose could be instantiated throughout the brain.

2.2. Lack of plausibility of ramping activity over very long periods: Seung (1996) and Wang (2002), among others, raised the issue of how neural firing rates could plausibly ramp up linearly over the course of even a one-second interval given the rapid dynamics of the neural membrane, which fluctuates on a time scale better measured in milliseconds. Gibbon et al. (1997) raised a similar challenge in discussing how neural implementations could employ precise, linear ramps over multiple seconds and even minutes. Simen et al. (2011a) discussed how this precision-tuning problem might be ameliorated, but it is clear that many researchers find

long-duration ramping of neural activity difficult to accept. The necessary experiments to test for such slow ramping, however, are much more difficult to run than ones to test for rapid ramping.

We have recently proposed a timing model that is a new variation on the PA theme (Rivest & Bengio, 2011; Simen et al., 2011b). The primary difference between this *drift–diffusion model* (DDM) of timing and the PA models just discussed is that it adapts the rate of its pacemaker to time different intervals, leaving its pulse count threshold fixed. In this respect, it shares the same defining feature as a third classic PA model, the Behavioral Theory of Timing (BeT) model of Killeen & Fetterman (1988). As we show, despite its apparently very different mathematical definition, the new model we have proposed contains BeT as a special case.

In our view, the adaptive-pacemaker approach offers several possible advantages over the adaptive-criterion approach of TPA and SET. One advantage is the straightforward way in which the adaptive-pacemaker approach has been mapped onto more explicitly neural substrates (Simen et al., 2011b). Another advantage is the way in which long durations are encoded, namely, with low pacemaker rates, as opposed to high pulse counts in TPA and SET. High pulse counts must be compressed in some way to fit within whatever limited range is available in the brain. Such compression is not implausible, but it seems to imply that the accumulation component is (for better or worse) nonlinear, and specific compression schemes have not been formalized for TPA or SET. A third advantage is the existence of an explicit learning rule that can be applied to govern the pacemaker rate (Rivest & Bengio, 2011; Simen et al., 2011b); such a rule is of course critical for the adaptive-pacemaker approach, but some form of pacemaker control might also be seen as necessary for the TPA/SET approach, because motivational factors in TPA (Treisman, 1963, Postulate 2, p. 19) and pharmacological factors in SET (Meck, 1996) are hypothesized to have important effects on the pacemaker's rate.

Whether neural systems can easily implement the digital counters and registers used by the creators of TPA and SET to describe implementations of their models is an open question. Clearly, the researchers who have proposed many recent, non-PA models in the neuroscience literature appear to feel that there is something implausible about the PA approach, as classically embodied in TPA or SET. This view may not be justified, given neural evidence adduced for both models: e.g., Treisman (1984), Treisman et al. (1990), and Treisman et al. (1992) for TPA; and, e.g., Gibbon et al. (1997), Meck (1996), and Meck (2006) for SET. We argue, however, that the neural implausibility charge carries even less weight against adaptive-pacemaker PA models. We therefore review the principles of these adaptive-pacemaker models, to make the case that the PA family currently provides the best account of timing behavior and its possible neural basis.

The particular form of the DDM developed below gives a simple analytical explanation of timescale invariance. It also accounts for one-trial learning of dura-

tions, and it parsimoniously reuses the DDM — a leading model of non-temporal, perceptual decision making (Smith & Ratcliff, 2004) — for the purpose of timing. Perhaps most importantly, it makes new predictions about the shape of response time distributions that are well supported. We now review the origins of this model historically and mathematically in terms of an opponent Poisson accumulation process, and we describe how it might overcome the major objections to counting models listed above.

1.3. A Brief History of the Drift–Diffusion Model

Diffusion models originated in the work of Einstein on the atomic/molecular basis of Brownian motion (Einstein, 1905), which is typified by the restless jiggling of a pollen grain in water as seen through a microscope (see Gardiner, 2004, for an excellent historical review). The distribution of first-passage times for diffusion models was later derived by physicists (Schrödinger, 1915; Smoluchowsky, 1915). These are the times when a particle under Brownian motion first exceeds some predefined distance from its starting point. These results were later generalized in the Fokker–Planck formalism widely used today, in which partial differential equations describe the time-evolution of probability distributions. Mathematicians (notably including Norbert Wiener, after whom Brownian motions are frequently termed *Wiener processes*) established their analytical bases as random processes, proving important theorems about their properties that underpin our explanation of timescale invariance.

Diffusion processes were given their most elegant representations as stochastic differential equations with the development of generalized rules for stochastic integration (in particular, the Ito calculus). Stochastic integration is the traditional operation of integration computed with respect to time and, simultaneously, with respect to a variable representing idealized Brownian motion (Gardiner, 2004). Because it bears directly on our narrative about accumulator models, we highlight the fact that the continuous mathematical formalism of diffusion — in which time and space are considered infinitely divisible, and particles never jump from one location to another without covering the intervening space — was motivated by what is essentially a *discontinuous* counting or accumulation process: i.e., the cumulative displacement of a particle by a sequence of collisions. The first Brownian motion observed under a microscope was the result of a large number of discrete, countable collisions of small water molecules with a larger grain of pollen, but it presumably looked as if it was being continually pummeled. If a movie were made of an idealized particle undergoing truly *continuous* Brownian motion (motion like that shown in Fig. 1), then slowing the movie down and examining only a tiny, magnified fragment of it would almost certainly reveal a trajectory with the same visual properties as the overall random walk (a form of self-similarity made famous by widely viewed computer animations that 'zoom in' on the Mandelbrot set). As-

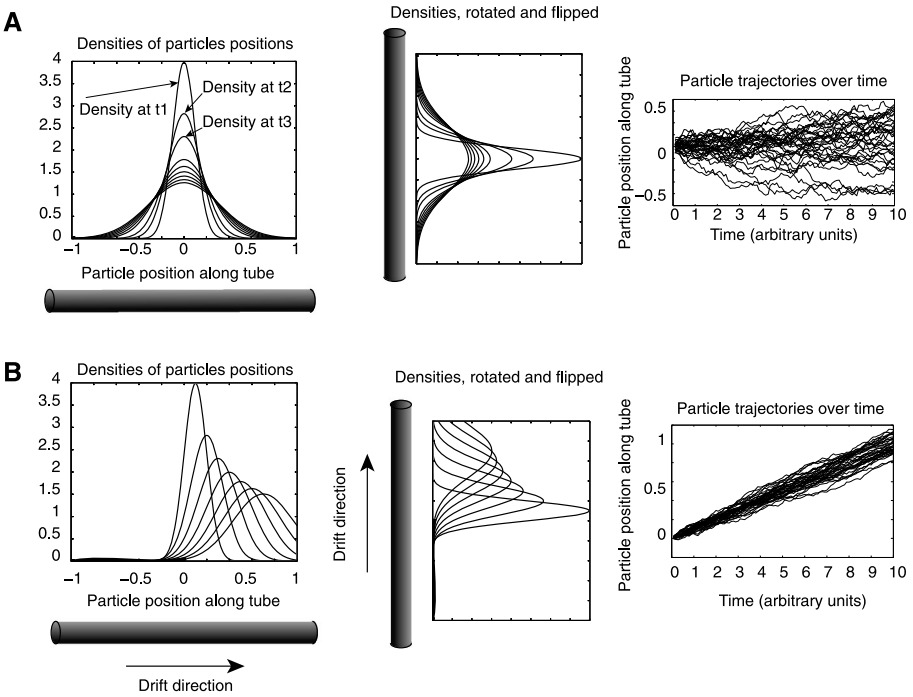

Figure 1. (A) Diffusion without any drift: Particle ensembles densities plotted against position at ten successive time steps (left) and particle trajectories plotted against time for ten time units (right). (B) Diffusion with drift, forcing these ensembles toward the right (top) end of the tube.

suming continuity, however physically implausible, has nonetheless turned out to be a very useful idealization for analyzing high-rate accumulation processes.

