*LETTER* Communicated by Xiao-Jing Wang

# Decision Confidence and Uncertainty in Diffusion Models with Partially Correlated Neuronal Integrators

**Rubén Moreno-Bote**
*rmoreno@bcs.rochester.edu*
*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, U.S.A., and Center for Neural Science, New York University, New York, NY 10003, U.S.A.*

**Diffusion models have become essential for describing the performance and statistics of reaction times in human decision making. Despite their success, it is not known how to evaluate decision confidence from them. I introduce a broader class of models consisting of two partially correlated neuronal integrators with arbitrarily time-varying decision boundaries that allow a natural description of confidence. The dependence of decision confidence on the state of the losing integrator, decision time, time-varying boundaries, and correlations is analytically described. The marginal confidence is computed for the half-anticorrelated case using the exact solution of the diffusion process with constant boundaries and compared to that of the independent and completely anticorrelated cases.**

## 1 Introduction

In most interesting situations, detection of a signal involves discriminating a meaningful from a random energy fluctuation. A standard procedure employed to discriminate between these two alternatives is to integrate over time such fluctuations and decide that there is a signal embedded in the noise if the accumulated sum reaches a predetermined threshold and that there is no signal otherwise. Although this is just the way some photosensors are used to detect dim stars, humans seem to employ a remarkably similar strategy to make decisions. Human performance and reaction time statistics in two-alternative forced-choice tasks are accurately fit by models consisting of a diffusion process of an underlying variable to decision boundaries (Ratcliff, 1978; Ratcliff & Smith, 2004; Luce, 1986; Usher & McClelland, 2001; Bogacz et al., 2006; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007; Wang, 2002; Beck et al., 2008). Moreover, diffusion to bound processes seem to be neurally hardwired in several areas of the brain, where the neurons' firing rate reflects integration of noisy information for alternatives and determines choices and decision times when it reaches a fixed threshold (Hanes &

 

Schall, 1996; Shadlen & Newsome, 1996; Platt & Glimcher, 1999; Ratcliff et al., 2007; see Schall, 2001, for a review).

Upon a decision, a crucial variable is the degree of confidence about the choice. In detecting a new particle or star, the estimated confidence will determine if we can believe in the result. Humans and animals are able to reliably estimate confidence or uncertainty about their decisions and use it to change or abort behaviors (Vickers, 1979; Juslin & Olsson, 1997; Vickers & Lee, 1998; Landy, Goutcher, Trommershuser, & Mamassian, 2007; Whiteley & Sahani, 2008; Kepecs, Uchida, Zariwala, & Mainen, 2008). Human subjects can be asked directly to report their confidence, but it is also possible to use choice behaviors in humans and animals to estimate their decision confidence. For instance, a subject can be presented with two choices: one that involves high but uncertain reward and another with low but certain reward. The confidence that the subject has about the uncertain reward at any trial will determine which option is taken, and therefore decision confidence can be estimated from its behavior at any trial. Despite the importance of decision confidence, very little is known about how it is encoded in the nervous system (but see Kepecs et al., 2008; Beck et al., 2008). Furthermore, the dependence of confidence on essential parameters of the decision process like reaction time or distance to boundary of the losing integrator has not been established. Knowing such dependencies on simple models would allow a quantitative comparison between theory and both behavioral and neurophysiological data.

This letter describes analytically decision confidence for a decision diffusion model (DDM) consisting of two neuronal integrators receiving partially correlated inputs. The main results are as follows: (1) decision confidence depends on the difference between the traveled distances by the two integrators in a general way, regardless of the correlation in the inputs and on the time dependence of the decision thresholds; (2) accumulation of evidence over time occurs without any leak or loss, that is, the final state of the integrators and elapsed time fully specify the a posteriori probability distribution over the parameters on which the decision is based; (3) marginal confidence is computed for the independent, half-anticorrelated, and completely anticorrelated cases as a function of the decision time, being at a maximum for independent integrators; and (4) the distributions of reaction times for the above cases are characterized. Models with several integrators similar to the one proposed here have been introduced before to describe qualitatively decision confidence (Vickers, 1979), but analytical expressions for confidence are in general not known for these simple models (Juslin & Olsson, 1997; Vickers & Lee, 1998; Van Zandt, 2000; Pleskac & Busemeyer, 2007). It is worth mentioning that the DDM considered here reduces to the classical DDM (Ratcliff & Smith, 2004; Luce, 1986) when the two integrators receive completely anticorrelated inputs, for which decision confidence as a function of the decision time is also completely characterized for arbitrary time-dependent boundaries.

## 2 Model

I consider a decision diffusion model (DDM) consisting of two neuronal integrators with variables $x_1$ and $x_2$ receiving independent and cross-correlated inputs,

$$\dot{x}_1 = \mu_1 + \sigma\ [\sqrt{1-\rho}\ \eta_1(t) + \sqrt{\rho}\ \eta_c(t)],$$
$$\dot{x}_2 = \mu_2 + \sigma\ [\sqrt{1-\rho}\ \eta_2(t) + \sqrt{\rho}\ \nu\ \eta_c(t)], \tag{2.1}$$

where $\eta_i(t), i = 1, 2, c$ are independent gaussian white noise processes with zero mean ($\langle \eta_i(t) \rangle = 0$) and unit variance ($\langle \eta_i(t)\eta_i(t') \rangle = \delta(t-t')$, where $\delta(t-t')$ is the Dirac's delta function).[1] The variables $x_1$ and $x_2$ can correspond to averaged firing rates across a population of neurons with similar tuning, or to population-averaged voltages (Mazurek et al., 2003). The model captures the approximately linear ramping activity observed in neurons in lateral intraparietal area (LIP) and frontal eye fields (FEF) in monkeys performing simple decision tasks (Hanes & Schall, 1996; Shadlen & Newsome, 1996). The $k$th neuronal integrator ($k = 1, 2$) drifts with rate $\mu_k$ and accumulates input white noises with total deviation $\sigma$. The correlation coefficient, $\rho \in [0, 1]$, determines the degree of correlation between the two integrators, being independent if $\rho = 0$ and completely correlated if $\rho = 1$. The correlation index, $\nu \in \{-1, 1\}$, determines the correlation sign between them, being positively correlated if $\nu = 1$ and anticorrelated if $\nu = -1$. The variables $x_1$ and $x_2$ diffuse from the initial values $a$ and $b$ until one of them reaches its threshold, $\Theta_1(t)$ and $\Theta_2(t)$, respectively, which can depend on time.

A two-alternative forced-choice task is modeled as follows. A decision for choice $k$ ($k = 1, 2$) is made whenever the variable $x_k$ reaches its threshold $\Theta_k(t)$ first. It can be assumed that the decision is correct if the boundary first reached corresponds to that of the integrator with the highest drift. This is a general situation that, for instance, could correspond to a task in which the goal is to detect the brighter of two lights. It also arises naturally in experiments in which the direction of motion of a random-dot motion display

[1] In numerical simulations, equation 2.1 is solved using the Euler method with small time step $\delta t$ as

$$\delta x_1 = \mu_1 \delta t + \sigma\ [\sqrt{1-\rho}\ w_1 + \sqrt{\rho}\ w_c]\sqrt{\delta t},$$
$$\delta x_2 = \mu_2 \delta t + \sigma\ [\sqrt{1-\rho}\ w_2 + \sqrt{\rho}\ \nu\ w_c]\sqrt{\delta t},$$

where $\delta x_i$ (i = 1, 2) are the variable increments in each step and $w_i$ ($i = 1, 2, c$) are independent across time and follow normal gaussians.

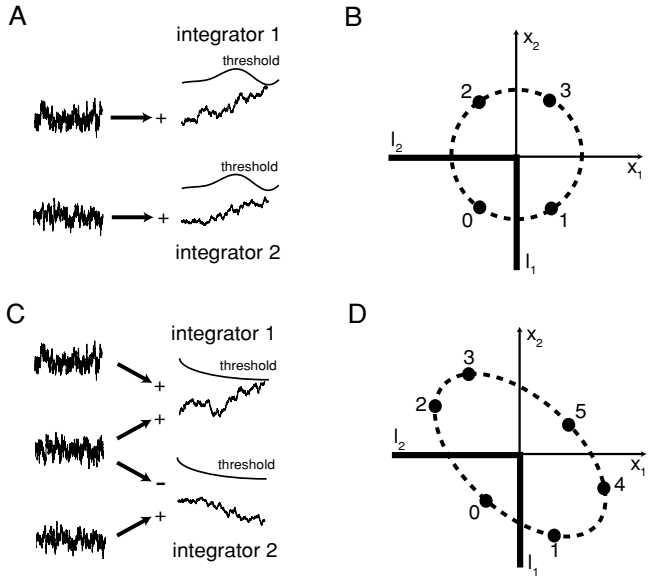


Figure 1: (A) Two integrators with independent white noise inputs. The system evolves until one of the integrators reaches its time-varying threshold—in this case, integrator 1. (B) The state probability density for the system in $A$ when the boundaries are constant over time consists of three images, equation A.9, arranged on a circle (see the appendix). The two semi-lines of the absorbing boundary are indicated with $l_1$ and $l_2$. (C) Two integrators with time-dependent boundaries and half-anticorrelated white noise inputs. (D) The state probability density for the system in panel $C$ with constant bounds consists of five images, equation A.8, arranged on an ellipse (see the appendix).

has to be determined during short presentations (Shadlen & Newsome, 1996; Mazurek et al., 2003). Here, the direction of motion can be modeled by the sign of the drift rates of the two integrators (i.e., $\mu = \mu_1 = -\mu_2 > 0$ indicates rightward motion, while $\mu = \mu_1 = -\mu_2 < 0$ indicates leftward motion). The difficulty of this task is controlled by the motion coherence, that is, the percentage of dots that move in the same direction as opposed to random directions, and it can be modeled by the magnitude of the drift rates, $|\mu| = |\mu_1| = |\mu_2|$. It will be assumed that the variance of the noise and correlation coefficient remains constant across all stimulation conditions. I assume that when a decision has been made, the state of the losing integrator can be read out. Figure 1 illustrates the model for the independent ($\rho = 0$ in Figure 1A) and half-anticorrelated ($\rho = 1/2$, $\nu = -1$, in Figure 1C) cases with time-varying thresholds.

