Journal of Mathematical Psychology
doi:10.1006/jmps.2002.1420, available online at http://www.idealibrary.com on IDEAL®

Hick's Law in a Stochastic Race Model with Speed–Accuracy Tradeoff

Marius Usher

Birkbeck College, University of London
E-mail: m.usher@bbk.ac.uk

Zeev Olami

Weizmann Institute of Science
E-mail: zeev@chemphys.weizmann.ac.il

and

James L. McClelland

Carnegie Mellon University and the Center for the Neural Basis of Cognition
E-mail: mcclelland+@cmu.edu

We present an analytic solution for a race model of n stochastic accumulators for multiple choice reaction time. We show that to maintain a constant level of accuracy, the response criterion needs to be increased approximately logarithmically with n, to compensate for the increase with n in the likelihood that an incorrect alternative will be most active after any fixed amount of time accumulating information. Assuming that participants monitor and maintain a constant level of performance can then explain the logarithmic dependency of the response latency on n as specified by Hick's law. Moreover, we show that for short time intervals, the Shannon information that observers extract from a stimulus, is predicted to increase linearly with processing time.

Key Words: Hick's law; choice-RT; race models; accumulators; speed–accuracy tradeoff; Wiener process; diffusion; Shannon information.

1. INTRODUCTION

Hick's law is one of the most robust regularities that has been reported in the choice response time (RT) literature (Hick, 1952; Merkel, 1885; see also Teichener & Krebs, 1974 and Welford, 1968, for reviews). In its simplest form, which applies

We thank Mario Feingold and Tony Marley for very helpful discussions.

Address correspondence and reprint requests to Marius Usher, School of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, United Kingdom.

0022-2496/02 $35.00

to reaction times in which a high level of accuracy is maintained, it relates the logarithm of the number of choice alternatives with the mean RT for a correct identification: $RT = a + b\log(n)$. In its complementary form, in which processing time is controlled and accuracy is the dependent variable, it relates the processing time of the stimulus with the amount of Shannon information extracted: $I(n) = k\,RT$. Due to its remarkable simplicity, Hick's law has attracted much attention in the choice RT literature. Originally Hick (1952) interpreted this regularity according to a sequential and hierarchical model of choice, but this was later severely criticized (Laming, 1968). Different explanations based on n parallel and *exhaustive* processes have also been developed (Christie & Luce, 1956; Laming, 1966). Later approachers to explain Hick's law were based on *self-terminating* race processes (Vickers, 1979; Vickers & Lee, 2000; Lacouture & Marley, 1991; Karpiuk, Lacouture & Marley, 1997). Those latter models, however, were specifically developed to address the case of the 1D absolute identification tasks, with stimuli organized on a 1-dimensional continuum.

In this article, we want to separate the factor that corresponds to the organizational structure of the input stimuli, from more general principles that relate to the accumulation of noisy information. In essence, our observation is that, within standard (and self-terminating) models of stochastic information accumulation (Ratcliff, 1978, 1988; Vickers, 1979; Usher & McClelland, 2001) Hick's law obtains even for stimuli which lack a specific similarity structure (i.e., they are equidistant), since as the number of alternatives increases, the chance that an incorrect response will receive more support than the correct response also increases. This is true for any fixed amount of information accumulation time t and can be compensated for by accumulating information for a longer time, since the probability that an incorrect alternative will have the most support decreases as more information is accumulated. Thus, if the observer adjusts a response criterion to maintain a constant level of accuracy as the number of alternatives increases, then reaction times will increase with n. Remarkably, our analysis shows that this increase with n closely approximates a function that grows linearly with the logarithm of n. Related to this, we also find that when the response time is controlled the amount of information extracted from the stimulus display increases linearly with time, at least for short time intervals. Here we present an analysis of this situation, based on a simple stochastic race model for choice. However, our claim is that Hick's law is not a property of any particular process model, but of the underlying statistics of the detection of signals in noise.[1]

The simplifying assumption of equidistance of input stimuli is inconsistent with experiments that use stimuli on a one-dimensional continuum; however, it serves as a useful approximation for experiments where the stimuli are visual patterns in a high dimensional input space (including visually presented digits, numerals, or words). While this is not the case for *all* the experiments that reported Hick's law regularities, there are experiments where this situation may apply. For example, in the original study performed by Merkel (1885) the stimuli were the set of numerals

[1] Just as signal detectability theory is based on the inherent statistical properties of a static sample of information perturbed by random noise, so our theory of Hick's law is based on the inherent statistical properties of a series of samples of such intrinsically noisy information.