We complete this brief historical account of diffusion by describing its importation into psychology. In parallel to the developments noted above, statistician/economist Abraham Wald developed theories of random processes for their application to decision problems (Wald, 1947). His work culminated in results such as the optimality, under certain conditions, of the sequential probability ratio test (SPRT). In the SPRT, the log-likelihood ratio of two competing hypotheses is computed in real time as samples of evidence are obtained, until the ratio exceeds a level that indicates a decision, positive or negative, can be made with some level of expected accuracy. We will refer to such a criterion level as a *decision threshold*. Wald showed that, in a very general sense, no other test was better than the SPRT for a generic class of two-choice problems: no faster test would be as accurate, and no more accurate test would be as fast (Wald & Wolfowitz, 1948). When the evidence samples are sufficiently small and occur sufficiently frequently, the SPRT approximates a drift–diffusion process: continuity becomes an acceptable assumption; diffusion occurs because subsequent samples of the evidence frequently do not equal

each other; and drift toward the correct threshold happens, on average, whenever samples of evidence favor the correct response. This approximation has practical importance because it is a source of great mathematical convenience. In the 'one-choice' timing context, in particular, assuming continuity yields a simple, two-parameter expression for the distribution of response times and its moments — the inverse Gaussian distribution — as discussed below. Thus, drift–diffusion processes have a longstanding, historical connection to decision theory.

Stone (1960) imported the discrete-sample SPRT into the psychological literature as a model of human reaction times in two-choice decisions. He noted one of its most psychologically plausible features, which is that it cannot help but produce a speed-accuracy tradeoff: A higher decision threshold leads to slower performance but higher accuracy (because the odds increase that the correct threshold will be crossed first, by the law of large numbers), whereas a lower threshold leads to greater speed and more mistakes. Few other phenomena are more robustly observed in human and non-human decision making than a tradeoff between speed and accuracy.

A strong prediction of the SPRT model and its diffusion approximation is that correct and error reaction times in two-choice tasks should be equal when the starting point is equidistant from both thresholds. Because these predictions are usually violated in behavioral data, the SPRT model was soon rejected by influential scientists (e.g., Laming, 1968). Subsequently, Ratcliff showed that generalizations of the DDM incorporating across-trial variability in the starting point of the diffusion process and in the rate of the drift allows either faster or slower errors than corrects (Ratcliff, 1985; Ratcliff & Rouder, 1998).

It is important to note that Ratcliff assumed no direct connection between his generalized DDM and the SPRT or its assumptions. Human participants are modeled as embodying drift–diffusion processes without some of the assumptions key to SPRTs (e.g., that participants have good estimates of stimulus likelihoods). In practical terms, Ratcliff and others showed that his generalized model had the capacity to fit an enormous range of human behavioral data, from memory retrieval (Ratcliff, 1978) to letter identification (Ratcliff & Rouder, 2000) to simple perceptual discrimination (Ratcliff & Smith, 2010), with the typical finding that fitted drift rates decrease as task difficulty increases. Diffusion models are now leading models of perceptual decision processes, applying to neural data as well as behavior (Gold & Shadlen, 2007; Ratcliff, Cherian, & Segraves, 2003); most of the alternatives to diffusion models are arguably subtle rather than radical variations of them (e.g., the Linear Ballistic Accumulator of Brown and Heathcote, 2008, which is basically the Ratcliff model, sans within-trial diffusion; or the Ornstein–Uhlenbeck model of Usher & McClelland, 2001, in which a feedback term either dampens or enhances drift away from the starting point). Nonetheless, DDMs have only recently been adapted as formal models of timing (Rivest & Bengio, 2011; Simen et al., 2011b). These adaptations were originally inspired, however, by classic PA models such as

Treisman's and Gibbon and Church's. In fact, as shown below, the type of time-adaptive DDM we propose closely approximates an extended form of Killeen and Fetterman's (1988) BeT model.

2. Formal Definition and Properties of the Drift–Diffusion Model (DDM)

We now define the DDM mathematically, describe a physical realization of it, and review the standard method for simulating it on a computer. (The description of the simulation technique is easier to understand than the equation of motion that it simulates.) Table 1 gives a glossary of DDM parameters.

The DDM is most compactly defined by a single, stochastic differential equation, where dt refers to time, and dW denotes, loosely speaking, Gaussian white noise added at every step of the simulation (a formalization of the Brownian molecular impacts):

$$dx = A \cdot dt + c \cdot dW. \tag{1}$$

Think of x as the position of a drop of food-color in a horizontal tube. Without loss of generality, we can define horizontal positions x along the tube in reference to its midpoint (position 0). Consider the right edge of the tube to be at position z, and the left edge to be at position $-z$, so that the total tube length is $2z$. The drift rate A determines how rapidly the particle will move to the right on average (i.e., the current in the tube). The noise amplitude c specifies the level of noise or random Brownian motion; it determines the expected distance traveled by the particle per unit of time under random perturbations alone. When the water is hot, for example, the momentum that the water molecules confer on the particle is likely to be larger than for cold water. Hotter water thus causes the food-color to diffuse more rapidly (c is larger).

We denote the horizontal position where the food-color is injected into the tube as x_0. In the context of using the model to time intervals, we are primarily interested in the response time (RT): that is, how long it takes for a pigment particle to reach the right edge of the tube for the first time. In the more typical application of the model — to decisions between two choices — there is an additional objective

Table 1.
Parameters of the DDM

Parameter symbol	Parameter name
A	Drift rate
c	Noise coefficient
z	Threshold
T_0	Non-decision latency

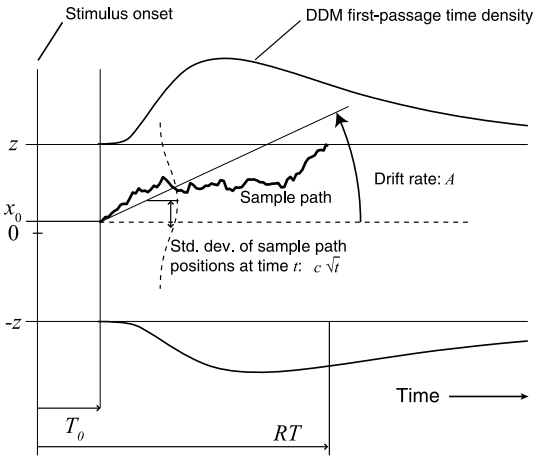

Figure 2. The drift–diffusion model of two-choice decision making. The drift, A, determines how rapidly particles rise upward on average; the noise level, c, determines how much random perturbation of the particle occurs per unit time; the starting point is x_0; the upper boundary z serves as a criterion for one type of response in two-alternative decision making, while the lower boundary, $-z$, serves as a criterion for the other. In the case of timing, z is assumed to be much larger relative to the noise than in perceptual decision making, and $-z$ is assumed to be at negative infinity, though in practice, as long as it is more than $0.1z$ below the starting point, lower-boundary crossings are extremely rare and can be ignored. T_0 represents a non-decision latency for sensory encoding and motor response that is added to the decision time of the diffusion model to yield a total response time.

of determining the probabilities of one edge being reached before the other, which yield the model's choice probabilities. In the timing context, fitting real data requires the drift A to be so strong relative to diffusion c and tube length $2z$ that the left edge is almost never reached before the right edge. It is this assumption that allows the use of the inverse Gaussian distribution, greatly simplifying the results as compared to the more common two-choice application of the model. Figure 2 shows an example of such a trajectory over time as it is usually represented, with time on the horizontal axis, and with the tube now oriented vertically (e.g., the former left edge is rotated down, and the former right edge to the top).