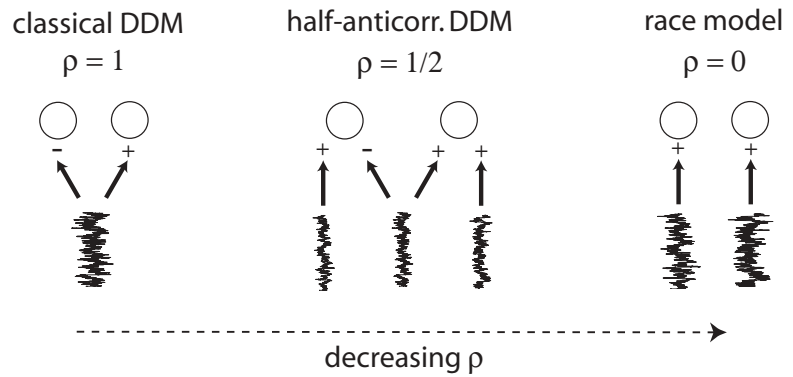


Figure 2: The DDM in equation 2.1 covers continuously the range between diffusion models (classical DDM) and race models as a function of the correlation coefficient between the input noises. A mixture model corresponds, for example, to the DDM with half-anticorrelated inputs. Circles represent integrators.

The classical DDM (Ratcliff & Smith, 2004; Luce, 1986) is a special case of equation 2.1 when the integrators receive completely anticorrelated inputs ($\rho = 1$, $\nu = -1$), $\mu = \mu_1 = -\mu_2$, and the thresholds are identical and time independent, $\Theta_1(t) = \Theta_2(t) = \Theta$. In that case, only motion parallel to the diagonal $x_1 = -x_2$ is possible, and therefore each integrator can be viewed as the antineuron of the other. Equations 2.1 reduce in this case to the system

$$\dot{x} = \mu + \sigma\ \eta(t), \tag{2.2}$$

where $x = x_1 = -x_2$ and the decision boundaries lie at $x = \pm\Theta$ (as before, $\eta(t)$ is a gaussian white noise process). Although very relevant to fit behavioral data, the classical DDM assumes perfect anticorrelation between the integrators, and such degree of correlation seems unrealistic in the brain (Zohary, Shadlen, & Newsome, 1994).

When $\rho = 0$ in equations 2.1, the two integrators become independent, and then the system becomes a race model (see, e.g., Vickers, 1979; Luce, 1986). Therefore, the model defined in equations 2.1 covers a wide spectrum of classical models of decision making, ranging from the classical DDM for $\rho = 1$ and $\nu = -1$, to the race model for $\rho = 0$, as shown in Figure 2. However, the model considered here does not include interactions or connections between the neuronal integrators, as in Usher & McClelland (2001). A discussion about these models has been recently presented in Bogacz (2007).

## 3 Results

**3.1 Independence of the Probability Distributions over the Drift Rates on Decision Boundaries.** A decision about which drift is higher ($\mu_1 > \mu_2$ or $\mu_2 > \mu_1$?) in the system defined in equations 2.1 will depend on the drift estimates at decision time $t$ given the observed trajectories of the integrators up to that time, denoted $\bar{x}_1$ and $\bar{x}_2$, with initial states $x_1(0) = a$ and $x_2(0) = b$ at $t = 0$. The critical quantity is then the probability density of having the drifts $\mu_1$ and $\mu_2$ given the trajectories $\bar{x}_1$ and $\bar{x}_2$ up to time $t$, $p(\mu_1, \mu_2 \mid \bar{x}_1, \bar{x}_2)$. Instead of working directly with this density, it is more convenient to introduce the new variables $u = x_1 + x_2$ and $v = x_1 - x_2$ and work with the probability density $p(\mu_u, \mu_v \mid \bar{u}, \bar{v})$ of having drifts $\mu_u = \mu_1 + \mu_2$ and $\mu_v = \mu_1 - \mu_2$ given the trajectories $\bar{u} = \bar{x}_1 + \bar{x}_2$ and $\bar{v} = \bar{x}_1 - \bar{x}_2$, with initial conditions $u_0 = a + b$ and $v_0 = a - b$. In the new variables, equations 2.1 transform into two independent Wiener processes, with means $\mu_u$ and $\mu_v$, and variances $\sigma_u^2 = 2\sigma^2(1 + \rho v)$ and $\sigma_v^2 = 2\sigma^2(1 - \rho v)$, making the calculations easier.

Let us first consider the case in which there are not boundaries. The case with boundaries will be considered immediately after this discussion. If the time is discretized in $n$ steps of size $\delta t$, such that $t = n\delta t$, and defining $u_i = u(i\delta t)$ ($v_i = v(i\delta t)$), $i = 1, \ldots, n$, then the probability density of having a variable step $\delta u_i = u_i - u_{i-1}$ ($\delta v_i = v_i - v_{i-1}$) at time $t = i\delta t$ follows a gaussian distribution with mean $\mu_u \delta t$ ($\mu_v \delta t$) and variance $\sigma_u^2 \delta t$ ($\sigma_v^2 \delta t$). Since each increment is independent of the others, the probability density of the trajectory $(\bar{u}, \bar{v})$ is

$$p(\bar{u}, \bar{v} \mid \mu_u, \mu_v) = lim_{n\to\infty}(2\pi\sigma_u\sigma_v\delta t)^{-n} \prod_{i=1}^{n} e^{-\frac{(\delta u_i - \mu_u \delta t)^2}{2\sigma_u^2 \delta t} - \frac{(\delta v_i - \mu_v \delta t)^2}{2\sigma_v^2 \delta t}}. \quad (3.1)$$

If $p(\mu_u, \mu_v)$ is the distribution of the drifts based on prior knowledge, Bayes' theorem states that $p(\mu_u, \mu_v \mid \bar{u}, \bar{v}) \propto p(\bar{u}, \bar{v} \mid \mu_u, \mu_v) p(\mu_u, \mu_v)$, where the proportionality is in relation to the drift rates. If the prior distribution is uniform, $p(\mu_u, \mu_v) = p_0$, Bayes' theorem and equation 3.1 give after normalization that the *free* (i.e., without boundaries) distribution over the drift rates is

$$p_{free}(\mu_u, \mu_v \mid \bar{u}, \bar{v}) = p_{free}(\mu_u, \mu_v \mid u - u_0, v - v_0, t) = \\ = \frac{1}{2\pi\sigma_u\sigma_v/t} e^{-\frac{(\mu_u - (u-u_0)/t)^2}{2\sigma_u^2/t} - \frac{(\mu_v - (v-v_0)/t)^2}{2\sigma_v^2/t}}, \quad (3.2)$$

where $(u, v)$ is the ending point of the trajectory $(\bar{u}, \bar{v})$ at time $t$. Then the a posteriori probability distribution over the drift in the variable $u$ ($v$) is a gaussian distribution with mean $(u - u_0)/t$ ($[v - v_0]/t$) and variance $\sigma_u^2/t$ ($\sigma_v^2/t$). Hence, the mean drift is just the slope of the trajectory, computed as the ratio between the distance from initial to ending points, and elapsed time. The variance of the probability distribution decays as a function of time, so a longer time results in a narrower distribution

and therefore in a better estimation of the drifts. Importantly, the identity between $p_{free}(\mu_u, \mu_v \mid \bar{u}, \bar{v})$ and $p_{free}(\mu_u, \mu_v \mid u - u_0, v - v_0, t)$ shows that all the information in relation to the drift rates available in the trajectories is preserved in the distances traveled by the integrators (i.e., their final state) and elapsed time. This is just the description of the fact that the traveled distances and elapsed time are sufficient statistics for the distribution of drift rates.