1–5 and I–V. Although there are likely to be some nonhomogeneities in the similarity structure of the stimuli in this and nearly all other experiments, it is still worth dissociating the effect of this factor from the most basic effect of increasing the number of choice alternatives. Moreover, the equidistance assumption is helpful in order to allow the derivation of analytical results.

Findings similar to the findings presented here were presented in a previous article (Usher & McClelland, 2001) where we proposed a neurocomputational model that addressed many aspects of the time course of information processing. The model was based on leaky, competing accumulators which integrate information and which generate a response at the moment when one of the accumulators reaches a response criterion. Within this model, we discovered that a close approximation to Hick's law arises if one assumes that subjects maintain a constant level of accuracy as the number of alternatives increases. Each alternative is associated with a dynamical activation value that undergoes a stochastic diffusion process, and as more alternatives are added the chance that the activation of an incorrect alternative will reach the criterion activation level before the correct alternative increases. To maintain a constant level of accuracy, the criterion needs to be adjusted in order to compensate for this. This assumption of the model is consistent with a feature of the procedure that Hick used in his 1952 experiment. Because he was interested in errorless performance, he required subjects to undergo preliminary blocks with each set size. During these blocks, subjects were instructed to adjust their performance to eliminate errors while still responding as rapidly as possible. In Hick's experiment, several blocks of trials were run at each set size, and the data reported were taken from the block following the first block of the given set size in which no errors were made. Given this aspect of the procedure, we adopted the assumption that subjects adjust the criterion as a function of number of alternatives so as to keep accuracy constant at a high level when the number of stimuli increases. Vickers (1979, Chap. 8; Vickers & Lee, 2000) used a similar approach, based on confidence (rather than accuracy) which is self regulated, in his account of Hick's law, as it is applied to stimuli arrayed along a single dimension.

In our previous article (Usher & McClelland, 2001) we relied on Monte Carlo simulations to show that a family of race models (race of accumulators with and without leakage and competition) satisfy Hick's logarithmic regularity. That is, the criterion needs to be increased as a logarithmic function of the number of choices in order to maintain a constant level of performance. Here we pursue the simplest member of this family of models, a race between nonleaky accumulators or diffusion processes (Ratcliff, 1978, 1988; Vickers, 1970, 1979), in order to provide a tractable formulation showing analytically how the adjustment of the criterion can compensate for the increased likelihood of error and leads to a logarithmic relation between RT and the number of choices. We also extend the analysis to show that the approach accounts for the time-controlled variant of Hick's law.

2. HICK'S LAW AT HIGH, CONSTANT ACCURACY

We assume that each alternative is represented by an accumulator, whose activation is x_i and which receives sensory input, I_i, proportional to the match of the accumulator's preferred input to the probe stimulus. The accumulation of activation is noisy; Gaussian noise, ξ, of zero mean and variance σ^2 is added to the activations x_i. As a result the activation of each unit performs a Wiener diffusion process with drift I_i and variance σ^2. This is expressed in the following stochastic differential equation, where the noise term is proportional to the square-root of the dt interval due to the fact that the variance and not SD is linear with dt,

$$dx_i = I_i dt + \xi \sqrt{dt}. \tag{1}$$

We can assume without loss of generality that the stimulus corresponds to the first response unit, x_1 and its drift is $I_1 = \mu$ (the drift value is kept here fixed, as we do not investigate variations in the stimulus quality). Following the equidistance assumption, we chose all the drifts of the incorrect responses to be equivalent, and for simplicity we consider the case where these drifts (I_i; $i > 1$) are all equal to 0. (This assumption is made to allow analytical derivations, but see Usher & McClelland, 2001). We also choose the variance of the diffusion process, $\sigma^2 = 1$. This is equivalent to measuring the random walk in units of σ. The choice terminating the trial and determining the RT is made when the activation of one of the accumulators reaches a boundary at distance θ from the origin (see Vickers, 1970).

For $n = 1$ (a single diffusion process) the distribution of arrival times (or first-passage times) corresponds to a Wald distribution (Ricciardi, 1977; Luce, 1986)

$$g(\theta, t) = \frac{\theta}{\sqrt{2\pi t^3}} \exp\left[-\frac{(\theta - \mu t)^2}{2t} \right], \tag{2}$$

where μ is the drift of the diffusion process and θ is the response criterion. In the Appendix we illustrate (Fig. 3) the density distribution of arrivals times for a diffusion with drift $\mu = 1$ (correct accumulator) and with drift $\mu = 0$ (incorrect accumulator), showing that the arrival time distribution of the process with zero drift (dashed curve) is slower.