The Euler–Maruyama method (Higham, 2001) is widely used for computer simulation of Equation (1) as a discrete-time random walk. It involves iteratively evaluating Equation (2) to obtain a new particle position, x_{new}, based on the previous one, x_{old}. It employs the standard normal random variable $N(0, 1)$ and time steps of size Δt:

$$x_{\text{new}} = x_{\text{old}} + A \cdot \Delta t + c \cdot \sqrt{\Delta t} \cdot N(0, 1). \tag{2}$$

The square root of the time step in the noise term must be included, because the mean square deviation of the particle position distribution increases linearly with time — a standard property of diffusive processes. If not included, then the random

walk simulation, Equation (2), systematically deviates from Equation (1) (viz., distributions of particle positions will have the wrong variance). The noise parameter c scales the standard deviation of pure Brownian motion whose variance is 1 after 1 unit of time (in general, the variance equals the amount of time elapsed — this and other key properties were proved by Wiener; see, e.g., Gardiner, 2004).

This model can be adapted to time responses in several ways. We could fix the drift A and time different durations T by setting the threshold z equal to AT, for example. This approach would be analogous to SET's IPI. Alternatively, we could fix the threshold and adapt the drift rate to time different intervals (as in BeT; Killeen & Fetterman, 1988). Finally, we could do both simultaneously. We will follow BeT and develop the fixed-threshold/adaptive-drift approach, but before describing why, it will help to review some properties of the DDM that are critical for its account of timescale invariance. First, see Fig. 3 for why the lower threshold at $-z$

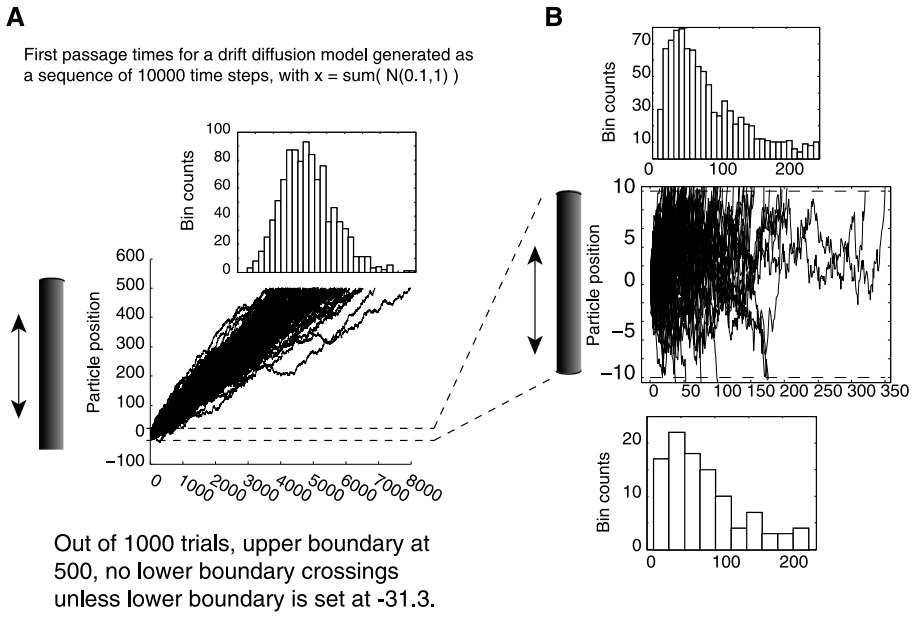

Figure 3. (A) A simulation of a 1000 trials of a drift–diffusion process, in which a normal random variable with mean 0.1 and std 1 is accumulated over 10 000 time steps, until a first-passage across the threshold value of 500 occurs. No trajectories ever descended below −31.3, so whether or not a lower boundary (reflecting or absorbing) is included at, say, −30 or less, the first-passage-time histogram would look almost identical. (B) The exact same set of trajectories, magnified, starting at position 0, with an upper threshold at 10 and a lower threshold at −10. The histogram of upper-threshold crossing times is shown above; the histogram of lower-threshold crossing times is shown below. Note that there are far fewer lower-threshold crossings, but that the histogram has the same shape as the one above.

can be ignored; when z and A are high relative to c, the lower boundary is almost never touched (for parameters that yield plausible levels of timing precision, the probability of crossing the lower threshold first is on the order of 10^{-22}).

2.1. *Timescale Invariance (or Lack Thereof) for Varieties of the DDM*

We first note that the DDM predicts non-normal distributions of response times. With a single absorbing boundary — in psychological terms, a single decision threshold — the DDM predicts an inverse Gaussian distribution (Luce, 1986). The inverse Gaussian (also known as the *Wald distribution*) is typically defined in terms of two parameters: $\mu = z/A$ and $\eta \doteq z^2/c^2$. Its density, defining the likelihood of first-passage at time t, is the following:

$$p(t, \mu, \eta) = \sqrt{\frac{\eta}{2\pi t^3}} \cdot \exp\left(\frac{-\eta(t-\mu)^2}{2\mu^2 t}\right).$$

The inverse Gaussian converges on a normal distribution as noise c decreases in relation to drift rate A (see Fig. 4 for an example of this convergence). With two thresholds, one above and one below the starting point, the response time distribution has no closed-form expression and is expensive to compute, requiring the evaluation of many terms in an infinite sum of sines (Feller, 1968; Tuerlinckx, 2004). As previously noted, however, with a large, positive ratio of drift to noise, the lower threshold can be ignored, as particle trajectories have almost 0 probability of crossing the lower threshold before the upper threshold, and the distribution quickly approximates the inverse Gaussian. Assuming only a single threshold is also more parsimonious because there is only one response to be made in simple timing tasks.

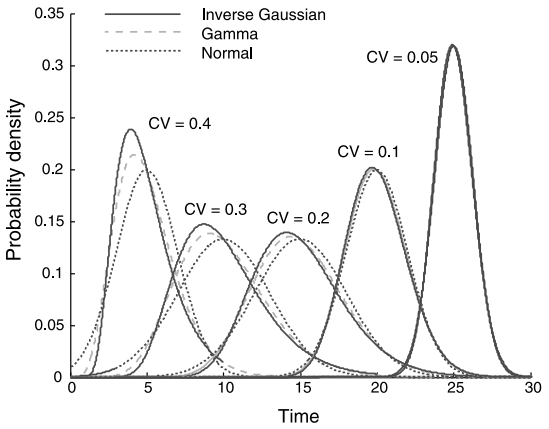

Figure 4. Comparison of inverse Gaussian (solid), gamma (dashed), and normal (dotted) probability densities with equivalent CVs. Densities with larger CVs (on the left) show greater deviation from normality and larger positive skewness. These densities were computed with greater noise for shorter durations.

The DDM with constant noise (i.e., with c fixed and thus independent of A and z) is not timescale invariant. In this case, the expected first-passage time T is simply the ratio of distance-to-threshold divided by the drift rate: $T = z/A$. The standard deviation of the first-passage times is $\sigma = \frac{c\sqrt{z}}{\sqrt{A^3}}$, as derived in Simen et al. (2011b) from the more standard expression in terms of μ and η: $\sigma^2 = \mu^3/\eta$ (e.g., Luce, 1986). The CV is therefore the following:

$$\mathrm{CV}_{\mathrm{DDM}} = \frac{c\sqrt{z}}{\sqrt{A^3}} \cdot \frac{A}{z} = \frac{c}{\sqrt{Az}}. \tag{3}$$

If we hold threshold z fixed and vary drift A to time intervals, so that $A = z/T$, we can eliminate A to obtain

$$\mathrm{CV}_{\mathrm{DDM}} = \frac{c\sqrt{T}}{z}. \tag{4}$$

In this case, with the threshold fixed, the CV grows as the square root of T; timescale invariance is violated, to the detriment of timing long intervals.

