Stochastic Accumulation by Cortical Columns May Explain the Scalar Property of Multistable Perception

Robin Cao,1,2 Jochen Braun,1 and Maurizio Mattia2,*

1Cognitive Biology, Center for Behavioral Brain Sciences, Otto von Guericke University, 39106 Magdeburg, Germany
2Department of Technologies and Health, Istituto Superiore di Sanità, 00161 Roma, Italy

(Received 30 January 2014; revised manuscript received 30 March 2014; published 27 August 2014)

The timing of certain mental events is thought to reflect random walks performed by underlying neural dynamics. One class of such events—stochastic reversals of multistable perceptions—exhibits a unique scalar property: even though timing densities vary widely, higher moments stay in particular proportions to the mean. We show that stochastic accumulation of activity in a finite number of idealized cortical columns—realizing a generalized Ehrenfest urn model—may explain these observations. Modeling stochastic reversals as the first-passage time of a threshold number of active columns, we obtain higher moments of the first-passage time density. We derive analytical expressions for noninteracting columns and generalize the results to interacting columns in simulations. The scalar property of multistable perception is reproduced by a dynamic regime with a fixed, low threshold, in which the activation of a few additional columns suffices for a reversal.

DOI: 10.1103/PhysRevLett.113.098103
PACS numbers: 87.19.ll, 02.50.Ga, 87.18.Sn, 87.19.Jc

Many perceptual and working memory tasks involve temporal integration of noisy evidence and their timing is often well modeled by simple stochastic processes (e.g., a Wiener process) [1]. Mean response times are fairly consistent and vary only by a few 100 ms between observers and stimulus conditions (e.g., [2]). The timing density (for a given observer and condition) typically resembles an inverse Gaussian distribution with a coefficient of variation $\approx 0.2$.

With ambiguous visual displays or auditory scenes, the viewer’s or listener’s perception often reverses suddenly and seemingly spontaneously between alternative interpretations (multistable perception [3]), perhaps reflecting a self-organized instability [4,5]. The timing of these stochastic reversals is unlike that of volitional responses. Mean periods of stable perception vary greatly (from $< 1$ s to several minutes) between observers and stimulus conditions (e.g., [5]). Nevertheless, higher moments of the timing density vary in particular proportions to the mean, so that a characteristic distribution shape is preserved: a Gamma distribution with a coefficient of variation $\approx 0.6$ [6–8].

This unique scalar property of reversal timing (i.e., maintaining distribution shape in spite of widely different means) suggests that, for a panoply of multistable phenomena in vision and audition, the distribution shape reflects very similar underlying processes, whereas the distribution mean is set independently by other, far more variable factors. Although the stochastic dynamics of multistable phenomena has been studied extensively [9,10], typically in terms of Ornstein-Uhlenbeck processes with nonlinear couplings [11], neither the large observer variation nor the scalar property have been explained.

Here we propose a simple stochastic process—which realizes a generalized Ehrenfest urn model—to resolve this puzzle. Its key ingredients are a stochastic accumulation of activity in a finite number of discrete neural nodes, such as cortical columns [12,13] and, internal to each node, a spontaneous dynamics modulated by sensory input. The stochastic accumulation to the threshold reproduces the Gamma-like distribution and the internal stimulus-dependent dynamics of nodes explains the large observer variation.

In treating multistable phenomena as a first-passage time (FPT) problem, we assume that the stochastic intervals between spontaneous reversals characterize an underlying random walk performed by microscopic states. We further assume that the conclusion of one stochastic accumulation coincides with the start of another such accumulation, so that a sequence of perceptual reversals is obtained. Several plausible reversal mechanisms have been proposed [9–11].

Our findings reveal an exceedingly simple physics of granular neural representations underlying the timing of cognitive events.

Results.—For each possible appearance of a multistable scene, we postulate a population of $N$ distinct neural nodes, such as cortical columns [12], representing the supporting sensory evidence [10,14]. Each node exhibits the stochastic internal dynamics of an attractor cell assembly [13]. Following previous work [14,15], we reduce this dynamics to spontaneous transitions between just two states (inactive and active), with rates $\nu_+$ (activation) and $\nu_-$ (inactivation). Transition rates $\nu_{\pm}$ are modulated by sensory input (i.e., the quality of the supporting evidence) differentially. Importantly, the transitions are fluctuation driven and occur independently in different columns.
When one perceptual appearance is replaced by another, the fraction of active nodes is effectively reset to an initial value $X_0$. (We are not concerned with the details of this process.) Thereafter, the fraction $X(t)$ relaxes stochastically towards its steady state $X_\infty$. The steady state depends on rates $\nu_\pm$ (see below), which in turn are set by the (constant) sensory input. When collective activity $NX(t)$ crosses a threshold $\theta$, the cycle begins anew. Thus, the intervals $T$ between reversals are modeled as FPTs of the stochastic variable $X(t)$.