We can now calculate the survival functions (the complement of the cumulative distribution functions) corresponding to the density distributions in Eq. (2), for $\mu = 0$, which corresponds to the probability that the diffusion process with zero drift has not yet arrived at the criterion θ by time t:

$$G_0(\theta, t) = \int_t^{\infty} g_0(\theta, t')\, dt' = \int_0^{\frac{\theta}{\sqrt{2t}}} \exp(-x_0^2)\, dx = Erf\left(\frac{\theta}{\sqrt{2t}} \right). \tag{3}$$

The probability for a correct response in the race process and the RT distribution of correct responses can be computed by considering the probability that the process with drift reaches the boundary before all the other processes; this requires multiplying the probability that the first accumulator reaches criterion in the time interval $(t, t + dt)$, $g_1(\theta, t)\, dt$, with the probability that none of the other accumulators reached the criterion by time t, $G_0(\theta, t)^{n-1}$. The probability density for a correct response at time t is: $\rho(n, \theta, t) = g_1(\theta, t)[G_0(\theta, t)]^{n-1}$. From this the probability for a correct response and the mean RT can be obtained,

$$P(n,\theta) = \int_0^\infty \rho(n,\theta,t)\,dt \tag{4}$$

$$\langle T(n,\theta)\rangle = \frac{1}{P(n,\theta)}\int_0^\infty t\rho(n,\theta,t)\,dt. \tag{5}$$

In the Appendix we develop an analytical approximation to Eq. (4), which reduces this relatively complex integral to a simpler formula:

$$P(n,\theta) \approx Erf^{n-1}\sqrt{\frac{\theta\mu}{2}}. \tag{6}$$

This equation shows that the accuracy is controlled by two parameters: the number of alternatives, n, and the product $\theta\mu$. Changes in drift, are thus equivalent (with regards to accuracy) with a corresponding scale of the criterion. As we focus on the dependency of accuracy and RT on n (and not on stimulus quality) we choose $\mu = 1$ in the discussion below. Increasing n and θ has opposite effects on the correct response probability, $P(n,\theta)$. For a specific θ value, increasing the number of choices leads to a decrease in performance, while for a specific n increasing the response threshold θ improves performance. The idea suggested here (see also Usher & McClelland, 2001) is that a logarithmic dependency of θ on n can maintain the correct response probability at an approximately constant level $P[n,\theta(n)] = const.$ We show this in two ways:

First, using a logarithmic dependency, $\theta(n) = 2.4 + 2\log_2 n$ (chosen so as to obtain a performance of about 96%), we plot in Fig. 1A the performance level that results when this dependency is substituted in Eq. (4). For n in the range of 2–10 the performance remains approximately constant (i.e., within a range of 0.005 around 0.96). Second, we solved numerically the equation $P[\theta(n), n] = 0.96$ (Eq. (4)) with a higher precision of 0.001. The obtained values of $\theta(n)$ are plotted (solid line) on log-linear scale in Fig. 1B. The response times predicted by the model from these θ values (given by Eq. (5)) are plotted (with dots) in the same figure. It is known that for a Wald distribution (Eq. (2)) the mean RT is $\langle T(1,n)\rangle = \theta(n)$, and the figure indicates that the same thing holds to a good approximation for the race process of accumulators. The threshold and the RT both show an approximately logarithmic

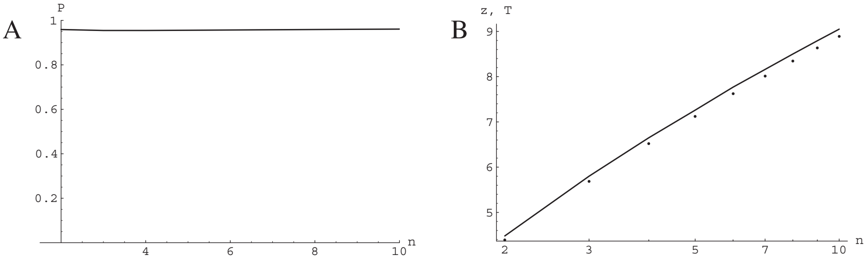

Fig. 1. (A) The percentage correct, according to Eq. (4) for $\theta(n) = 2.4 + 2\log n$. (B) The threshold $\theta(n)$ (solid line) and the corresponding RT (dots) (Eq. (5)) on log-linear scale for $\theta(n)$ that is chosen so as to maintain $P = 0.96$ at a precision of 0.001.

growth with n (i.e., linear on a log-linear scale). In the Appendix we present an intuitive explanation for the logarithmic dependency (based on approximations, valid for large θ, i.e., high P-levels, and large or moderate n).