If we hold drift A fixed and vary threshold z to time intervals, then

$$\mathrm{CV}_{\mathrm{DDM}} = \frac{c}{A^2 T} = \frac{c}{A\sqrt{T}}. \tag{5}$$

In that case, with variable drift or pacemaker speed, the CV *decreases* as the square root of T; again, timescale invariance is violated, to the benefit of timing long intervals. This version is similar to classic Poisson timing (Gibbon, 1992), in which a fixed-rate Poisson process times different durations via different pulse counts, producing $\mathrm{CV} = 1/\sqrt{\lambda T}$, where λ is the Poisson rate parameter.

The final possibility is to fix the threshold, adapt drift to time intervals, and allow the noise term to vary with the drift (and therefore with T). Reasons at both the neural and the behavioral level favoring this approach are noted below along with a formalization of this scheme in terms of two learning rules. Because the model is a special case of a DDM formed by two competing Poisson processes (as discussed in the next section), we refer to it as an *opponent Poisson DDM*; and because it is time-adaptive, using the learning rules defined in Rivest and Bengio (2011) and Simen et al. (2011b) to adapt the drift (discussed next), we refer to the model as a *time-adaptive*, *opponent Poisson DDM* (TOPDDM).

2.2. Timescale Invariance in the TOPDDM

First, we review how the TOPDDM yields timescale invariance, appealing to a principle of optimal behavior (see Table 2 for a glossary of TOPDDM parameter symbols). We then review how the model emerges from a set of simple assumptions about neural activity. Mathematically, these assumptions extend the assumptions underlying BeT, which can be considered a special case of the model we propose. BeT's physical interpretation differs from that of our model, but its mathematical account of timescale invariance carries over mostly unchanged.

Table 2.
Parameters of the TOPDDM

Symbol	Parameter name
A	Drift
z	Threshold
γ	Inhibition-to-excitation ratio
m	Noise-to-drift proportionality constant
λ	Pulse rate of the pacemaker
μ	Inverse Gaussian mean parameter
η	Inverse Gaussian shape parameter

Equation (3) implies that for any DDM, maximizing A and z (or minimizing c) minimizes the CV. Increasing precision in this way would seem to be advantageous for survival, up to biologically achievable speeds and criteria. We assume that thresholds are always kept near their maximum value because this allows the drift rate to decrease in order to time long intervals and to asymptotically approach its minimum as durations grow very long. The alternative strategy, fixing drift and adapting threshold, would lead to a linear approach of the threshold toward its maximum as durations grow, causing all longer durations to be perceived equal to the maximum time-able duration. There are limits on the durations we can precisely time in a stopwatch-like manner, but such limits are manifest as failing precision, not as a compression of all long counts toward some smaller, maximum count. Evidence is seen in the large CVs that approach 1 for 500-s intervals in Fig. 3 of Gibbon et al. (1997), or that increase regularly at long times (e.g., Bizo et al., 2006). The drift-adaptation strategy of variable A with z fixed near maximum permits more graceful degradation with long durations. Furthermore, we are unaware of evidence for an abrupt onset of under-estimates — stuck clocks — as durations increase. Finally, it could be argued that the physical limits on threshold height (e.g., maximum firing rates) are more constraining than physical lower bounds on ramp-up speed on a multiple-second timescale.

For the DDM, Equation (3) and the assumption of a constant threshold together imply that a constant CV is achieved when drift rate is proportional to the square of noise rate. But why would we make this assumption? This cornerstone of the DDM's explanation of timescale invariance follows from four assumptions about the kind of neural activity we take the DDM to represent:

1. The neurons used for timing emit action potentials (spikes) with exponentially distributed inter-spike times with rate λ;

2. Spike-times are uncorrelated between different pacemaker units in a population; pacemaker units are those in populations that fire at a constant rate during an interval and send spikes to the accumulator;
3. A neural integrator/accumulator sums this pacemaker's spikes;
4. Excitation and inhibition are balanced, such that any increase in excitatory inputs to a neural accumulator is accompanied by a proportional increase in inhibitory inputs.

Assumption 1 is an old one in psychology and neuroscience, and though its interpretation is different, such Poisson counting is also central to BeT. In BeT, the Poisson assumption applies to behavioral-state transitions (e.g., from grooming to exploring) rather than action potentials. The response-time distribution for BeT is a gamma distribution of first-passage times. BeT, like the DDM we propose, also fixes the threshold and adapts the rate of its 'spikes' (behavioral-state transitions) to time different intervals. As a Poisson process model, the mean and variance of the 'spike' count at a given time T are both equal to the rate parameter, λ, times T. To maintain constancy of CV, the rate parameter must be inversely proportional to the duration being timed: $\lambda = z/T$. For the gamma distribution, the shape parameter z and scale parameter $\theta = 1/\lambda$, yield a mean of $z\theta = T$, and a variance of $z\theta^2$. This implies a CV as follows:

$$\mathrm{CV}_{\mathrm{BeT}} = \frac{\sqrt{z}}{\lambda} \cdot \frac{1}{T} = \frac{\sqrt{z}}{z} = \frac{1}{\sqrt{z}}. \tag{6}$$

As long as z is held constant, the CV is constant. In general, for a gamma-distributed variable $X \sim \Gamma(z, \theta)$, any multiple qX is distributed as $\Gamma(z, q\theta)$. Dividing X by $T = z\theta$ yields $X/T \sim \Gamma(z, 1/z)$, which is invariant as T changes — i.e., it is timescale invariant. This is why $\theta = 1/\lambda$ is called the *scale* parameter. The shape parameter z gets its name because as it grows large, the resulting distribution changes shape — it takes on an almost normal shape with skew approaching 0; as z decreases, the distribution becomes increasingly positively skewed (see Fig. 4).

BeT's transitions between behavioral states cannot readily be re-interpreted as neural processes. In order to achieve a CV of 0.1 — good performance for non-human animals, typical performance for humans — z must equal 100. For a duration of 100 s, this implies that the clock speed is 1 pulse per second — far too low to be interpreted as an inter-spike time even for a single neuron, let alone a population.

The TOPDDM, in contrast, approximates the running sum of an excitatory Poisson process and a concurrent inhibitory process. The rate of the latter is some proportion γ of the excitatory process rate, and both process rates are adapted to time different intervals (when $\gamma = 0$, the model is equivalent to BeT). The distribution of net spike counts for such a model is a Skellam distribution, which

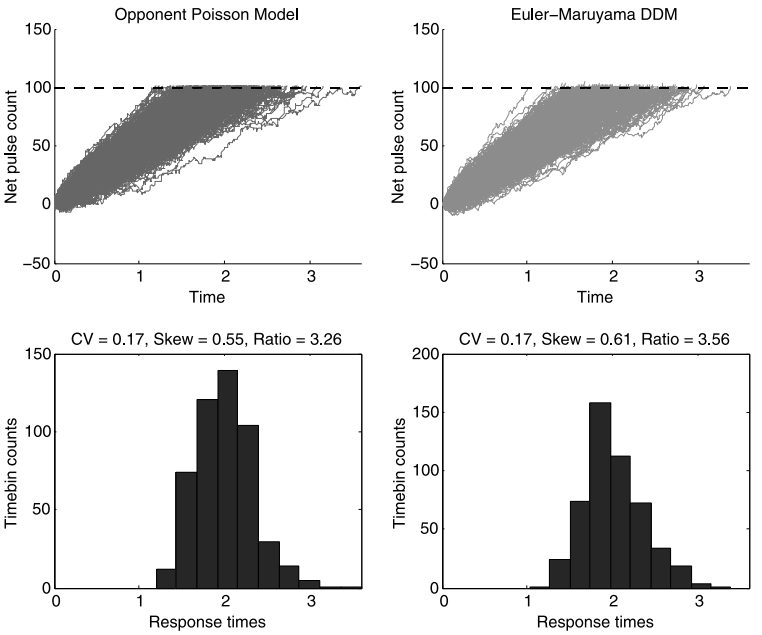

Figure 5. Comparison of trajectories and response time distributions for 500 simulated trials of an opponent Poisson process (left panels) and the Euler–Maruyama simulation of the corresponding DDM parameterization (right panels). CVs of the RT distributions (shown in titles of bottom panels) are identical out to two decimal places; skewness is similar; and the ratio of skewness-to-CV ('Ratio'), predicted to be 3 on average for the DDM, is close to 3 in both cases.