Now I consider the case in which there are boundaries $\Theta_k(t)$ $(k = 1, 2)$. In this case, only trajectories that do not cross the boundaries at any time before $t$ (i.e., $(\bar{u}, \bar{v})$ such that $u + v < 2\Theta_1(\tau)$ and $u - v < 2\Theta_2(\tau)$ for all $\tau < t$) are realizable, while the trajectories that cross the boundaries before time $t$ are absorbed. In what follows, I show the central result that the probability distribution of the drift rates given the traveled distances by the integrators and the elapsed time when boundaries are present is identical to the probability distribution of the drift rates when there are not boundaries, equation 3.2.

Two steps will be followed in the proof. First, it is shown that the probability density of having the state $(u, v)$ at time $t$ when arbitrary boundaries are present can be factorized in two terms: one that depends on the drift rates but not on the boundaries and another that depends on the boundaries but not on the drift rates. Second, it is shown that this factorization and Bayes' theorem naturally lead to equation 3.2 for uniform a priori distributions over the drift rates. This proof is similar to that derived by Beck et al. (2008) for Poisson-like distributions.

The probability density of observing the state $(u, v)$ of the integrators at time $t$ given initial conditions $(u_0, v_0)$ and drift rates $\mu_u$ and $\mu_v$ can be expressed as the integral of the probability of having the trajectory $(\bar{u}, \bar{v})$, equation 3.1, over all possible trajectories with initial point $(u_0, v_0)$ and final point $(u, v)$, which do not cross the boundaries at any time before $t$,

$$p_{bound}(u, v, t \mid \mu_u, \mu_v) = \int_{\Omega} d(\delta\vec{u})\, d(\delta\vec{v})\, lim_{n\to\infty} \prod_{i=1}^{n} \frac{e^{-\frac{(\delta u_i - \mu_u \delta t)^2}{2\sigma_u^2 \delta t} - \frac{(\delta v_i - \mu_v \delta t)^2}{2\sigma_v^2 \delta t}}}{2\pi\sigma_u\sigma_v\delta t}, \tag{3.3}$$

where $d(\delta\vec{u}) = d(\delta u_1), \ldots, d(\delta u_n)$ and $d(\delta\vec{v}) = d(\delta v_1), \ldots, d(\delta v_n)$ (remember that a trajectory $(\bar{u}, \bar{v})$ is defined in terms of the differential increments $(\delta u_i, \delta v_i)$ for $i = 1, \ldots, n$ in time steps $\delta t = t/n$), and the region $\Omega$ is defined as the set of trajectories $(\bar{u}, \bar{v})$ that satisfies $u_0 + v_0 + \sum_{i=1}^{j}(\delta u_i + \delta v_i) < 2\Theta_1(j\delta t)$ and $u_0 - v_0 + \sum_{i=1}^{j}(\delta u_i - \delta v_i) < 2\Theta_2(j\delta t)$ for $j = 1, \ldots, n$ (i.e., they do not cross the boundaries before time $t$), along with the constraint that $\sum_{i=1}^{n} \delta u_i = u - u_0$ and $\sum_{i=1}^{n} \delta v_i = v - v_0$ (i.e., the initial and final points are $(u_0, v_0)$ and $(u, v)$, respectively). By expanding the squared terms in the

exponents of equation 3.3 and using that $\sum_{i=1}^{n} \delta t = t$, $\sum_{i=1}^{n} \delta u_i = u - u_0$, and $\sum_{i=1}^{n} \delta v_i = v - v_0$, it is possible to extract out from the integral the terms that depend on the drift rates as

$$p_{bound}(u, v, t \mid \mu_u, \mu_v) = e^{-\frac{\mu_u^2 t - 2\mu_u(u-u_0)}{2\sigma_u^2} - \frac{\mu_v^2 t - 2\mu_v(v-v_0)}{2\sigma_v^2}} R(u, v), \tag{3.4}$$

where

$$R(u, v) = \int_{\Omega} d(\delta\vec{u})\, d(\delta\vec{v})\, lim_{n\to\infty} \prod_{i=1}^{n} \frac{e^{-\frac{\delta u_i^2}{2\sigma_u^2 \delta t} - \frac{\delta v_i^2}{2\sigma_v^2 \delta t}}}{2\pi\sigma_u\sigma_v\delta t}. \tag{3.5}$$

Note that $R(u, v)$ depends on the boundaries but not on the drift rates.

Since the quantity of interest is the probability distribution of the drift rates given the final state of the integrators, $p_{bound}(\mu_u, \mu_v \mid u, v, t)$, and this is proportional to equation 3.4 through Bayes' theorem, then only the dependencies of equation 3.4 on the drift rates are explicitly considered. This means, interestingly, that to determine the distribution over the drift rates, it is not required to compute explicitly the complicated integral in equation 3.5, a task achievable only in particular cases. Therefore, after using Bayes' theorem $p_{bound}(\mu_u, \mu_v \mid u, v, t) \propto p_{bound}(u, v, t \mid \mu_u, \mu_v)$ (proportionality is in relation to the drift rates; here a uniform a priori distribution over the drifts rates is assumed, but a similar result holds for any arbitrary a priori distribution) and normalizing the expression, it is found that

$$p_{bound}(\mu_u, \mu_v \mid u - u_0, v - v_0, t) = \frac{1}{2\pi\sigma_u\sigma_v/t}\, e^{-\frac{(\mu_u - (u-u_0)/t)^2}{2\sigma_u^2/t} - \frac{(\mu_v - (v-v_0)/t)^2}{2\sigma_v^2/t}} \tag{3.6}$$

regardless of the boundary conditions, which is identical to equation 3.2. This proves that $p_{bound}(\mu_u, \mu_v \mid u - u_0, v - v_0, t) = p_{free}(\mu_u, \mu_v \mid u - u_0, v - v_0, t)$.

Finally, one needs to show that all information contained in the trajectory $(\bar{u}, \bar{v})$ is preserved in the final state of the integrators and elapsed time in the presence of boundaries. Since for any trajectory that does not cross the bound at any time before $t$, $(\bar{u}, \bar{v})$, one has $p_{bound}(\mu_u, \mu_v \mid \bar{u}, \bar{v}) = p_{free}(\mu_u, \mu_v \mid \bar{u}, \bar{v})$, then using equations 3.2 and 3.6, $p_{bound}(\mu_u, \mu_v \mid u - u_0, v - v_0, t) = p_{bound}(\mu_u, \mu_v \mid \bar{u}, \bar{v})$. Therefore, the integrators accumulate evidence with any leak or loss. This interestingly means that, as in the case without boundaries, the shape of the trajectory does not need to be remembered, since the final state of the integrators and elapsed time contain all necessary information in order to perform inference about the drift

rates. More formally, the distances traveled by the integrators along with the elapsed time are sufficient statistics for the drift rates.

The equivalence between probability distributions with and without boundaries can be understood intuitively as follows. Note, first, that any trajectory that does not cross the boundaries at any time before $t$ and has the same initial and final points leads to the same probability distribution over the drift rates, equation 3.2. Note, second, that the effect of the boundaries is to reduce the number of trajectories that can reach a given ending point from the same initial condition compared to the case with no boundaries. However, since any of the remaining realizable trajectories (i.e., those that do not cross the boundaries) with the same initial and ending points lead to the same probability distribution of the drift rates, the shape of the boundaries does not affect the information available about the drifts at time $t$.

The result that the probability distribution over the drift rates does not depend on the temporal profile of the decision boundaries before the decision time is nevertheless surprising. Let us assume, for instance, that the decision boundaries drop sharply to very negative values at short times and then recover back to some finite value and that the final state of one integrator is positive, $x_1 > 0$. Then one could argue that the a posteriori probability distribution over the drift rate $\mu_1$ of that integrator has to be skewed to very large negative values (i.e., negative drifts are more likely than positive ones) because the trajectory of the integrator first needed to reach very low values before ending at the final state $x_1 > 0$. Then, in general, one would conclude that the probability distribution over the drift rates depends not only on the final state of the integrators, but also on the temporal shape of the boundaries. Equation 3.6 shows that the above reasoning is incorrect and that in fact the probability distribution over the drift $\mu_1$ has a larger probability mass at positive values. The paradox can be resolved by noticing that the traveled distances by the integrators and elapsed time are sufficient statistics for the drift rates $\mu_k$ ($k = 1, 2$) in the decision process, while the shape of the paths traveled provides no additional information about the drift rates. Therefore, the state of the integrators, along with the elapsed time, summarizes all information one needs to know to make a decision that involves any inequality over their drift rates.