3. HICK'S LAW IN TIME CONTROLLED EXPERIMENTS

Hick (1952) noticed that the relationship $T = K \log_2 (n)$, typically obtained when subjects operate at a high and fixed level of accuracy can be generalized in terms of the amount of information, I, extracted from the display as a function of processing time. The extracted information is a measure of the decrease in the subject's uncertainty regarding the stimulus, after inspecting the display. When subjects perform with perfect accuracy, no uncertainty remains after they perceive the display, and the extracted information is equal to the original uncertainty. For n equiprobable stimuli, this uncertainty is given by the entropy function, $I(n) = \log_2(n)$. Hick's law can therefore be generalized to

$$T \propto I(n). \tag{7}$$

In his second and third experiment, Hick (1952) required subjects to perform on a given response-set ($n = 10$) under various speed instructions. The result, confirmed by later experiments that controlled the speed accuracy tradeoff using response-deadlines at various set-sizes (Pachella *et al.*, 1968; Pachella & Fisher, 1972), is that Eq. (7) holds also for speeded responses. In this case, however, the extracted information equals the initial uncertainty minus the uncertainty left after the presentation of the display and reflected in the pattern of response errors. If one denotes by p_{ij} the joined probability of generating response i after perceiving stimulus j (i.e., $p_{ij} = P(i \mid j)\, p_j$; and $\sum_{ij} p_{ij} = 1$), one can calculate the extracted information (see, e.g., Hick, 1952; Welford, 1968), according to

$$I = \sum_{ij} p_{ij} \log_2 \frac{p_{ij}}{p_i\, p_j}, \tag{8}$$

where p_j and p_i are the marginal probabilities for the presentation of stimulus j and for the generation of response i, respectively. Here we consider the case of equiprobable stimuli and responses, thus $p_i = p_j = 1/n$. Furthermore we assume that response matrix is homogeneous (i.e., that when a subject makes an error he is equally likely to respond with any of the incorrect response alternatives).

Thus we assume that diagonal elements p_{ii} are

$$p_{ii}(n, T) = \frac{p(n, T)}{n}, \tag{9}$$

where $p(n, T)$ is the probability for a correct response in an n-alternative choice task performed in time T (i.e., this probability is equally divided among all the diagonal elements). Similarly, for the nondiagonal elements (there are $n(n-1)$ of them), p_{ij}, $i \neq j$ is

$$p_{ij} = \frac{1 - p(n, T)}{n(n-1)}. \tag{10}$$

With this, p_{ij} satisfies: $\sum_{ij} p_{ij} = 1$. From Eqs (8), (9), and (10) one obtains

$$I(n, T) = \log_2 (n) + p(n, T) \log_2 (p(n, T)) + (1 - p(n, T)) \log_2 \left[\frac{1 - p(n, T)}{n - 1} \right]. \quad (11)$$

Here we show that the probability, $p(n, T)$, can be computed from the Wiener diffusion process of the race model. We begin by noting that for very short times, this process gives a chance level of performance (the correct and incorrect activations have distributions with identical means), $P(n, 0) = 1/n$, and that at long times the performance level reaches 1 ($\lim_{t\to\infty} P(n, t) = 1$—the correct/incorrect distributions are well separated.[2]) One can check that in those two limits, $I(n, 0) = 0$ (zero information at time zero) and that $\lim_{t\to\infty} I(n, t) = \log_2(n)$ (the entropy of the set).

If we assume that at $t = 0$ all the accumulators are initialized as $x_i(0) = 0$ and they evolve according to a Wiener diffusion process (1) the probability distributions of the correct accumulator $P(x_1, t)$ and of the incorrect ones, $P(x_j, t)$ ($j \neq 1$) are Gaussian distributions with mean t and 0, respectively (corresponding to the noiseless trajectories) and with standard deviations of $SD = \sqrt{t}$ (Ricciardi, 1977; Luce, 1986).