quickly approximates a normal distribution after a few spikes, and thereby justifies a drift–diffusion approximation, which assumes a normal distribution of noise added at each 'time step' (as in Equation (2)). Figure 5 illustrates how closely the DDM approximation matches its corresponding opponent Poisson process. The expected value of this sum is simply the difference between the expected values of each; these equal λt (excitatory) and $\gamma\lambda t$ (inhibitory). The variance of this sum is the sum of the variances. Each variance, given the Poisson assumption, equals the number of expected spikes at time t. Thus the variance of the spike count is $\lambda T + \gamma\lambda T$. In the description of Equation (2), we noted that the variance of the particle position distribution increases linearly with time T; specifically, this variance is c^2T. Therefore, c must approximate $(1+\gamma)\lambda$ in the following drift–diffusion approximation of the spike-count process:

$$\mathrm{d}x = (1-\gamma)\lambda \cdot \mathrm{d}t + \sqrt{(1+\gamma)\lambda} \cdot \mathrm{d}W. \tag{7}$$

By defining $m = \sqrt{\frac{(1+\gamma)}{1-\gamma}}$, and by defining $A = (1-\gamma)\lambda$, we obtain

$$\mathrm{d}x = A \cdot \mathrm{d}t + m\sqrt{A} \cdot \mathrm{d}W. \tag{8}$$

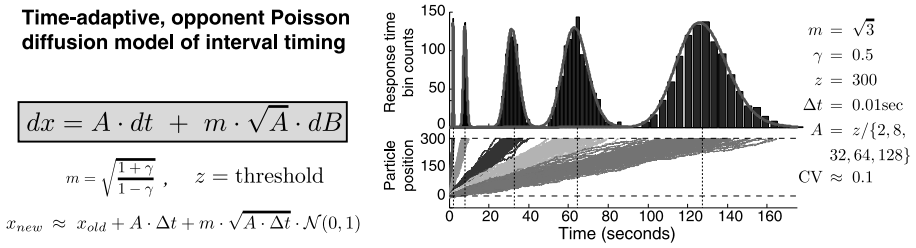

Figure 6. Simulations of the TopDDM for five different interval durations (2, 8, 32, 64, and 128 s). Histogram bin widths increase as the durations increase. The CVs of these distributions are all approximately 0.1.

Thus, the simple assumptions of an opponent Poisson process outlined above yield the right functional form for an account of timescale invariance. Figure 6 shows timescale-invariant simulations of the TOPDDM (with a CV near 0.1 in all conditions).

The CV of the TOPDDM is:

$$\text{CV}_{\text{TOPDDM}} = \frac{m}{\sqrt{z}}. \tag{9}$$

The assumption of proportional inhibition entails a free parameter m, ranging between 1 (for $\gamma = 0$) and infinity (for $\gamma = 1$). This free parameter allows CVs to take on arbitrarily large values even for large spike rates. Setting z to a nominal maximum level minimizes CV for all interval durations, and also produces a constant CV. As with the gamma distribution of BeT, fixing z further predicts timescale invariance of the complete distribution. If X is inverse Gaussian-distributed (iG) with parameters μ and η, $X \sim \text{iG}(\mu, \eta)$, then $qX \sim \text{iG}(q\mu, q\eta) = \text{iG}(qz/A, qz^2/c^2)$ (see Chhikara & Folks, 1989), which in the opponent Poisson case is $\text{iG}(qz/A, qz^2/(m^2A))$. With $q = 1/T = A/z$, this yields $qX \sim \text{iG}(1, z/m^2)$. This parameterization is independent of T and is thus timescale invariant.

It would seem best, if possible, for the animal to reduce the proportion of inhibition γ toward 0, at least for the purposes of timing intervals, as that minimizes the CV. It would drive the DDM toward BeT. But it would also require an implausibly large value for the criterion, or clock rates much too slow to be considered neural processes. In addition, reducing balancing inhibition to 0 could yield a host of unpleasant consequences for the animal: too great a reduction of the inhibitory neurotransmitter GABA in the brain, produces seizures as neurons begin to synchronize their firing in a cascade (Telfeian & Connors, 1998). Thus, γ must be nonzero.

2.3. Skewness Predictions of Four Varieties of PA Model

Skewness statistics are usually ignored in the discussion of empirical timing data — indeed, disregarding skewness has been explicitly recommended (Gibbon & Church, 1990). Yet some positive skewness is typically observed in timing data, and much larger positive skewness is usually observed in the decision making RT distributions to which the DDM is usually applied (see Fig. 3). This greater skewness may be due to slower rates of information gain (A), and in some cases to brief lapses of attention (Killeen, 2013), but is simply explained as the result of much lower threshold-to-noise and threshold-to-drift ratios in the two-choice context (Fig. 3B) relative to the timing context (Fig. 3A). Giving a coherent account of RT skewness is therefore important for a model like ours that seeks to unify timing and decision making in terms of a common process. In general, skewness predictions may serve to discriminate between different timing models that mimic each other in other respects.

The TPA model makes no commitment to particular inter-pulse duration distributions, and so is not constrained with respect to what it predicts about response time distributions. This makes TPA very flexible, but possibly unfalsifiable so far as skewness observations are concerned. SET is typically taken to predict Gaussian response time distributions due to the memory comparison noise that swamps noise in its accumulator. Still, that accumulator noise would tend to lend a slight amount of skewness to the response time distributions. Again, though, the lack of a specific prediction about skewness makes falsification difficult.

For BeT, the skewness of the gamma distribution is much more specific: it is twice its CV (see Simen et al., 2011b, for a derivation). For the single-threshold DDM with a high inhibitory spike rate, the skewness of the inverse Gaussian is three times its CV. Of course, as the inhibitory spike-rate proportion of the opponent Poisson model approaches zero, and the TOPDDM approaches BeT, the predicted skewness must approach twice the CV. This fact demonstrates that the continuous DDM approximation of the discrete accumulation process begins to yield a small amount of systematic error as the inhibition proportion γ approaches 0. Unpublished simulations that we have conducted show that the DDM is nevertheless a good approximation of the spike counting process even in such cases. A close match between the simulated RT distributions when $\gamma = 0$, despite a 50% increase in predicted skewness of the DDM (3) over BeT (2) for a given CV, suggests that skewness statistics are likely to be noisy; large samples are needed to estimate skewness accurately. This also suggests that skewness statistics based on real data are likely to be very sensitive to contaminants, such as responses made when attention is diverted away from the timing task. Nevertheless, both BeT and the TOPDDM predict that skewness should be reliably greater than 0, and should be proportional to CV across all durations.

Despite the apparent sensitivity of the skewness statistic to noise, we have observed in analyses of existing rat data provided by Church and colleagues (Church, Lacourse, & Crystal, 1998) that the ratio of skewness to CV was significantly greater than 0, between 2 and 3, usually closer to 3 than to 2 (as predicted; Simen et al., 2011b). Skewness-to-CV ratio, then, may serve to discriminate timing models from each other if CVs are not too small. In this respect, some non-human animal data are currently better accounted for by the TOPDDM than by its competitors (Simen et al., 2011b).