One can check that this is true in very simple cases, where analytical solutions are known for the whole probability distribution over the final state of the integrators. Let us consider, for instance, the diffusion of a single integrator, which obeys equation 2.2, and let us assume that there is only a time-dependent boundary at $\Theta(t) = A + Bt$. Then the probability distribution at time $t$ over the state $x$ of the integrator with starting point $x = 0$ and drift rate $\mu$ follows,

$$p(x, t \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \left[ e^{-\frac{(x-\mu t)^2}{2\sigma^2 t}} - e^{-D - \frac{(x-2A-\mu t)^2}{2\sigma^2 t}} \right], \tag{3.7}$$

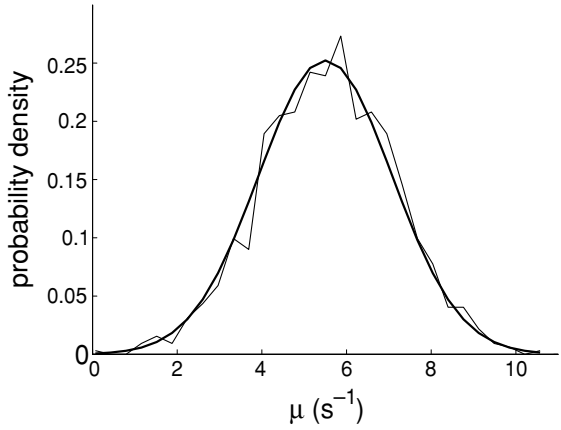


Figure 3: Distribution of the drift rate for a one-variable diffusion process, equation 2.4, with $\sigma^2 = 0.5\ \mathrm{s}^{-1}$ from numerical results (thin line) and from equation 3.8 (thick line) at decision time $t = 200$ ms. The boundary is defined as $\Theta(t) = A + Bt^2$ with $A = 1.3$ and $B = -5\ \mathrm{s}^{-2}$. The mean and variance obtained from the simulations are $5.57\ \mathrm{s}^{-1}$ and $2.50\ \mathrm{s}^{-1}$, respectively, close to the predicted values $5.50\ \mathrm{s}^{-1}$ and $2.50\ \mathrm{s}^{-1}$. Sample trajectories were generated using drift rates uniformly distributed in the range between $-200\ \mathrm{s}^{-1}$ and $200\ \mathrm{s}^{-1}$. The trajectories that reached the threshold $\Theta(t)$ within the time window $t = 200$ ms and $t = 201$ ms were used to compute the distribution of the drift rate.

with $D = 2A(B - \mu)/\sigma^2$. (It is easy to show that $p(x, t \mid \mu)$ vanishes at the time-dependent threshold, and since it is composed of a sum of two free solutions of the associated Fokker-Planck equation, it is the solution with a time-dependent absorbing boundary $\Theta(t)$; Cox & Miller, 1965.) Finally, using Bayes' theorem with uniform a priori distribution over the drift rate, $p(\mu \mid x, t) \propto p(x, t \mid \mu)$, one finds that

$$p(\mu \mid x, t) = \frac{1}{\sqrt{2\pi\sigma^2/t}}\, e^{-\frac{(\mu - x/t)^2}{2\sigma^2/t}}, \tag{3.8}$$

that is, the a posteriori probability distribution over the drift rate $\mu$ is a gaussian with mean $x/t$ and variance $\sigma^2/t$, and it does not depend on the parameters $A$ and $B$ that define the time-dependent boundary. (Note that equation 3.8 is a particular case of equation 3.6 when restricted to the dynamics of a single integrator.) The same result holds when there are two side time-independent boundaries, a case for which analytical expressions are also known (Cox & Miller, 1965), and in fact, as it has been shown before, equation 3.8 is valid for any decision boundaries.

Figure 3 shows the probability density function over the drift rate $\mu$ computed from numerical simulations of a single integrator that obeys

equation 2.2 in the presence of a nonlinear boundary $\Theta(t)$, a case for which there are no generally valid analytical solutions. Sample trajectories with a broad distribution of drift rates were generated, and from the trajectories that ended at threshold at times within a short window $t$ and $t + \Delta t$, the distribution of drift rates at decision time was computed. As the theory predicts, equation 3.8 should be valid in the presence of any threshold, and therefore the distribution of drift rates should be a gaussian with mean $\Theta(t)/t$ and variance $\sigma^2/t$ when the decision is made at time $t$ regardless of the shape of the boundary. The simulations are in perfect agreement with the theory.

A special feature of the DDM considered in equations 2.1 is that all information is readily available in the final state of the integrators and the elapsed time, equation 3.6. Indeed, this property is violated in neuronal integrator models with leak terms (negative linear $x_k$s), as considered, for example, in Usher and McClelland (2001). It can be shown that in these models, the a posteriori probability density distribution over the drift rates depends not only on the final states of the integrators but also on the mean value of the trajectories. Furthermore, if expressed only as a function of the final state of the integrators (along with the initial state and elapsed time), the probability density over the drift rate depends on the temporal profile of the decision boundaries $\Theta_k(t)$. It is interesting to note, then, that brain areas like LIP and FEF show ramping activity that resembles perfect, leakless integration of evidence.

**3.2 Decision Confidence.** Upon a decision, confidence can be expressed as the estimated probability that the highest signal has been correctly detected given the observed trajectories of the integrators up to decision time $t$. Given the final state of the integrators, the join probability density of the drift rates is given by equation 3.6. From this probability density, it is easy to compute whether the probability of $\mu_1 > \mu_2$ is larger or lower than the probability of $\mu_2 > \mu_1$, and the choice $\mu_1 > \mu_2$ or $\mu_2 > \mu_1$ will be made accordingly. In order to associate uniquely one decision boundary with one choice, the initial states of the integrators and the bounds are taken to be identical for the two integrators, $a = b$ and $\Theta(t) = \Theta_1(t) = \Theta_2(t)$. Therefore, if the boundary of integrator $k = 1$ is reached first, then at decision time, $x_1(t) = \Theta(t) > x_2(t)$. This implies that the case $\mu_1 > \mu_2$ has higher probability, and hence it is chosen (and, conversely, if integrator $k = 2$ reaches its bound first). Without loss of generality, in the following, integrator $k = 1$ is assumed to reach its boundary first, $x_1 = \Theta(t)$, and then the choice $\mu_1 > \mu_2$ has been made. Decision confidence can then be written using equation 3.6 as

$$p(\mu_1 > \mu_2 \mid x_2, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\Delta(x_2)}{\sigma_v\sqrt{t}}} dz\, e^{-z^2/2}, \tag{3.9}$$

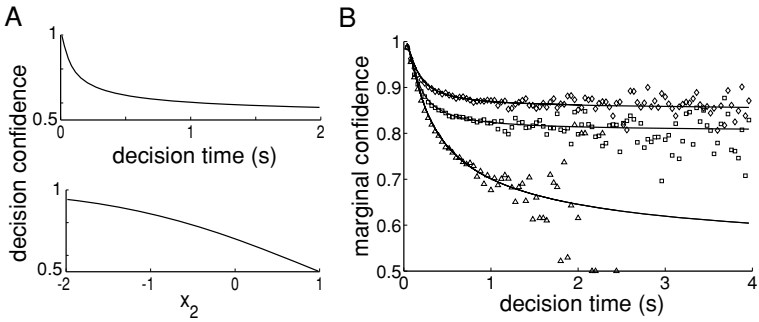


Figure 4: (A) Confidence for the decision $\mu_1 > \mu_2$ as a function of decision time for $x_2 = 0.5$ and as a function of $x_2$ for $t = 1$ s for independent integrators using equation 3.9. (B) Marginal decision confidence as a function of decision time for independent (top, equations 3.20 and 3.22), half-anticorrelated (middle, equations 3.20 and 3.21), and completely anticorrelated (classical DDM; bottom, equation 3.11) integrators. Lines correspond to analytical expressions and data points to simulations. For all cases, $\sigma^2 = 3.8\,\mathrm{s}^{-1}$, $\Theta_1 = \Theta_2 = 1$ and $a = b = 0$.

where $\Delta(x_2) = \Theta(t) - x_2$ and $x_2$ is the final state of integrator 2 at decision time $t$. Equation 3.9 is a cumulative normal gaussian with the argument being proportional to the ratio between the difference in distance traveled by the two integrators and the squared root of the decision time. The same expression holds if the prior distribution of the drifts is $p(\mu_1, \mu_2) = p(\mu_1)\delta(\mu_1 - \mu_2)$, where $p(\mu_1) = p_0$ ($p_0$ is a constant), which states that $\mu_1$ is distributed uniformly along with the constraint $\mu_1 = -\mu_2$. Decision confidence depends, then, on the difference between the states of the two integrators at decision time. The correlation between the integrators affect only $\sigma_v$, and hence the slope of the confidence function, but it does not change the functional dependence on the states of the integrators and elapsed time.

Figure 4A shows decision confidence as a function of decision time for a fixed state of the losing integrator and as a function of the state of the losing integrator, $x_2$, for fixed decision time. Constant bounds, $\Theta(t) = \Theta$, for the integrators are used in this example. As a function of decision time, confidence decays monotonically from one to one-half. In the limit of long $t$, equation 3.9 converges to $1/2 + (\Theta - x_2)/\sqrt{2\pi\sigma_v^2 t}$. The monotonic decay is due to the fact that fast responses are likely to happen in trials in which large drift rates are used, which leads to better performance and, hence, increased confidence. As a function of the state of the losing integrator, decision confidence decreases from one at large and negative $x_2$ and approaches one-half when $x_2$ is close to threshold.