When the RT is controlled (subjects are required to make a selection at time t), the probability for a correct response, $p(n, t)$, can be assumed to reflect the probability that the activation of the correct accumulator (with drift $\mu = 1$) is larger than the activation of all the other accumulators (with drift $\mu = 0$) at time t (Ratcliff, 1978). Thus one needs to multiply the density probability distribution of the first accumulator at activation level x with the cumulative distribution probabilities of all the other accumulators at x, $[\frac{1}{2}(1 + Erf(\frac{x}{\sqrt{2t}}))]^{n-1}$, and integrate across all possible x values, to obtain the probability that at time t, x_1 has an activation larger than that of all the other units.

$$p(n, t) = \frac{1}{\sqrt{2\pi t}} \frac{1}{2^{n-1}} \int_{-\infty}^{\infty} \left[1 + Erf\left(\frac{x}{\sqrt{2t}} \right) \right]^{n-1} \exp\left[-\frac{(x - \mu t)^2}{2t} \right] dx \quad (12)$$

In Fig. 2A, we show the time-controlled accuracy curves for a race of $n = 2$, $n = 3$, $n = 4$, $n = 5$ accumulators. One can see at $t = 0$, the performance corresponds to the chance level (guessing) and it gradually increases toward perfect performance at long times. The information extracted is computed, for each of these n-values, (Eq. (11)) and displayed in Fig. 2B. One can see that the increase in the information is approximately linear in time and that the slopes are very similar (except for $n = 2$ that has a relatively lower slope).

[2] In choice experiments where the stimuli are confusable, performance does not reach a perfect level even at long times (Swensson, 1972). This factor can be accounted for by modifications of the diffusion model, such as variability between trials in the drift parameter (Ratcliff, 1978) or leakage of the activation that results in an Ornstein–Uhlenbeck (OU) diffusion process (Busemeyer & Townsend, 1993; Usher & McClelland, 2001). Here we focus, on choice between well discriminable stimuli sets (which are constructed in a high dimensional space), and where the Wiener diffusion process may provide a reasonable first approximation [see however, Usher and McClelland (2001) for a discussion of Hick's regularity in the OU-process]. In particular for processing times short relative to the time OU scale, the OU process can be approximated by a Wiener diffusion process.

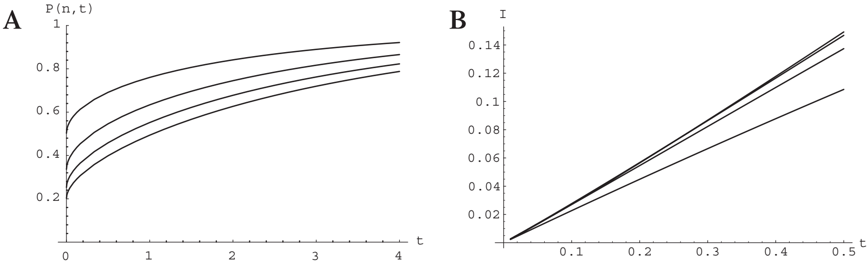

FIG. 2. (A) The performance level $p(n, t)$ corresponding to Eq. (12) is displayed for $n=2, 3, 4, 5$. (B) The information extracted according to Eq. (11), where $p(n, t)$ corresponds to the curves in A (the line with the lowest slope corresponds to $n=2$).

To demonstrate that, for small t, the information increases with linearly with time, we develop $p(n, t)$ (Eq. (12)) in powers of t. Neglecting (for $t<1$) the third term in the exponent $\exp(-\frac{x^2}{2t}+x\mu-\frac{\mu^2 t}{2})$, and developing $\exp(x\mu) \approx 1+x\mu$, we obtain[3]

$$p(n, t) \approx \frac{1}{n}+f_n \mu \sqrt{t}, \tag{13}$$

where

$$f_n=\sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty}\left[\frac{1}{2}(1+Erf(y))\right]^{n-1} \exp(-y^2)\, y\, dy. \tag{14}$$

For short times the extracted information can be calculated (using Eqs. (13) and (11)),

$$I(n, t) \approx \frac{n^2}{n-1} f_n^2 \mu^2 t. \tag{15}$$

It it interesting that despite the fact that the probability $p(n, t)$ has a singularity (infinite derivative) at $t=0$ and therefore does not grow linearly with time, the extracted information is linear in time (for short time intervals) in accordance with Hick's law.