2.4. Adaptation Rules

The time adaptation of the time-adaptive DDM (TDDM) family of models, including the TOPDDM, arises through simple learning rules that correct the timing error incurred after exposure to a new duration, within trial-to-trial sampling error (Rivest & Bengio, 2011; Simen et al., 2011b). Some of the authors have examined versions of TOPDDM that make no reference to the opponent Poisson assumption, and have therefore used the abbreviation TDDM in other publications (Luzardo, Ludvig, & Rivest, 2013).

If the accumulation fails to reach threshold by the end of an interval, the following late-timer correction rule increases the rate of accumulation:

$$\Delta A = A \cdot \frac{z - x_{\text{end}}}{x_{\text{end}}}. \tag{10}$$

The learning rule does this without depending on accurate knowledge of the duration itself, but only on its current level at the end of the interval (x_{end}) and the drift A and threshold z.

To show that this rule works, consider the duration t_n predicted on trial n after one application of the late-timer learning rule on trial $n-1$, in which the actual, new duration is T:

$$t_n = \frac{z}{A_n} = \frac{z}{A_{n-1} + A_{n-1} \cdot (\frac{z - x_{\text{end}}}{x_{\text{end}}})} = \frac{x_{\text{end}}}{A_{n-1}} = T. \tag{11}$$

If the accumulation hits threshold early, an early-timer correction rule is needed to decrease the rate of accumulation. One possibility for such a learning rule would be if the diffusion process were to rise to some level above threshold z, then a rule complementary to Equation (10) could tune the drift for more accurate performance. This option, though simple to effect, conflicts however with the assumption that organisms employ a near-maximal threshold.

Thus, we consider a learning rule that can correct the timing error without direct reference to the actual duration T, and without the diffusion process exceeding z. Equation (12) gives a differential equation for reducing A in real time in order to learn the new, longer duration:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -\frac{A^2}{z} \quad \text{with initial condition } A(t_{\text{early}}) = \frac{z}{t_{\text{early}}}. \tag{12}$$

Equation (12) should be integrated from the moment the diffusion process reaches z, i.e., t_{early}, to the end of the duration T on trial $n-1$.

This early-timer rule learns the accumulation rate that leads to T in a single exposure, as we can see by moving the A^2 to the left-hand side of the equation and integrating both sides with respect to t:

$$\left[\int_{A(t_{\text{early}})}^{A(T)} -A^{-2}\cdot \mathrm{d}A = \int_{t_{\text{early}}}^{T} \frac{1}{z}\mathrm{d}t\right] \Rightarrow \frac{z}{A(T)} = T. \tag{13}$$

Although these rules allow learning in one shot, a learning rate less than 1 can also be implemented by rescaling the correction term (see Rivest & Bengio, 2011; Simen et al., 2011b). The advantage of damping the process in such a way would be to reduce the effect of sampling error on rate adjustments.

The TOPDDM can be summarized as consisting of Equations (8), (10), and (12).

3. Recent Successes and Remaining Challenges for the TDDMs

Since the publication of Rivest and Bengio (2011) and Simen et al. (2011b), new capabilities have been modeled with time-adaptive DDMs.

3.1. Empirical Learning Rates

In addition to the core timing phenomena of timescale invariance and Weber's Law, the PA family of models has accounted for a number of other behavioral observations regarding the speed with which humans and other animals learn to time (those applying learning rules Equations (10) and (12)). For example, Simen et al. (2011b) showed that human participants learn to encode durations in a single trial without covert verbal counting or tapping.

In addition, Rivest and Bengio (2011) showed that slight variations on this learning rule cause the model to learn either the arithmetic mean of the observed intervals or the harmonic mean (i.e., the overall rate of the events being timed). Estimating this event rate is of particular importance in modeling the speed of learning in conditioning where a number of theories are predicated on that rate with respect to each stimulus (Balsam, Drew, & Gallistel, 2010; Gallistel & Gibbon, 2000). Moreover, Rivest and Bengio (2011) showed that for any fixed learning rate $0 < \alpha < 1$, the drift rate is actually an exponentially weighted moving average of the observed event rate (or of the time intervals, depending on the exact formulation of the learning rule).

More recently, the TDDM family of models was shown to account for animal behavior in situations in which intervals are continuously changing using cyclic schedules (Luzardo et al., 2013). In this work, not only was it shown that TDMMs perform as well as, and sometimes even better than, the multiple time scales (MTS) model (Staddon, Chelaru, & Higa, 2002; Staddon & Higa, 1999b), but it was

also shown how the addition of a single constant parameter in the threshold provides an account for the lack of time-scale invariance present in dynamic-schedule data sets. This addition does not preclude time-scale invariance because the basic model is a special case of the more general model (when that extra constant is set to 0).

3.2. The Peak Interval Task

There are several features of timing data that existing DDM-based accounts have not yet addressed. One is the dynamics within trials of the TOPDDM in commonly used tasks such as the peak-interval (PI) task (Catania, 1970; Roberts, 1981). Extending the model to account for this type of data is no more difficult for the TOPDDM than it is for SET, for example, as we briefly demonstrate.

The DDM and related models are used in the perceptual decision making and psychophysics literatures to model single responses to discrete stimuli. These models are not typically applied to the onset and offset of periods of high-rate responding (Church, Meck, & Gibbon, 1994). The necessary modification of the model to account for these data is as simple in this case as it is for SET's IPI. Although there are multiple ways in which to do this, one simple approach is to use a single drift–diffusion process and two different thresholds: a low threshold for starting a Poisson response process, and a higher threshold for stopping it. This model predicts timescale invariance of the start and stop times across trials, as well as a variety of other patterns involving spread of the high-rate response period and correlations between these variables. Its performance is illustrated in Fig. 7, where a 10-s schedule and a 30-s schedule are both performed.

This dual-threshold version of the TOPDDM predicts timescale-invariant inverse Gaussian distributions of the start times and stop times across trials. Frequently, stop-time CVs are smaller than start-time CVs, as measured from the start of the interval (Balci et al., 2009; Gallistel, King, & McDonald, 2004). TOPDDM can account for this phenomenon, because the lower thresholds on the start-time estimates imply a higher CV. When 1000 trials were simulated using the same parameters used to generate Fig. 7, the start time CVs were 0.127 and 0.126 for short and long durations respectively, and stop time CVs were respectively 0.085 and 0.086.

4. General Discussion

Like many models developed during the heyday of early artificial intelligence and the contemporaneous cognitive revolution in psychology, Treisman's (1963) timing model was influenced by computer-design principles, and the experiments he used to test the model were based exclusively on human behavior. Since then, theoretical behaviorists have applied models to the experimental analysis of behavior in non-human animals (Staddon, 2001). This work has established the existence