It is possible to compute decision confidence for the classical DDM employing the general expression in equation 3.9. Here I use the prior

distribution $p(\mu_1, \mu_2) = p(\mu_1)\delta(\mu_1 - \mu_2)$ with $p(\mu_1) = p_0$ ($p_0$ is a constant), which ensures the constraint $\mu = \mu_1 = -\mu_2$; in addition, constant bounds $\Theta(t) = \Theta$ are used. Then the state of one integrator is determined by the state of the remaining integrator, since when integrator $k = 1$ reaches its threshold, $x_1 = \Theta$, integrator $k = 2$ is at $x_2 = -\Theta$. At decision time, the probability distribution over the drift rate $\mu$ can be written using equation 3.6 as

$$p(\mu \mid t) = \frac{1}{\sqrt{2\pi\sigma^2/t}}\, e^{-\frac{(\mu-\Theta/t)^2}{2\sigma^2/t}}, \tag{3.10}$$

and from this equation or from the general expression in equation 3.9 with $\sigma_v = 2\sigma$, one gets that decision confidence is

$$p(\mu > 0 \mid t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Theta/\sigma\sqrt{t}} dz\, e^{-z^2/2}, \tag{3.11}$$

which depends on only decision time (the dependence on the threshold and noise variance is not made explicit). For long decision times, decision confidence converges to $1/2 + \Theta/\sqrt{2\pi\sigma^2 t}$. Experimental measurements of decision confidence have been typically fit by $a/t + b$ functions (Vickers, 1979), which are mathematically different from the $a'/\sqrt{t} + b'$ dependences reported here. However, the two expressions are not very different numerically one from each other when they are used in a limited range of decision times. It should be noted that distinguishing between the two expressions requires large amounts of data for very long reaction times, which are not typically available. Nevertheless, it would be desirable to design experiments that allow the estimation of confidence for a wide range of decision times to test the analytical expressions found in this letter.

The classical DDM can be extended to the case in which the boundaries are time dependent, $\Theta(t)$. In this case, the probability distribution over the drift rate $\mu$ at decision time is written using equation 3.6 as

$$p(\mu \mid t) = \frac{1}{\sqrt{2\pi\sigma^2/t}}\, e^{-\frac{(\mu-\Theta(t)/t)^2}{2\sigma^2/t}}, \tag{3.12}$$

and decision confidence then becomes

$$p(\mu > 0 \mid t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Theta(t)/\sigma\sqrt{t}} dz\, e^{-z^2/2}. \tag{3.13}$$

Considering again the general DDM model, equation 2.1, if the drift rates are not uniformly distributed but can take only two possible values

$\mu_1 = -\mu_2 = \pm\mu_0$ with probability one-half,

$$p(\mu_1, \mu_2) = \frac{1}{2}\,\delta(\mu_1 - \mu_0)\,\delta(\mu_2 + \mu_0) + \frac{1}{2}\,\delta(\mu_1 + \mu_0)\,\delta(\mu_2 - \mu_0), \tag{3.14}$$

then following similar steps as in equations 3.6 and 3.9, it can be shown that

$$p(\mu_1 > \mu_2 \mid x_2) = \frac{1}{1 + e^{-4\mu_0\Delta(x_2)/\sigma_v^2}}. \tag{3.15}$$

At decision time, decision confidence depends on the difference between the traveled distances by the integrators, $\Delta(x_2) = \Theta(t) - x_2$, but not explicitly on decision time. As in equations 3.6 and 3.9, confidence is independent of the boundaries used to generate decisions. For the classical DDM with time-dependent boundaries, $\Theta(t)$, decision confidence becomes

$$p(\mu_1 > \mu_2 \mid t) = \frac{1}{1 + e^{-2\mu_0\Theta(t)/\sigma^2}}, \tag{3.16}$$

which depends on only decision time through the boundary time dependence. It is interesting to note that for this case, decision confidence equals the expression for the probability of reaching first the correct boundary (Cox & Miller, 1965). Therefore, confidence and performance are identical here. Although this result can be shown to be rather general, it is not true for all boundary conditions.

Experimentally, human verbal reports of confidence are close but typically lower than the actual performance shown by the subjects (Vickers, 1979); in general, the slope of the confidence versus difficulty curve is lower than that of the probability correct versus difficulty curve. One could argue that verbal reports provide biased estimates of the true internal confidence because of their discretized nature. A second, and more interesting, possibility is that subjects' belief does not fully correspond to their performance because the subjects have not completely learned the task's contingencies. In either case, confidence would not correspond to the probability of being correct. The last situation can be described with the DDM presented here if, for instance, the variance $\sigma$ in equation 2.1 is only approximately known by the subject and is taken to be larger than the real one. For the problem corresponding to equation 3.16, probability correct and confidence will be correlated and have the same functional dependences on $\mu_0$ (which measures the difficulty of each trial), but the slope of the performance versus difficulty curve will be higher than that of the confidence-versus-difficulty curve.

**3.3 Marginal Confidence.** Decision confidence depends in general on both the decision time and the state of the losing integrator. However, for comparison with behavioral data, it is necessary to know the marginal decision confidence where the state of the losing integrator has been integrated out, since the latter is not typically available. Marginalization requires knowing the probability density of $x_k$ ($k = 1, 2$) at decision time, which satisfies a Fokker-Planck equation (FPE) with absorbing boundaries. The FPE associated with equations 2.1 is

$$\frac{\partial}{\partial t}P = \left[-\mu_1\frac{\partial}{\partial x_1} - \mu_2\frac{\partial}{\partial x_2} + \frac{1}{2}\sigma^2\left(\frac{\partial^2}{\partial^2 x_1} + 2\rho\nu\frac{\partial^2}{\partial x_1\partial x_2} + \frac{\partial^2}{\partial^2 x_2}\right)\right]P\ , \tag{3.17}$$

where $P \equiv P(\vec{x}, t \mid \vec{x}_0)$ is the joint probability density of finding the integrators in state $\vec{x} = (x_1, x_2)$ at time $t$ given that the initial condition at time $t = 0$ was $\vec{x}_0 = (a, a)$. The FPE has to be solved along with the condition that at the boundaries, defined as the semi-lines $x_1 = \Theta(t)$ for all $x_2 \leq \Theta(t)$ ($l_1$) and $x_2 = \Theta(t)$ for all $x_1 \leq \Theta(t)$ ($l_2$), $P$ has to be zero at all times. This condition defines the semi-lines $l_1$ and $l_2$ as absorbing boundaries. Solutions to the FPE, equation 3.17, are known for the independent and completely anticorrelated (classical DDM) cases with time-independent boundaries (Cox & Miller, 1965). In the appendix, I provide details for the derivation of the solution for two half-anticorrelated integrators with time-independent boundaries. The above three solutions allow a direct comparison of marginal confidence between different cases. In the following, $\Theta = 0$ and $a < 0$. Then $|a|$ measures the initial distance to threshold.

In order to determine the marginal confidence (one needs to derive the joint probability density that integrator $k$ reaches its threshold at time $t$ and integrator $j \neq k$ is at state $x_j$ ($x_j \leq 0$) at that time, along with the condition that no integrator has crossed its threshold before, given the initial condition $\vec{x}_0$. This quantity, denoted $g_k(t, x_j \mid \vec{x}_0, \vec{\mu})$, is written as

$$g_k(t, x_j \mid \vec{x}_0, \vec{\mu}) = -\frac{1}{2}\sigma^2\frac{\partial}{\partial x_k}P(\vec{x}, t \mid \vec{x}_0)|_{x_k=0}, \tag{3.18}$$

which can be obtained from the probability flux vector expression associated with the FPE (see equation 3.17; Risken, 1989). The dependence on the drift rates $\vec{\mu} \equiv (\mu_1, \mu_2)$ has been made explicit for later convenience. Its integral over $x_j$ is the conditional first passage times distribution for integrator $k$. For instance, using equation A.9 in the appendix, for two independent integrators equation 3.18 takes the expression

$$g_1(t, x_2 \mid \vec{x}_0, \vec{\mu}) = \frac{|a|}{2\pi\sigma^2 t^2}\, e^{-\frac{(a+\mu_1 t)^2}{2\sigma^2 t}}\left(e^{-\frac{(x_2-a-\mu_2 t)^2}{2\sigma^2 t}} - e^{-\frac{2b\mu_2}{\sigma^2} - \frac{(x_2+a-\mu_2 t)^2}{2\sigma^2 t}}\right), \tag{3.19}$$

and similarly for $g_2(t, x_1 \mid \vec{x}_0, \vec{\mu})$.