4. DISCUSSION

We have presented a simple scheme that provides a robust explanation for Hick's regularity in choice RT, on the basis of the assumption that subjects in such experiments strive to maintain a constant level of accuracy as the number of alternatives increases. This scheme shows that a multiple-accumulator model requires an approximately logarithmic increase in the response criterion (and therefore in RT) in order to maintain a constant level of accuracy with increasing set-size. This same

[3] The first term, $\frac{1}{n}$, results from the integral with $\exp(-\frac{x^2}{2t})$: $\frac{1}{\sqrt{2\pi t}} \frac{1}{2^{n-1}} \int_{-\infty}^{\infty} [1+Erf(\frac{x}{\sqrt{2t}})]^{n-1} \exp(-\frac{x^2}{2t})\, dx=\frac{1}{n}$.

scheme also explains the linear relation between the amount of extracted information (which is a nonlinear function of the probability of correct response and of the set size) and processing time in time controlled paradigms. In our previous work (Usher & McClelland, 2001) we provided simulation results that indicate that this scheme is robust enough to hold for less restrictive type of models (leaky competing accumulators, as well as for a race of diffusion processes coupled with the use of a criterion based on the difference in activation between accumulators, as used in random-walk type models).

This mathematical analysis indicates that Hick's law regularity is a consequence of the statistical properties of adding more accumulators on a stochastic information accumulation process. Together with simulations showing that the same results hold up with various additional assumptions in place (Usher & McClelland, 2001), it demonstrates that a principle of a constant rate of information transmission is not needed to obtain the Hick's law regularity. We suggest that a major determinant of this regularity is the intrinsic increase in likelihood of spuriously reaching threshold as a function of an increase in the number of alternatives in conjunction with the attempt to maintain a fixed level of performance. Unlike the principle of fixed rate of information transmission that was shown to be unplausible for human choice performance (Laming, 1968), these principles are standard within the choice RT theory (Ratcliff, 1978; 1988; Vickers, 1979) and consistent with neurocomputational principles (Usher & McClelland, 2001). Moreover, the analysis we presented here predicts that the logarithmic dependency of RT on n (at high accuracy) depends critically on the maintenance of an approximately constant level of performance and that the slopes of information extracted on processing time may show small variations with n.

Several aspects of the results require some caution. First, the logarithmic dependency between set size and RT is only an approximation (careful examination of the graph in Fig. 2B may reveal a slight underestimation at $n = 2$.) Due to the noise fluctuations present in most data sets, it is not clear whether actual data could ever detect so small a discrepancy. It should also be noted that our analysis does not take into account any possible similarity relations among the stimuli, even though such similarities do clearly play some part in real experiments. To address this issue, a variety of encoding assumptions will need to be explored, extending our analyses as well as those of Vickers (1979; see also Vickers & Lee, 2000) and Lacouture & Marley (1991; see also, Karpiuk, Lacouture, & Marley, 1997). Future research on Hick's law should also examine more discriminative measures of the RT (such as error latencies and distributional properties (Roberts & Pashler, 2000)). Finally, the use of internal metacognitive signals such as confidence (Vickers, 1979; Vickers & Lee, 2000) may be used by participants when feedback about performance is not available. Research addressig these issues are likely to capture the fundamental nature of the basis for Hick's regularity.

APPENDIX

The density distribution of arrivals times (Eq. (2)) for a diffusion with drift $\mu = 1$ (correct accumulator) and with drift $\mu = 0$ (incorrect accumulator) are illustrated in

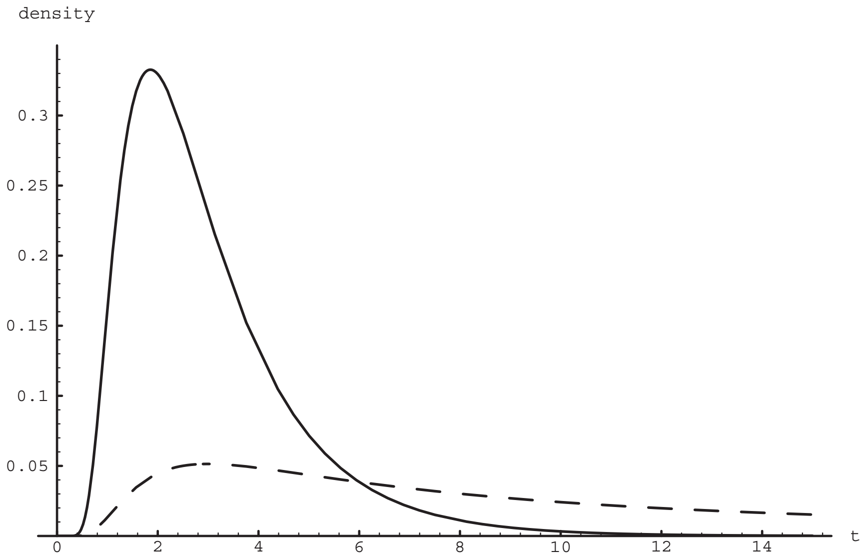

FIG. 3. Distribution densities for arrivals times in a diffusion process, Eq. (2), for $\mu = 1$ and $\mu = 0$ (no drift). The response threshold θ is set to 3.