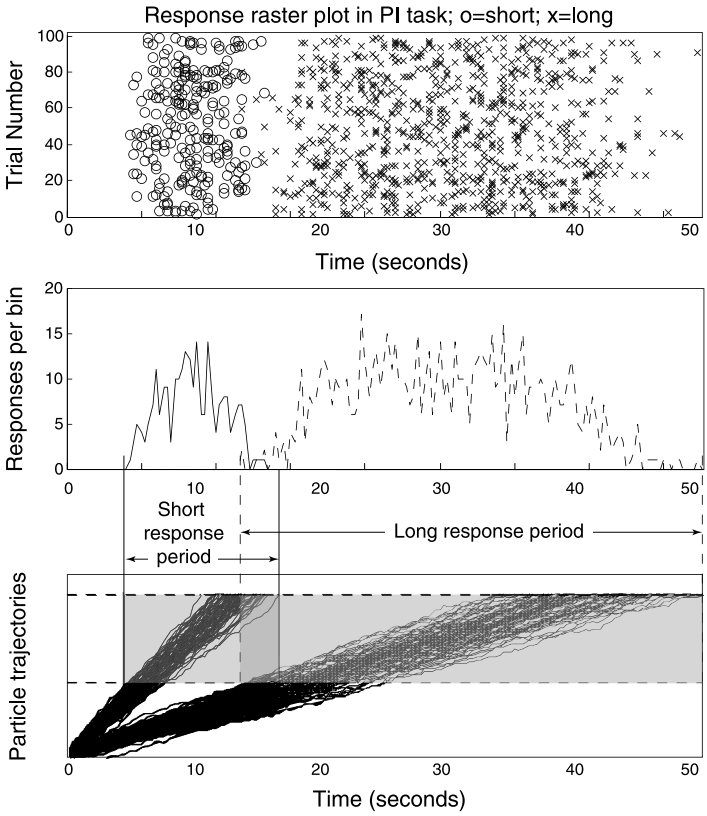

Figure 7. Peak interval performance under two different fixed interval schedules (10 s and 30 s). The top panel shows 10-s responses as o's; 30-s responses as *x*'s. The middle panel shows the binned response rates (solid = 10 s; dashed = 30 s). The bottom panel shows trajectories of the accumulation process. 100 trials were simulated. CVs for start times of periods of high rates responding are 0.121 and 0.116 for short and long durations respectively; CVs for stop times are 0.090 and 0.085 respectively. Thus start times are timescale invariant, stop times are scale invariant, but start times are more variable than stop times.

of cross-species generalities manifesting themselves as laws of behavior, such as timescale invariance. There appears to be little consensus, though, about what these generalities are based on.

We have reviewed how a slight variation on an old idea in psychology and neuroscience — that Poisson spike rates linearly encode important quantities, as in TOPDDM — can give rise to Weber's law for time, and we have noted that it can furthermore account for Weber's law for any paired comparison of two stimulus intensities (see Link, 1992, and a restatement of his argument in the Appendix). We have also reviewed how a different assumption — that the pacemaker runs fast in some trials and slow in others, as in TPA — yields a separate account of both forms

of Weber's law (Treisman, 1963, 1964). Finally, we have reviewed how SET explains the timing version of Weber's law by reducing it to Weber's law for two-alternative choice (Gibbon, 1992).

The TOPDDM in particular can account for Weber's law for time, and at the same time, it makes novel predictions regarding the shape of response time distributions. They should be inverse Gaussians — a prediction that seems in many cases more consistent with empirical data than the typical normal distribution, given the positive skew in so much timing data (Gibbon & Church, 1990; Guilhardi, Yi, & Church, 2007).

The TPA model has great flexibility in terms of the shape of the response time distributions it predicts. Our simulations suggest that as long as its key pacemaker-variability assumption is made, both exponential and Gaussian distributions of inter-pulse durations yield timescale invariance. However, the shape of the response time distribution varies dramatically depending on the inter-pulse duration distribution. It can thus account for the approximately Gaussian shapes predicted by SET, as well as the inverse Gaussian shapes predicted by TOPDDM.

Taken together, these models as a class are likely to give as good a quantitative account of behavioral timing data as any other model on the market.

4.1. Summary

Timing is just one piece of the puzzle confronted by psychologists and neuroscientists. It is arguably such a critical piece, however, that establishing a strong theory of timing represents major scientific progress.

Where are we in terms of reaching consensus on the mechanisms by which we time our behavior? The Golden Age of any science arguably occurs when the feedback between theory creation and empirical testing reaches maximum velocity, fed by attraction toward a stable, useful theory. It is hard to say whether we are at that point, but the conversation between models and data is clearly more extended and more sophisticated than ever before.

The PA family is unlikely to provide a final treatment of the brain's milliseconds to minutes-range timing mechanism(s). It does not address the other time scales in which we live (though we feel the TOPDDM in particular is a strong contender for some of those timescales). Even so, the increasing adoption by timing researchers of mathematical tools such as stochastic differential equations is consistent with the kind of theoretical and empirical acceleration that scientists seek. That kind of acceleration was clearly triggered by Treisman's adoption of computing concepts to model the brain's internal clock in 1963, though in recent decades, the TPA model has not, in our view, received the attention it deserves. Today, the increasing use of diffusion and other concepts in combination with TPA suggests that the Golden Anniversary of Treisman (1963) may very much herald a Golden Age of timing research.

Acknowledgements

This work was supported in part by FP7 Marie Curie (#PIRG08-GA-2010-277015) and TÜBİTAK (#111K402) grants to FB and an RMC start-up fund to FR. We thank two anonymous reviewers for very insightful reviews.

References

Ahrens, M. B., & Sahani, M. (2008). Observers exploit stochastic models of sensory change to help judge the passage of time. *Curr. Biol.*, *21*, 1–7.

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

Almeida, R., & Ledberg, A. (2010). A biologically plausible model of time-scale invariant interval timing. *J. Comput. Neurosci.*, *28*, 155–175.

Balci, F., Freestone, D., Simen, P., deSouza, L., Cohen, J. D., & Holmes, P. (2011). Optimal temporal risk assessment. *Front. Neurosci.*, *5*, 56.

Balci, F., Gallistel, C. R., Allen, B. D., Frank, K. M., Gibson, J. M., & Brunner, D. (2009). Acquisition of peak responding: What is learned? *Behav. Process.*, *80*, 67–75.

Balsam, P. D., Drew, M. R., & Gallistel, C. R. (2010). Time and associative learning. *Comp. Cogn. Behav. Rev.*, *5*, 1–22.

Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behav. Process.*, *71*, 201–210.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognit. Psychol.*, *57*, 153–178.

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

Chhikara, R. S., & Folks, L. (1989). *The inverse Gaussian distribution: theory, methodology, and applications*, Vol. 95. Boca Raton, FL: CRC.

Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *J. Exp. Psychol.*, *3*, 216–228.

Church, R. M., Lacourse, D. M., & Crystal, J. D. (1998). Temporal search as a function of the variability of interfood intervals. *J. Exp. Psychol. Anim. Behav. Process.*, *24*, 291–315.

Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *J. Exp. Psychol. Anim. Behav. Process.*, *20*, 135–155.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, *17*, 549–560.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York, NY: Wiley.

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

Gallistel, C. R., King, A., & McDonald, R. (2004). Sources of variability and systematic error in mouse timing behavior. *J. Exp. Psychol. Anim. Behav. Process.*, *30*, 3–16.

Gardiner, C. W. (2004). *Handbook of stochastic methods* (3rd ed.). New York, NY: Springer-Verlag.

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

Gibbon, J. (1992). Ubiquity of scalar timing with a Poisson clock. *J. Math. Psychol.*, *35*, 283–293.

Gibbon, J., & Church, R. M. (1984). Sources of variance in an information processing theory of timing. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 465–488). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibbon, J., & Church, R. M. (1990). Representation of time. *Cognition*, *37*, 23–54.

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

Gibbon, J., Malapani, C., Dale, C., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Curr. Opin. Neurobiol.*, *7*, 170–184.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Gooch, C. M., Wiener, M., Hamilton, A. C., & Coslett, H. B. (2011). Temporal discrimination of sub- and suprasecond time intervals: A voxel-based lesion mapping analysis. *Front. Integr. Neurosci.*, *5*, 59.

Grossberg, S., & Schmajuk, N. A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Netw.*, *2*, 79–102.

Guilhardi, P., Yi, L., & Church, R. M. (2007). A modular theory of learning and performance. *Psychon. Bull. Rev.*, *14*, 543–559.

Haß, J., Blaschke, S., Rammsayer, T., & Herrmann, J. M. (2008). A neurocomputational model for optimal temporal processing. *J. Comput. Neurosci.*, *25*, 449–464.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *Soc. Ind. Appl. Math. Rev.*, *43*, 525–546.

Jozefowiez, J., Staddon, J. E. R., & Cerutti, D. T. (2009). The behavioral economics of choice and interval timing. *Psychol. Rev.*, *116*, 519–539.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron*, *53*, 427–438.