Marginal confidence is defined as the confidence in a decision made at time $t$ when the state of the losing integrator and the values of the drift rates are unknown or, more formally,

$$p(\mu_1 > \mu_2 \mid t) = \int dx_2 \; p(\mu_1 > \mu_2 \mid x_2, t) \; P(x_2, t \mid \vec{x}_0), \tag{3.20}$$

where decision confidence $p(\mu_1 > \mu_2 \mid x_2, t)$ is given in equation 3.9 and $P(x_2, t \mid \vec{x}_0)$ is the probability distribution of $x_2$ at time $t$ given that the first integrator has reached threshold at time $t$ averaged over the distribution of drift rates. The latter is defined as $P(x_2, t \mid \vec{x}_0) = C(t)\, g_1(t, x_2 \mid \vec{x}_0)$, where $C(t)$ is a normalization constant; $g_1(t, x_2 \mid \vec{x}_0) = \int d\vec{\mu}\; g_1(t, x_2 \mid \vec{x}_0, \vec{\mu})\; p(\vec{\mu})$, which includes the average over the distribution of drift rates; and $p(\vec{\mu})$ is the a priori distribution over the drift rates. One could also define marginal confidence for a fixed set of values of drift rates, as they can be controlled experimentally, but here I focus on the averaged quantity. For two half-anticorrelated integrators, using a uniform prior along with the constraint that $\mu_1 = -\mu_2$, and equations A.8 and 3.16, it is found that the probability density of finding the second integrator at state $x_2$ given that the first integrator has reached its boundary at time $t$ takes the expression

$$P(x_2, t \mid \vec{x}_0) \propto e^{-(x_2 - 2a)^2/2\sigma^2 t} + e^{-(x_2 + 2a)^2/2\sigma^2 t} - 2\, e^{-(x_2^2 + 4a^2)/2\sigma^2 t}, \tag{3.21}$$

where the proportionality is in relation to the variable $x_2$. For two independent integrators, using equations A.9 and 3.17, with the constraint that $\mu_1 = -\mu_2$, it is obtained

$$P(x_2, t \mid \vec{x}_0) \propto e^{-x_2^2/4\sigma^2 t}(e^{a x_2/\sigma^2 t} - e^{-a x_2/\sigma^2 t}). \tag{3.22}$$

Figure 4B shows that marginal confidence (computed using equations 3.20 to 3.22, for two independent and half-anticorrelated integrators and using equation 3.11 for the classical DDM with the same threshold and noise values in the three cases) is at a maximum for the whole range of decision times when the integrators are independent, nearly so when they are half-anticorrelated, and at a minimum for the classical DDM. Therefore, decision confidence in integrators with even strong correlations is close to that in independent integrators and higher than that in the classical DDM. The large difference in marginal confidence between the classical DDM and the half-anticorrelated or independent integrators is due to the fact that as decision time increases, the losing integrator in the latter two cases separates from its boundary (a reminder that the losing integrator has negative drift). Using the fact that confidence increases as the distance between the integrators increases, equation 3.9, it is now clear why marginal confidence

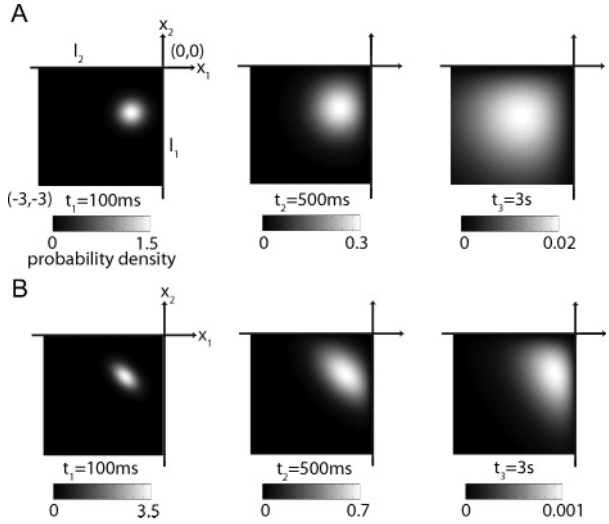


Figure 5: Time evolution of the probability density distribution over $\vec{x}$ for two integrators with independent (A) and half-anticorrelated (B) white noise inputs. The gray-coded distributions are plotted at three different times, using the solutions in equations A.9 and A.8. Parameters for the independent integrators ($\rho = 0$) are $a = -1.5$, $b = -1$, $\mu_1 = 0.2\,\mathrm{s}^{-1}$, $\mu_2 = 0.1\,\mathrm{s}^{-1}$, and $\sigma^2 = 1\,\mathrm{s}^{-1}$. Parameters for the half-anticorrelated integrators ($\rho = 1/2$, $\nu = -1$) are $a = b = -1$, $\mu_1 = 0.75\,\mathrm{s}^{-1}$, $\mu_2 = 0\,\mathrm{s}^{-1}$, and $\sigma^2 = 0.5\,\mathrm{s}^{-1}$. Besides, $\Theta_1 = \Theta_2 = 0$.

becomes higher for the independent and half-anticorrelated cases than for the classical DDM.

**3.4 Reaction Time Distributions.** The time evolution of the probability density function of finding the neuronal integrators in a given state can be used to gain additional insight into the decision process. Figure 5 shows three snapshots of the time evolution of the probability density distribution over $\vec{x}$ for two independent (see Figure 5A) and two half-anticorrelated integrators (see Figure 5B). In both cases, I use $\mu_1 > \mu_2 \geq 0$. The distribution broadens over time and moves toward the boundaries. When the distributions get close to the boundaries, their peaks remain approximately stationary. For the parameter values used, the distributions approach the boundary $l_1$ more rapidly. Therefore, most of the trajectories will escape through the semi-line $l_1$. However, many others will cross $l_2$ first. It can be proved that one of the two boundaries is crossed at some finite time with probability one if the drift of at least one integrator is larger than zero. The distribution spreads circularly for two independent integrators, while for

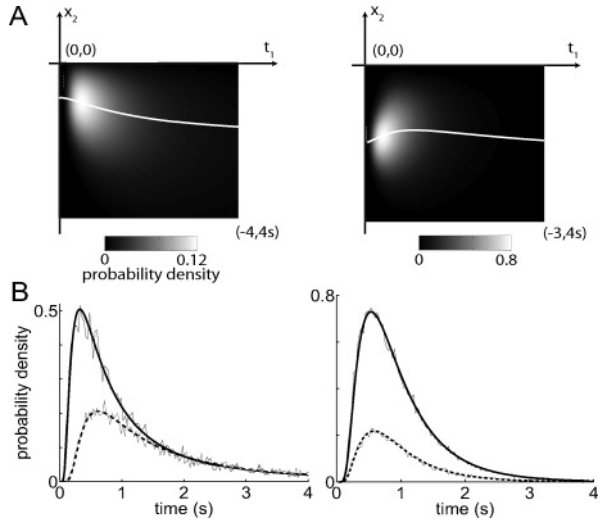


Figure 6: (A). First passage times-state joint probability distribution, $g_1(t, x_2 \mid \vec{x}_0)$, for two independent (left) and two half-anticorrelated (right) integrators (see equation 3.19, and equations 3.18 and A.8, respectively). White lines are the mean value of $x_2$ as a function of decision time. (B) First passage times probability density distributions, $g_k(t)$, $k = 1, 2$, for independent (left) and half-anticorrelated (right) integrators. Full and dashed lines correspond to the integrator with higher and lower drift, respectively. Fluctuating lines correspond to numerical simulations of equations 2.1. Parameters are as in Figure 5.

two half-anticorrelated integrators, the distribution spreads more along the nonprincipal diagonal.

The first passage times-state joint probability distribution, equation 3.18, can be used to describe the state of the losing integrator given that a decision has been made. Figure 6A shows a typical example of the joint probability distribution. The mean value of $x_2$ (the white line) diverges monotonically from threshold as a function of the crossing time of integrator $k = 1$. This is because, as shown in Figure 5A, trajectories tend to disperse at long times. A different picture emerges when the two integrators are half-anticorrelated, equations 3.18 and A.8, as shown in the right panel. The mean value of $x_2$ approaches threshold at short decision times, while it diverges from threshold at long decision times. The first rise is due to the presence of negative correlations in the inputs to the two integrators and can be understood as follows. If integrator $k = 1$ reaches its threshold at a late time, this delay is likely due to a negative input fluctuation to that integrator. Since the two integrators have anticorrelated inputs, the integrator $k = 2$ has likely experienced a positive input fluctuation, explaining why the mean value

of $x_2$ approaches the threshold at short decision times. At long decision times, the diffusion of trajectories due to noise dominates over the previous mechanism, and therefore the mean value of $x_2$ diverges from threshold in that regime.

Figure 6B shows the probability distributions of first passage times for each integrator along with the condition that it crossed its threshold first. This first passage time distribution is computed by integrating equation 3.18 with the probability distribution of the losing integrator. The integrator with the higher drift crosses first the threshold with a probability larger than one-half (full lines). The distributions are positively skewed, similar to the reaction times distributions typically found in two-alternative forced-choice tasks (Ratcliff & Smith, 2004; Luce, 1986).

## 4 Discussion

A decision diffusion model (DDM) consisting of two neuronal integrators receiving independent as well as correlated inputs has been studied. Decision confidence has been shown to depend on the difference between the traveled distances of the two integrators at decision time. Importantly, leakless integrators accumulate evidence without loss, since their final state and elapsed time fully specify the probability distribution over the signals, regardless of the stopping rule used to make decisions. Different behaviors are obtained from leak integrators, since the probability distribution over the drift rates depends not only on the final states of the integrators and elapsed time. These results can have important consequences about neuronal coding and integration of information in lateral intraparietal area (LIP) and frontal eye fields (FEF), where almost linear, perfect integration of evidence is observed.