Fig. 3. The arrival time distribution of the process with zero drift (dashed curve) is slower. (At larger t it decays as a power law, $t^{-3/2}$ while the process with drift (solid curve) has a faster exponential tail. The mean RT for a Wald distribution, $g(\theta, t)$, with drift 1 and response threshold, θ, is $\langle T \rangle = \theta$.

The criterion, $\theta(n)$ that maintains a fixed accuracy, $P[n, \theta(n)]$ is (Eq. (4))

$$P(n, \theta) = \frac{\theta}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{t^3}} \exp\left[-\frac{(\theta - \mu t)^2}{2t} \right] Erf^{n-1} \left[\frac{\theta}{\sqrt{2t}} \right] dt$$

by changing the integration variable from t to $x = \theta / \sqrt{2t}$, one obtains

$$P(n, \theta) = \frac{2}{\sqrt{\pi}} \exp(\theta\mu) \int_0^\infty \exp\left[-\left(x^2 + \frac{\theta^2 \mu^2}{4x^2} \right) \right] Erf^{n-1}(x)\, dx. \tag{16}$$

Since the integrand is a multiplication of a strongly peaked function $F(x) = \exp[-(x^2 + \frac{\theta^2\mu^2}{4x^2})]$ with a slowly varying function $G(x) = [Erf^{n-1}(x)]$, one can approximate the integral using the *steepest descent* method of approximation (e.g., Arfken, 1970, pp. 373–376)

$$\int F(x)\, G(x)\, dx \approx \sqrt{2\pi} \frac{F(x_{max})\, G(x_{max})}{\sqrt{\left.\frac{d^2 \log_2(F(x))}{d^2 x}\right|_{x = x_{max}}}},$$

where, x_{max} is the value of x at the maximum of the $F(x)$ function: $\frac{dF(x)}{dx}\big|_{x = x_{max}} = 0$.

Applying this to Eq. (16), with $x_{max}^2 = \frac{\theta\mu}{2}$ leads to

$$P(n, \theta) \approx Erf^{n-1} \sqrt{\frac{\theta\mu}{2}}. \tag{17}$$

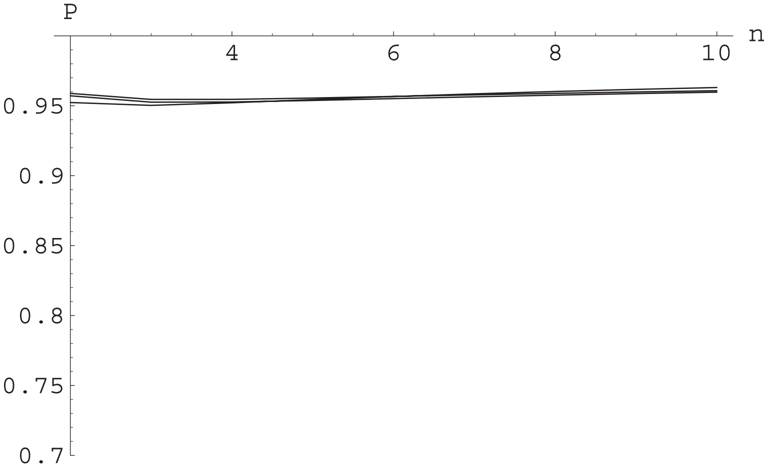

FIG. 4. The performance as function of n, according to Eqs. (4), (17), (18), for $\theta(n) = 2.4 + 2\log_2(n)$ and $\mu = 1$. The approximation in terms of the Erf–function is virtually indistinguishable from the integral formula. Very slight deviations appear in the second approximation (underestimation at low n and over estimation at large n) that can be removed by introducing a second order in the expansion.