Killeen, P. R. (2013). Absent without leave: A neuroenergetic theory of mind wandering. *Front. Psychol.*, *4*, 373.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychol. Rev.*, *95*, 274–295.

Killeen, P. R., Fetterman, J. G., & Bizo, L. A. (1997). Time's causes. In C. M. Bradshaw & E. Szabadi (Eds.), *Time and behaviour: psychological and neurobehavioural analyses* (pp. 79–131). Amsterdam: Elsevier Science Publishers.

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychol. Rev.*, *94*, 455–468.

Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York, NY: Wiley.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Luce, R. D. (1986). *Response times: their role in Inferring Elementary Mental Organization*. New York NY: Oxford University Press.

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2008). Stimulus representation and the timing of reward-prediction errors. *Neural Comput.*, *20*, 3034–3054.

Luzardo, A., Ludvig, E. A., & Rivest, F. (2013). An adaptive drift–diffusion model of interval timing dynamics. *Behav. Process.*, *95*, 90–99.

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

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

Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cogn. Brain Res.*, *3*, 227–242.

Meck, W. H. (2006). Neuroanatomical localization of an internal clock: A functional link between mesolimbic, nigrostriatal, and mesocortical dopaminergic systems. *Brain Res.*, *1109*, 93–107.

Miall, R. C. (1989). The storage of time intervals using oscillating neurons. *Neural Comput.*, *1*, 359–371.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*, 59–108.

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychol. Rev.*, *92*, 212–225.

Ratcliff, R., Cherian, A., & Segraves, M. A. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two choice decisions. *J. Neurophysiol.*, *90*, 1392–1407.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.*, *9*, 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *J. Exp. Psychol. Hum. Percept. Perform.*, *26*, 127–140.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *J. Exp. Psychol. Gen.*, *139*, 70–94.

Rivest, F., & Bengio, Y. (2011). Adaptive drift–diffusion process to learn time intervals. arXiv:1103.2382v1.

Roberts, S. (1981). Isolation of an internal clock. *J. Exp. Psychol. Anim. Behav. Process.*, *7*, 242–268.

Schrödinger, E. (1915). Zur Theorie der Fall und Steigversuche an Teilchen mit Brownscher Bewegung. *Phys. Z.*, *16*, 289–295.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proc. Natl Acad. Sci. U.S.A.*, *93*, 13339–13344.

Shankar, K., & Howard, M. (2012). A scale-invariant internal representation of time. *Neural Comput.*, *24*, 134–193.

Simen, P., Balci, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011a). Interval timing by long-range temporal integration. *Front. Integr. Neurosci.*, *5*, 28.

Simen, P., Balci, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011b). A model of interval timing by neural integration. *J. Neurosci.*, *31*, 9238–9253.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci.*, *27*, 161–168.

Smoluchowsky, M. V. (1915). Notiz über die Berechning der Brownschen Molkularbewegung bei des Ehrenhaft-millikanchen Versuchsanordnung. *Phys. Z.*, *16*, 318–321.

Staddon, J. E. R. (2001). *The new behaviorism*. Philadelphia, PA: Psychology Press.

Staddon, J. E. R., Chelaru, I. M., & Higa, J. J. (2002). A tuned-trace theory of interval-timing dynamics. *J. Exp. Anal. Behav.*, *77*, 105–124.

Staddon, J. E. R., & Higa, J. J. (1996). Multiple time scales in simple habituation. *Psychol. Rev.*, *103*, 720–733.

Staddon, J. E. R., & Higa, J. J. (1999a). Time and memory: Towards a pacemaker-free theory of interval timing. *J. Exp. Anal. Behav.*, *71*, 215–251.

Staddon, J. E. R., & Higa, J. J. (1999b). The choose-short effect and trace models of timing. *J. Exp. Anal. Behav.*, *72*, 473–478.

Stone, M. (1960). Models for choice reaction time. *Psychometrika*, *25*, 251–260.

Telfeian, A. E., & Connors, B. W. (1998). Layer-specific pathways for the horizontal propagation of epileptiform discharges in neocortex. *Epilepsia*, *39*, 700–708.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the 'internal clock'. *Psychol. Monogr.*, *77*, 1–31.

Treisman, M. (1964). Noise and Weber's Law: The discrimination of brightness and other dimensions. *Psychol. Rev.*, *71*, 314–330.

Treisman, M. (1966). A statistical decision model for sensory discrimination which predicts Weber's law and other sensory laws: Some results of a computer simulation. *Percept. Psychophys.*, *1*, 203–230.

Treisman, M. (1984). Temporal rhythms and cerebral rhythms. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences: timing and time perception*, Vol. 423 (pp. 542–565). New York, NY: New York Academy of Sciences.

Treisman, M., Cook, N., Naish, P. L. N., & MacCrone, J. K. (1992). The internal clock: Electroencephalographic evidence for oscillatory processes underlying time perception. *Q. J. Exp. Psychol. A*, *47*, 241–289.

Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705–743.

Tuerlinckx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behav. Res. Methods Instrum. Comput.*, *36*, 702–716.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.*, *108*, 550–592.

Uttal, W. R. (2008). *Distributed neural systems: beyond the new phrenology*. Cambridge, MA: Sloan Educational Publishing.

Wackermann, J., & Ehm, W. (2006). The dual klepsydra model of internal time representation and time reproduction. *J. Theor. Biol.*, *239*, 482–493.

Wald, A. (1947). *Sequential analysis*. New York, NY: Dover.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Stat.*, *19*, 326–339.

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*, 955–968.

Wearden, J. H., & Bray, S. (2001). Scalar timing without reference memory? Episodic temporal generalization and bisection in humans. *Q. J. Exp. Psychol. B*, *54*, 289–309.

Appendix: The TOPDDM Account of Weber's Law in the Nontemporal, Two-Choice Case

In contrast to the manner in which Weber's law is 'built in' to later versions of SET's IPI (Objection 2.1), Weber's law emerges naturally in the case of two-choice perceptual discriminations using the DDM. Link (1992) showed that the opponent Poisson DDM of two-choice tasks gives Weber's law for accuracy of responses, as long as it is assumed that evidence for one choice is represented by a Poisson spike process and that the spike rate is a linear representation of the strength of evidence.

Indeed, for this model, if correct responses are equivalent to first passages across the upper threshold when drift is positive, and errors are first passages across the lower threshold, then the generic DDM has the following, particularly simple expression for accuracy. This expression gives the average proportion of first passages above the upper threshold when the starting point is equidistant between upper

and lower thresholds, as would be expected when a participant is not biased toward either response (see, e.g., Luce, 1986):

$$\text{Accuracy} = \frac{1}{1 + e^{2Az/c^2}}. \tag{A1}$$

For an opponent Poisson DDM, $1/m^2$ substitutes for A/c^2 in Equation (A1), yielding:

$$\text{Accuracy} = \frac{1}{1 + e^{2z/m^2}}. \tag{A2}$$

Note that this accuracy level is constant only across conditions in which the threshold z is constant and the negative spike rate is a fixed proportion γ of the positive spike rate — that is, across conditions in which the two stimulus intensities are both multiplied by the same factor, whatever that factor may be. Accuracy should thus be constant across precisely those conditions that do produce constant accuracy according to Weber's law: i.e., when the two stimuli being compared are scaled up or down by the same factor. Intuitively, this relationship holds because as the spike rate increases in order to represent a greater stimulus intensity, the Poisson noise it contributes to the decision process increases commensurately. Thus, Weber's law in its original formulation *emerges* from the opponent Poisson assumptions, as it does in Treisman's formulation from different assumptions (Treisman, 1963, 1964). There is no need in this case to assume a logarithmic representation of subjective intensity, which Fechner derived from the scale invariance of just-noticeable-differences.