Previous models have proposed that decision confidence is encoded in the difference between the state of two integrators that accumulate evidence in favor of two alternatives (Vickers, 1979; Juslin & Olsson, 1997; Vickers & Lee, 1998; Van Zandt, 2000), like the DDM studied here. However, it was not known how confidence was mathematically related to the difference of the states (balance of evidence) of the integrators. As equation 3.9 shows, decision confidence is a cumulative gaussian whose argument is proportional to the balance-of-evidence term. In addition, it explicitly shows time dependences that were not anticipated by these preliminary works. Pleskac and Busemeyer (2007) have considered a model in which confidence is computed after the decision has been made. However, an interesting alternative, as proposed by the model presented here, is that that when a decision is made, there is immediate access to decision confidence and the evaluation of confidence does not require additional deliberation. Ratcliff and Starns (2009) recently presented a model about how verbal reports of decision confidence can be generated by human subjects. The model requires multiple additional integrators (one for each confidence rating—for example, high,

medium, or low), whose role is to determine which rating will be reported. The DDM model that I have described here makes the interesting prediction that decision confidence is represented in the brain as a continuous value: if verbal reports about confidence are not forced, subjects might use confidence only implicitly without the necessity of discretizing it.

Biophysically inspired neuronal network models for decision making have been studied in Wang (2002) and Beck et al. (2008). These attractor models show almost linear ramping activity at low firing rates, like the DDM studied here, but they display nonlinearities that become evident when the firing rate approaches the decision boundary. Consistent with these models, Huk and Shadlen (2005) and Wong, Huk, Shadlen, and Wang (2007) have shown that brief motion pulses (100 ms) injected at different times from the onset of a dynamic random dot display affect the firing rate of LIP neurons in a time-dependent manner, which earlier pulses producing larger changes in firing rate than later pulses. As Huk and Shadlen (2005) pointed out, this result is also consistent with a perfect integrator model of inputs with time-decaying boundaries, like the DDM considered in this letter. Further experiments are required to test which is the correct model. These experiments would likely involve injecting motion pulses at different times conditioned to the same firing rate of the LIP neurons for all times. In this hypothetical experiment, the attractor model predicts time-shift invariance of the induced changes in firing rate, while the DDM with time-decaying boundaries predicts its violation. Further theoretical work is also required to connect realistic neuronal attractor networks with simplified integrator models and to determine to what extent nonlinearities that arise near the decision boundaries can be well approximated by hard decision boundaries (see Roxin & Ledberg, 2008).

Decision confidence has been shown to display two apparently contradictory behaviors (Vickers, 1979). First, confidence decays as a function of decision time when subjects are free to choose the stopping time (Henmon, 1911). In contrast, confidence increases as a function of decision time when subjects are externally forced to make a decision at an experimentally controlled time (Vickers, 1979). The first situation corresponds directly to the case that I have considered in Figure 4B: since trials for which the difference between the drift rates is large lead typically to short decision times, confidence will be higher at shorter than at longer decision times. The second situation can also be described with the DDM introduced here. This can be seen by noting that in this case, equation 3.9 becomes

$$p(\mu_1 > \mu_2 \mid x_1, x_2, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\Delta(x_1,x_2)}{\sigma_v\sqrt{t}}} dz\, e^{-z^2/2}, \tag{4.1}$$

where $\Delta(x_1, x_2) = x_1 - x_2$, $x_1$, and $x_2$ are the states of the integrators at time $t$, and it is assumed that the two integrators are allowed to evolve freely until the stopping time is externally forced (i.e., bounds are not present, as

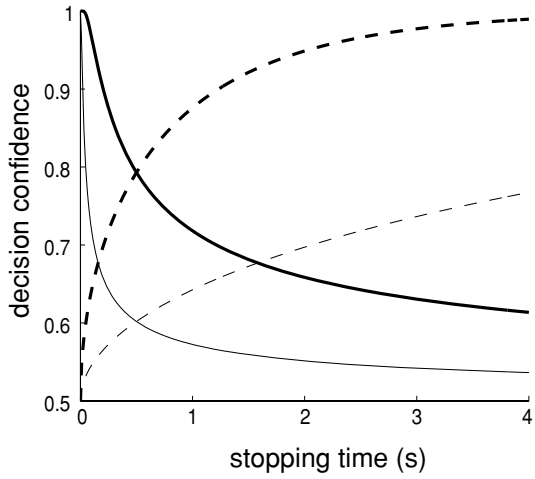


Figure 7: Decision confidence decreases with decision time when the stopping time is internally determined by a threshold crossing (continuous lines) but increases with time if the decision time is externally forced and there is no threshold (dashed lines). In both cases, the classical DDM is considered (i.e., a DDM with completely anticorrelated fluctuations), with bounds at $\Theta = 1$ when the decision time is internally generated and no bounds when the stopping time is externally forced (for the the latter case, $\mu_0 = 2\,\mathrm{s}^{-1}$). The noise variance is $\sigma^2 = 3\,\mathrm{s}^{-1}$ and $30\,\mathrm{s}^{-1}$ for the thick and thin lines, respectively.

in Vickers, 1979, and the state of the integrators is read out at the forced stopping time). Since the difference between the states of the integrators grows in mean linearly with time and proportionally to the difference in their drift rates, that is, $\langle \Delta(x_1, x_2) \rangle = (\mu_1 - \mu_2)t$, the argument of equation 4.1 grows in mean as a function of time, and therefore confidence is expected to increase. Figure 7 shows that decision confidence decreases as a function of decision time for the classical DDM when decision time is internally generated, equation 3.11, and that decision confidence in the same model without boundaries for any given drift rate $\mu_v = \mu_1 - \mu_2 = 2\mu_0 > 0$, approximated as

$$p(\mu_1 > \mu_2 \mid t, \mu_0) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\mu_0 t}{\sigma\sqrt{t}}} dz\, e^{-z^2/2}, \tag{4.2}$$

increases as a function of the externally forced stopping time. As expected, the curves move downward with increasing input noise $\sigma$, since the difficulty of the task increases with it. Mathematically, this is because both equation 3.11 and equation 4.2 are decreasing functions of $\sigma$.

In summary, I have provided analytical tools to describe decision confidence and uncertainty in decision making using the DDM. The

analytical expressions for decision confidence in partially correlated neuronal integrators found in this letter might allow estimating confidence from multirecording neurophysiological data in decision areas like LIP and FEF (Hanes & Schall, 1996; Shadlen & Newsome, 1996; Platt & Glimcher, 1999) and compare it with behaviorally measured confidence values. The results can be extended to multiple integrators and to more general forms of noise, therefore opening a window for studying how decision uncertainty is represented in neuronal populations in the brain.

## Appendix: Solution for a DDM with Half-Anticorrelated Integrators

This appendix provides details for the derivation of the solution of the FPE, equation 3.17, for the case of half-anticorrelated integrators. It also gives the derivation for the case in which the integrators are independent.

The strategy is to find the solution of the FPE, equation 3.17, with boundary conditions as specified in the main text as a linear superposition of free solutions,

$$P_e(\vec{x}, t \mid \vec{x}') = \frac{1}{\pi \sigma_u \sigma_v t}\, e^{-\frac{(x_1+x_2-x_1'-x_2'-\mu_u t)^2}{2\sigma_u^2 t} - \frac{(x_1-x_2-x_1'+x_2'-\mu_v t)^2}{2\sigma_v^2 t}}, \tag{A.1}$$

with initial condition $\vec{x}' = (x_1', x_2')$, using the well-known method of the images (Cox & Miller, 1965; Risken, 1989). Note that a free solution is a bivariate gaussian with independent variables $u = x_1 + x_2$ and $v = x_1 - x_2$. The free solutions satisfy the FPE, equation 3.17, for any initial condition. However, none of them satisfies the vanishing boundary condition at the semi-lines $l_1$ and $l_2$.

As a preliminary step, the FPE, equation 3.17, with absorbing boundary on the line $x_1 = 0$ is solved. I seek a solution with the form

$$P(\vec{x}, t \mid \vec{x}_0) = P_e(\vec{x}, t \mid \vec{x}_0) + r\ P_e(\vec{x}, t \mid \vec{x}_1), \tag{A.2}$$

consisting of a source at the initial condition $\vec{x}_0 = (a, b)$ (here it is not assumed that $a = b$, as in section 3.3) and an image source at $\vec{x}_1 = (a_1, b_1)$ weighted with the coefficient $r$. By imposing the boundary condition $P(0, x_2) = 0$ for all $x_2$ (i.e., the line $x_1 = 0$ is an absorbing boundary), I find that the image coordinates are $\vec{x}_1 = (-a, \alpha a + b)$ with $\alpha = 2(1 - \sigma_u^2/\sigma_v^2)/(1 + \sigma_u^2/\sigma_v^2)$, and

$$r = -e^{\mu_u(u_1-u_0)/\sigma_u^2 + \mu_v(v_1-v_0)/\sigma_v^2}, \tag{A.3}$$

where $u_0 = a + b$, $v_0 = a - b$, $u_1 = a_1 + b_1$ and $v_1 = a_1 - b_1$. The solution satisfies the initial condition, $P(\vec{x}, 0 \mid \vec{x}_0) = \delta(\vec{x} - \vec{x}_0)$. If the absorbing boundary is on the line $x_2 = 0$, the solution instead is

$$P(\vec{x}, t \mid \vec{x}_0) = P_e(\vec{x}, t \mid \vec{x}_0) + s\ P_e(\vec{x}, t \mid \vec{x}_2), \tag{A.4}$$

with the image at $\vec{x}_2 = (a_2, b_2) = (a + \alpha b, -b)$ and

$$s = -e^{\mu_u (u_2 - u_0)/\sigma_u^2 + \mu_v (v_2 - v_0)/\sigma_v^2}, \tag{A.5}$$

where $u_2 = a_2 + b_2$ and $v_2 = a_2 - b_2$. In both cases, the image is located on the ellipse,

$$\sigma_v^2 \; u_0^2 + \sigma_u^2 \; v_0^2 = \sigma_v^2 \; u_i^2 + \sigma_u^2 \; v_i^2, \tag{A.6}$$