One more approximation can be obtained by replacing the Erf function with its asymptotic expansion at large x (valid for P close to one): $Erf(x) \approx 1 - \frac{\exp(-x^2)}{\sqrt{\pi}\,x}$;

$$P[n, \theta] \approx \left[1 - \frac{\exp\left(-\frac{\theta\mu}{2}\right)}{\sqrt{\frac{\pi\theta\mu}{2}}} \right]^{n-1}. \tag{18}$$

The accuracy of these approximations is illustrated in Fig. 4, where we plot the performance obtained from the full integral formula 4, as well as from the two approximations Eqs. (17)–(18), for $\theta(n)$ chosen according to the logarithmic dependency, $\theta(n) = 2.4 + 2\log_2(n)$. All three lines show an approximately constant performance (see also Fig. 1). An intuitive justification for the logarithmic dependency of $\theta(n)$ can be obtained by considering $\ln P[n, \theta]$ in Eq. (18). Fixing the drift values to $\mu = 1$ and neglecting to a first approximation the dependency of θ on n in the denominator, for large θ, this gives

$$\ln[P(n, \theta(n))] = (n-1)\ln\left[1 - \exp\left(-\frac{\theta(n)}{2}\right)\right] \approx -(n-1)\exp\left[-\frac{\theta(n)}{2}\right].$$

A logarithmic dependency, $\theta(n) = a + b\ln(n)$ is therefore needed to balance the effect of the number of incorrect choices, $n-1$ and to maintain P at an approximately constant level.[4]

[4] This consideration works well at large n ($n \approx n-1$) when $b = 2$ [i.e., $exp[-\theta(n)/2] = n^{-1}$]. For n in the realistic range of 2–10, a constant P is obtained using the dependency: $\theta(n) = 2.4 + 2\log_2(n)$. In this case, $ln[1/P(n)] \approx (n-1)/n^{1.44}$. While the variable $(n-1)/n$ increases monotonically with n (from 0.5 to 1), when n (at denominator) is raised to a power larger than 1, this leads to a diminishing effect on P. For powers in the range of 1.2–1.5, the two effects balance and the change in the value of P is relatively small.

REFERENCES

Arfken, G. (1970). *Mathematical methods for physicists*. Academic Press: New York.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory. *Psychological Review*, **100**, 432–459.

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics*, **18**, 89–112.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, **4**, 11–26.

Karpiuk, P., Jr., Lacouture, Y., & Marley, A. A. J. (1997). A limited capacity, wave equality, random-walk model of absolute identification. In A. A. J. Marley (Ed.), *Choice, decision and measure: Essays in Honor of R. Duncan Luce*. Mahwah, NJ: Erlbaum.

Laming, D. R. (1966). A new interpretation for the relation between choice reaction time and the number of equiprobable alternatives. *British Journal of Mathematical and Statistical Psychology*, **19**, 139–149.

Laming, D. R. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Lacouture, Y., & Marley, A. A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science*, **3**, 401–433.

Luce, D. R. (1986). *Response times*. New York: Oxford University Press.

Merkel, J. (1885). Die Zeitlichen Verhältnisse der Wiellensthätigkeit. *Philosophische Studien*, **2**, 73–127.

Pachella, R. G., Fisher, D., & Karsh, R. (1968). Absolute judgments in speeded tasks: quantification of the trade-off between speed and accuracy. *Psychonomic Science*, **12**, 225–226.

Pachella, R. G., & Fisher, D. (1972). Hick's law and the speed-accuracy trade-off in absolute judgment. *Journal of Experimental Psychology*, **92**, 378–384.

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

Ratcliff, R. (1988). Continuous versus discrete information processing; modeling the accumulation of partial information. *Psychological Review*, **92**, 212–225.

Ricciardi, L. (1977). *Diffusion processes and related topics in biology*. New York: Springer-Verlag.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, **107**, 358–367.

Swensson, R. G. (1972). The elusive trade-off: Speed versus accuracy in visual discrimination tasks. *Perception and Psychophysics*, **12**, 16–32.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review*, **81**, 75–98.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky competing accumulator model. *Psychological Review*, in press.

Vickers, D. (1970). Evidence for an accumulator of psychophysical discrimination. *Ergonomics*, **13**, 37–58.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multi-choice tasks. *Non-linear Dynamics, Psychology and Life Sciences*, **4**, 1–31.

Welford, A. T. (1968). The measurement of sensory-motor performance: survey and reappraisal of twelve years of progress. *Ergonomics*, **3**, 189–230.

Received May 24, 2001