Naturally, this scenario is highly oversimplified. Presumably, multistable phenomena involve at least two populations of columns (one for each perceptual appearance) switching roles during reversals. While one population (associated with the currently dominant appearance) habituates downward, another population (currently suppressed appearance) recovers upward. When the latter gains a differential advantage over the former, the cycle begins anew [14].

We now present a general expression for higher-order moments $\langle T^n \rangle$ of FPTs for a population of independent columns or nodes (i.e., in the absence recurrent interactions). In this case, $X(t)$ realizes a well-known birth-death process, namely, the generalized Ehrenfest urn model [16,17]. Assuming independent Poisson statistics for the death process, namely, the generalized Ehrenfest urn dynamics is described by the master equation for the columns or nodes (i.e., in the absence recurrent interactions, $X(t)$ relaxes stochastically.

$$\partial_t P_n(t) = (N-n+1)\nu_+ P_{n-1}(t) + (n+1)\nu_- P_{n+1}(t) - [(N-n)\nu_+ + n\nu_-] P_n(t),$$

for $n \in [0,N]$. For constant $\nu_\pm$, $P_n(t)$ evolves with a characteristic time $\tau = 1/(\nu_+ + \nu_-)$ towards a binomial distribution with an average fraction of active nodes $X_\infty = \nu_+/(\nu_+ + \nu_-)$ [16,17]. The dynamics is stochastic for small $N$ and becomes deterministic for $N \to \infty$. Note that $X_\infty$ provides a convenient (if indirect) measure of input strength.

We wish to solve the FPT problem with respect to an absorbing threshold at a certain number $\theta \in [0,N]$ of active nodes. Closed-form expressions for the higher-order moments of the FPT density exist [17,18], but include nested sums and products that offer no advantage over direct numerical integrations [19]. More convenient representations are known only for special cases, such as $X_0 = 0$, $\theta = N$, or the limit of $N \to \infty$ [20].

Here we overcome this difficulty and derive manageable expressions for the moments of the FPT density. To this end, we rewrite Eq. (1) in matrix notation

$$\tau \partial_t \tilde{P}(t) = \tilde{A} \tilde{P}(i)$$

and consider the spectral decomposition of the tridiagonal transition matrix $A$. The diagonal elements of $A$ are $A_{n,n} = -A_{n,n+1} - A_{n,n-1}$ and its off-diagonal elements are $A_{n,n+1} = (N-n)X_\infty$ and $A_{n,n-1} = n(1-X_\infty)$. The eigenvectors and eigenvalues of $A$ can be expressed in terms of Krawchouk polynomials $K_n(x)$ [16,21]:

$$K_n(x) = \sum_{k=0}^{n} \frac{(-n)_k (x)_k 1}{(-N)_k k! X_\infty^n},$$

where the Pochhammer symbol $(x)_n = x(x+1)\ldots(x+n-1)$ is defined as a rising factorial.

Given an initial number of active populations of $i = NX_0$, the statistics of the FPT $T_{i\theta}$ to an absorbing barrier at $\theta > i$ can be inferred from its Laplace transform $\mathcal{L}[T_{i\theta}]$, which is given by the following ratio [22]:

$$\mathcal{L}(T_{i\theta}) = \frac{K_i(-\theta)}{K_0(-\theta)}.$$

The $n$th order moment $\langle T^n_{i\theta} \rangle$ can be obtained recursively in terms of lower-order moments, which follows from the $n$th derivative of a quotient of functions [23]:

$$\langle T^n_{i\theta} \rangle = (-1)^n \mathcal{L}[T_{i\theta}]^{[n]}(0) = \tau^n K_i^{[n]}(0) - \sum_{j=0}^{n-1} \binom{n}{j} \tau^{n-j} K_0^{[n-j]}(0) \langle T^j_{i\theta} \rangle,$$

where we have used $K_0(0) = 1$. Derivatives $K_i^{[n]}(0)$ of Krawchouk polynomials essentially reduce to the derivatives of Pochhammer symbols evaluated at zero, which may be expressed in terms of unsigned Stirling numbers of the first kind $\binom{n}{m}$:

$$\lim_{x \to 0} x^m = m! \left[ \begin{array}{c} k \\ m \end{array} \right] = m!(k-1)!\omega(k,m-1),$$

where $\omega(n,m)$ is given by a recursive expression in terms of the $r$th order harmonic numbers $H_{n,r} = \sum_{k=1}^{n} k^{-r}$:

$$\omega(n,m) = \sum_{k=0}^{m-1} (1-m)_k H_{n-1,k+1} w(n,m-1,k),$$

with $\omega(0,0) = 1$ [24]. Combining Eqs. (5) and (6), we obtain the first three moments of the FPT as follows:

$$\langle T_{i\theta} \rangle = \tau \sum_{k=1}^{\theta} \frac{(-\theta)_k (-i)_k 1}{k(-N)_k} X_\infty^k,$$

$$\langle T^2_{i\theta} \rangle = -\tau^2 \sum_{k=1}^{\theta} \frac{(-\theta)_k (-i)_k 2H_{k-1,1} X_\infty^k + 2\langle T_{i\theta} \rangle \langle T^1_{i\theta} \rangle}{k(-N)_k},$$

(8)
\[
\langle T_{\theta}^2 \rangle = \sum_{k=1}^{\theta} \frac{(-\theta)_k - (-i)_k \frac{3(H_{k-1,1} - H_{k-1,2})}{k(-N)_k} X_{\theta}^k}{X_{\theta}^0} + 3\left(\langle T_{\theta} \rangle - 2\langle T_{\theta} \rangle^2\langle T_{\theta} \rangle + 3\langle T_{\theta} \rangle \langle T_{\theta} \rangle^3\right). \tag{9}
\]

Even higher-order moments can be worked out with comparable effort. The recursive expression above and in Eqs. (7)–(9) are new. Only the first moment, Eq. (7), was published previously [25].

In Fig. 1, direct simulations of \(N\) stochastic bistable columns are compared to the predictions of Eqs. (7)–(9). Specifically, the mean \(\langle T_{\theta} \rangle\), coefficient of variation \(c_v = \sqrt{\langle T_{\theta}^2 \rangle / \langle T_{\theta} \rangle^2 - 1}\), and skewness \(\gamma_1 = \langle (T_{\theta}^3) / \langle T_{\theta} \rangle^3 - 3c_v^2 - 1 \rangle / c_v^3\) of the FPT density are shown for different initial conditions \(X_0 \in \{0, 0.2, 0.4\}\) and different values of \(\nu_+\). The latter values were chosen to set the steady-state activation \(X_{\infty}\) at various locations below or above the threshold \(\theta/N\) \((X_{\infty} - \theta/N \in [-0.2, +0.2])\). The comparison verifies Eqs. (7)–(9).

In spite of its highly idealized nature, our analytical treatment of a population of cortical columns offers a tenable model for the scalar property of multistable reversals. Typically, these are characterized by a Gamma-like distribution with a coefficient of variation \(c_v \approx 0.6\) and a skewness \(\gamma_1 \approx 2c_v\). Importantly, different strengths of visual or auditory inputs do not change these values [7,8,26–28]. In contrast, volitional response times approximate inverse Gaussian distributions with much smaller coefficients of variation \(c_v \approx 0.2\) and significantly higher skewness \(\gamma_1 \approx 3c_v\) [1,2].

Recall that steady-state activity \(X_{\infty}\) (indirectly) represents sensory input. For different combinations of \(X_{\infty}\) and threshold \(\theta/N\), the analysis reveals curves of constant \(c_v\) with two branches [Figs. 2(a) and 2(b)]: a relatively low, fixed-threshold regime (suprathreshold \(X_{\infty}\)) and a high, variable-threshold regime (subthreshold \(X_{\infty}\)). In the former regime, \(c_v\) remains nearly constant when the threshold \(\theta/N\)
is fixed slightly above the initial activity $X_0$ [red dashed line, Fig. 2(c)], with $c_v = 1/\sqrt{\theta/N - X_0}$. In the latter regime, $c_v$ remains nearly constant when $\theta/N$ is adjusted to remain slightly above steady-state activity $X_\infty$ [blue dashed line, Fig. 2(c)]. This qualitative picture changes with neither population size $N$ nor initial conditions [Fig. 2(b)].

The two regimes differ dramatically in terms of mean FPT [Fig. 2(e)]: mean perceptual durations ($T$) are comparatively fast ($< \tau$) in the suprathreshold regime—as only the stochastic accumulation of few additional active columns is needed in order to breach threshold—and comparatively slow ($\gg \tau$) in the subthreshold regime, as fluctuations must carry activity above the steady state [Fig. 2(c)]. A telling difference concerns the distribution shape [Figs. 2(d) and 2(f)]: whereas the suprathreshold population will provide additional input to individual columns, interacting columns. Recurrent interactions within a population will provide additional input to individual columns, making transition rates $\nu_+$ dependent on collective activity $X(t)$ in a mean-field approximation [29].