$i = 1, 2$, which includes the source at the initial condition. It is convenient to introduce the matrices

$$\Pi_1(\alpha) = \begin{pmatrix} -1 & 0 \\ \alpha & 1 \end{pmatrix}, \quad \Pi_2(\alpha) = \begin{pmatrix} 1 & \alpha \\ 0 & -1 \end{pmatrix},$$

which define a mapping from the initial condition to the image positions through the lines $x_1 = 0$ and $x_2 = 0$, respectively, $\vec{x}_i = \Pi_i(\alpha)\vec{x}_0$, $i = 1, 2$. Repeated and alternated applications of $\Pi_1$ and $\Pi_2$ to the initial condition lead to points that lie on the same ellipse, equation A.6.

A combination of the two previous images with the source at the initial condition,

$$P(\vec{x}, t \mid \vec{x}_0) = P_e(\vec{x}, t \mid \vec{x}_0) + c_1 \; P_e(\vec{x}, t \mid \vec{x}_1) + c_2 \; P_e(\vec{x}, t \mid \vec{x}_2), \tag{A.7}$$

is not a solution to the FPE, equation 3.17, with absorbing boundaries on the semi-lines $l_1$ and $l_2$, for any values of $c_1$ and $c_2$. This is because the source at the initial condition plus one of the images can vanish on one semi-line, but not simultaneously on the other, implying the necessity of additional images. One possibility is to augment the number of images to try to match the boundary condition at the two semi-lines simultaneously. In fact, I find that a finite number of images is enough for a DDM with two half-anticorrelated integrators, corresponding to ($\rho = 1/2$ and $\nu = -1$ in equation 2.1). The solution consists of five images,

$$P(\vec{x}, t \mid \vec{x}_0) = P_e(\vec{x}, t \mid \vec{x}_0) + \sum_{i=1}^{5} c_i \; P_e(\vec{x}, t \mid \vec{x}_i), \tag{A.8}$$

with $\sigma_u^2 = \sigma^2$ and $\sigma_v^2 = 3\sigma^2$; positions $\vec{x}_1 = (-a, a + b)$, $\vec{x}_2 = (a + b, -b)$, $\vec{x}_3 = (b, -a - b)$, $\vec{x}_4 = (-a - b, a)$, and $\vec{x}_5 = (-b, -a)$; and coefficients $c_1 = -e^{-2a\mu_1/\sigma^2}$, $c_2 = -e^{-2b\mu_2/\sigma^2}$, $c_3 = e^{-[2a\mu_1 + 2(a+b)\mu_2]/\sigma^2}$, $c_4 = e^{-[2(a+b)\mu_1 + 2b\mu_2]/\sigma^2}$, and $c_5 = -e^{-[2(a+b)(\mu_1+\mu_2)]/\sigma^2}$. The solution is schematically represented in Figure 1D. The images are located on an ellipse, equation A.6. Their positions are obtained by applying iteratively and alternatively the matrices

$\Pi_i(\alpha)$, $i = 1, 2$, with $\alpha = 1$ to the initial condition. After defining $\hat{\Pi}_i = \Pi_i(1)$, $i = 1, 2$, we have $\vec{x}_1 = \hat{\Pi}_1\vec{x}_0$, $\vec{x}_2 = \hat{\Pi}_2\vec{x}_0$, $\vec{x}_3 = \hat{\Pi}_2\hat{\Pi}_1\vec{x}_0$, $\vec{x}_4 = \hat{\Pi}_1\hat{\Pi}_2\vec{x}_0$, and $\vec{x}_5 = \hat{\Pi}_1\hat{\Pi}_2\hat{\Pi}_1\vec{x}_0 = \hat{\Pi}_2\hat{\Pi}_1\hat{\Pi}_2\vec{x}_0$. Images 1 and 2 are the images of the source at the initial condition through lines $x_1 = 0$ and $x_2 = 0$, respectively. Images 3 and 4 can be thought of as the images of the first and second images through the lines $x_2 = 0$ and $x_1 = 0$, respectively, and image 5 as the image of images 3 and 4 through the lines $x_1 = 0$ and $x_2 = 0$. The coefficients have been found by using the previous results for single line boundaries, which allow imposing for each pair of image and its image through one semi-line that the boundary condition is met on that semi-line.

A finite number of free solutions suffices too when the white noise inputs are independent across the two integrators, corresponding to $\rho = 0$ in equations 2.1. The solution consists of three images,

$$P(\vec{x}, t \mid \vec{x}_0) = P_e(\vec{x}, t \mid \vec{x}_0) + \sum_{i=1}^{3} c_i \; P_e(\vec{x}, t \mid \vec{x}_i), \tag{A.9}$$

with $\sigma_u^2 = \sigma_v^2 = 2\sigma^2$; positions $\vec{x}_1 = (-a\,, b)$, $\vec{x}_2 = (a\,, -b)$, and $\vec{x}_3 = (-a\,, -b)$; and coefficients $c_1 = -e^{-2a\mu_1/\sigma^2}$, $c_2 = -e^{-2b\mu_2/\sigma^2}$, and $c_3 = c_1c_2$. A schematic representation of the solution is shown in Figure 1B. All images and initial source lie on a circle, equation A.6. The positions of the two first images are obtained by application of the matrices $\Pi_i(0)$, $i = 1, 2$ ($\alpha = 0$) to the initial condition, $\vec{x}_i = \Pi_i(0)\vec{x}_0$. Their coefficients can be obtained from the coefficients of the images required to solve the FPE, equation 3.17, with absorbing boundaries on the lines $x_1 = 0$ or $x_2 = 0$, equations A.4 and A.5. The third image can be thought of as the image of images 1 and 2 through the lines $x_2 = 0$ and $x_1 = 0$, respectively. It is uniquely defined because of the transitivity of the matrix product $\Pi_1(0)\Pi_2(0)$, that is, $\vec{x}_3 = \Pi_1(0)\Pi_2(0)\vec{x}_0 = \Pi_2(0)\Pi_1(0)\vec{x}_0$. The coefficient of the third image has been chosen in order to match the boundary condition on both semi-lines simultaneously. The solution for this case can also be obtained by multiplying the known individual solutions for two one-dimensional Wiener processes with absorbing boundaries (Cox & Miller, 1965).

## Acknowledgments

I thank Alex Pouget and Jan Drugowitsch for insightful discussions. This work has been partially supported by the Swartz Foundation.

## References

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron, 60*, 1142–1152.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behavior. *Trends in Cognitive Sciences, 11*(3), 118–125.

Bogacz, R., Brown, E., Moehlis, J., Hu, P., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review, 113*, 700–765.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes.* London: Methuen.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*, 427–430.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychol. Rev., 18*, 186–201.

Huk, A. C., & Shadlen, M. N. (2005). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Journal of Neurocience, 25*(24), 10420–10436.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswickian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychol. Rev., 104*(2), 344–366.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 445*, 227–231.

Landy, M. S., Goutcher, R., Trommershuser, J., & Mamassian, P. (2007). Visual estimation under risk. *J. Vis., 7*(6), 1–15.

Luce, R. D. (1986). *Response time: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision-making. *Cerebral Cortex, 13*, 1257–1269.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 22*, 9475–9489.

Pleskac, T. J., & Busemeyer, R. D. (2007). A dynamic, stochastic theory of confidence, choice, and response time. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 563–568). Austin, TX: Cognitive Science Society.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev., 85*(2), 59–108.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. Neurophysiol., 97*(2), 1756–1774.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychol. Rev., 116*(1), 59–83.

Risken, H. (1989). *The Fokker-Planck equation* (2nd ed.). Berlin: Springer-Verlag.

Roxin, A. & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput. Biol., 4*(3), e1000046.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience, 2*, 33–42.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Sciences (USA), 93*, 628–633.

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

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

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences, 2*(3), 169–194.

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

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *J. Vis., 8*(3), 1–15.

Wong, K.-F., Huk, A. C., Shadlen, M. N., & Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience, 1*, 6.

Van Zandt, T. (2000). ROC curves and confidence judgements in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn., 26*(3), 582–600.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature, 370*, 140–143.

---

Received December 27, 2008; accepted October 29, 2009.