In the regimes with high and constant $c_v$, it is sufficient to consider a linear dependence of excitatory couplings

$$\nu_+(X) = \nu_+^{(0)} + \nu_+^{(1)} X,$$

as $X(t)$ remains mostly in a narrow range just below threshold [see Fig. 2(c)]. Given this constraint, the FPT distribution would not change qualitatively if higher-order terms of the Taylor expansion were to be included in Eq. (10).

Interestingly, greater coupling strength $\nu_+^{(1)}$ leaves the $c_v$ of FPT essentially unchanged [Fig. 3(a), top]. Coupling decreases mean FPT, but mostly in the suprathreshold regime [Fig. 3(a), bottom]. Remarkably, the difference between the distribution shapes of sub- and suprathreshold regimes, as expressed in the ratio $\gamma_1/c_v$, is invariant with coupling strength and initial conditions [Fig. 3(b)]. Recurrent interactions between columns compress the FPT distribution in time, without changing its shape. In effect, interactions scale $\nu_+$ just as a progressive compression of the unit of time would do. This holds for both regimes exhibiting a scalar property.

Discussion.—Multistable phenomena are thought to expose a self-organized instability of perceptual representations [4-5], which ensures that all tenable interpretations of a sensory scene are explored [30]. The unique scalar property of multistable reversals may reflect these functional imperatives.

![FIG. 3](color online). Higher-order statistics of FPT for populations of interacting columns. (a) Top: comparison of coefficients of variation for interacting columns, $c_v$ (dress), and independent columns, $c_v^{\text{free}}$, for various coupling values $\nu_+^{(1)}$. Bottom: comparison of mean FPT, $\langle T_{\text{int}} \rangle$ and $\langle T_{\text{free}} \rangle$. Colors distinguish supra- and subthreshold regimes and hues indicate different initial conditions $X_0$, as in Figs. 2(e) and 2(f). (b) Difference in the shape of FPT distribution shape between sub- and suprathreshold regimes, as expressed by $\Delta \gamma_1/c_v$. Shading distinguishes initial conditions. The baseline rates $\nu_+^{(0)}$ were chosen to obtain $X_\infty = 0.8$ and $\tau = 1$ with zero coupling $[\nu_+^{(1)} = 0]$.

We have shown that a stochastic accumulation of collective activity in a finite population of (idealized) cortical columns can explain the scalar property. With or without recurrent interactions between columns, there are two threshold settings for which the coefficient of variation remains constant at $c_v \approx 0.6$ as the distribution mean shifts by an order of magnitude with sensory input [6-8,26-28]. A low and fixed threshold, placed just above initial activation ($\theta/N - X_0 \ll 1$), produces short mean durations (fractions of the relaxation time $\tau$) and a Gamma-like skewness $\gamma_1 \approx 2c_v$, precisely matching experimental observations. A high and variable threshold, placed just above the steady-state activation ($\theta/N - X_\infty \ll 1$), predicts far longer mean durations (multiples of $\tau$) and a significantly higher skewness $\gamma_1 \gtrsim 3c_v$, contrary to experimental evidence.

In the present framework, the scalar property of the stimulus-dependent FPT distribution emerges at the population level, being set by population size $N$ and threshold placement $\theta$. In contrast, the mean FPT reflects the intrinsic dynamics of individual columns, notably the relaxation time $\tau$. This additional degree of freedom may help explain the spurious variation of reversal times between observers [6,7].

An appealing feature of the proposed framework is its simplicity. To obtain an equivalent scalar property with Gamma-like distributions from other stochastic processes (e.g., Wiener or Ornstein-Uhlenbeck process), the noise
would have to be both discrete and state dependent. Specifically, noise would have to change by a factor of $\sqrt{\nu_+}/\nu_-$ as collective activity $X$ grows from zero to unity. A discrete Ehrenfest process, reflecting an underlying granularity of neural representations [28], provides both features in a natural way.

Large, sudden events alternating with long quiescent periods often reflect a stochastic collective dynamic among interacting microscopic components, as in chemical and biochemical reactions [17,31,32], neuron firing dynamics [33], or multistable physical systems [34–36]. We conclude by noting the close parallel between the scalar property in multistable perception and in timing cognition [29], both of which are particularly pronounced instances of Weber’s law in sensory psychophysics [37].

This work was supported in part by EU-project CORONET (FP7-ICT-269459). We gratefully acknowledge unpublished data on auditory multistability shared by István Winkler, Sue Denham, Dávid Farkas, and Alexandra Bendixen.

*maurizio.mattia@iss.it*

[23] C. Xenophontos (private communication).