Visual motion processing and perceptual decision making

Aziz Hurzook (ahurzook@uwaterloo.ca)
Oliver Trujillo (otrujill@uwaterloo.ca)
Chris Eliasmith (celiasmith@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Abstract

Perceptual decision making is a fundamental cognitive process widely studied in the behavioural sciences (Gold & Shadlen, 2007; Wang, 2008). We present a novel, biologically plausible model of visual motion processing and perceptual decision making, which is independent of the number of choice categories or alternatives. The implementation is presented in the form of a large-scale spiking neural circuit consisting of three main processes: 1) a velocity filter that uses the principle of oscillator interference to determine the direction and speed of pattern motion using networks of V1 simple cells; 2) a retinotopic representation of motion evidence in the middle temporal area (MT); and 3) competition-less integration of sensory 'evidence' over time by a higher-dimensional attractor network in the lateral intraparietal area (LIP). The mechanisms employed in 1) and 3) are new. We demonstrate the model by reproducing behavioral and neural results from classic perceptual decision making experiments that test the perceived direction of motion of variable coherence dot kinetograms. Specifically, these results capture monkey data from two-alternative forced-choice motion decision tests. We note that without any reconfiguration of the circuit, the implementation can be used to make decisions among a continuum of alternatives.

Keywords: perceptual decision making, continuous decision making, motion processing

Introduction

An important function of the mammalian brain is the ability to make decisions based on sensory input, and to take action based on these decisions. Organisms are constantly receiving sensory stimuli from their environment, and in order to choose sensible actions, they must sense and accumulate data over time until enough information exists to make a decision.

In this work, we offer two primary contributions in the computational modelling of a classic perceptual decision test. First, we take as our modelling starting point the visual intensity signals falling on the retina, from stimuli like those used in mammalian studies. Second, we show that the structure of the decision task is not relevant to the structure of the percept represented in the association cortex, and propose a novel mechanism to make decisions based on this structure.

A start-to-finish visual motion and perceptual decision circuit. We simulate the essential components of the primate motion perception and decision pathway using biologically plausible techniques at each stage of circuit modelling. From random-dot motion movies we generate burst signals known to occur in LGN (spatiotemporal derivatives of image intensity with noise reduced), the model then extracts velocity (direction and speed) information using a recurrently connected network of V1 simple cells, it then generates maps of optical flow in MT, and finally it integrates this evidence in LIP using an *n*-dimensional integrator from which the representation of perceived structure emerges, regardless of task structure. Unlike motion energy models and some related proposals (Adelson & Bergen, 1985; Rust, Mante, Simoncelli, & Movshon, 2006; Simoncelli & Heeger, 1998), the velocity selection mechanism we describe shows how recurrently connected spiking neurons can generate the observed spatiotemporal dynamics in V1 simple cells; that is, we show where the phase evolution of separable and inseparable Gabor-like V1 tunings comes from. Also new is our elimination of divisive normalization in the decoding of integrated vector quantities (Simoncelli & Heeger, 1998), and the use higher dimensional integration in MT. We are not aware of any past spiking neural models that include all of these stages of processing.

Decision making from the temporal integration of structured percepts. Past work employing integrators to explain perceptual decision making assumes that scalar evidence is integrated to a threshold (Wang, 2008). Many separate scalar integrators are proposed to mutually inhibit one another to explain more complex tasks (e.g. deciding between two, four, eight, etc. possible directions of motion). Here, we propose that a single vector integrator can account for any number of directions of motion. The concept of vector addition is simple: when two opposing vectors are added, they cancel; when two similar ones are added, they reinforce. If the vectors are time-dependent, then at any point in the time course of the integration we have the current state of perception (a vector). Thus, 'competition' among alternatives is misleading—there is no 'race' among 'competing' choice alternatives, as is typical of past models (M. E. Mazurek & Shadlen, 2003). Moreover, the percept vector is *independent* of the decision structure. In other words, the number of alternatives (two choices, *n* choices, a continuum) is irrelevant to the evidence accumulation process. Hence, the DV can be more generally interpreted as the *decision radius* ('DR', perhaps) of a percept vector evolving through integration in a higher dimensional *sphere* rather than a point on a line. The percept evolves over time as evidence accumulates, eventually crossing a *decision surface* ('DS', perhaps, rather than a decision threshold) if enough sensory evidence is accumulated. In the two-alternative forced choice task we use in our simulation, motion signals are integrated in two dimensions ($n = 2$) yet produce a binary decision, without reconfiguration of the circuit.

Our model suggests that the evidence that is accumulating

for perceptual decisions is a task-independent, *n*-dimensional *percept structure* (a vector) and not simply a task-dependent, one-dimensional category value (or decision variable, 'DV'). Since the percept structure can be interpreted as any time-dependent evidence state for any sensory modality, the circuit could provide a more general approach for the analysis of integrate-to-threshold processes. It could thus be applicable to arbitrary decision processes in the brain, of which the motion evidence domain is only one example. In what follows, we provide a summary of the theoretical principles supporting the model, a description of the model itself, experimental details and results.

Principles of model design

We use the leaky integrate-and-fire (LIF) neuron as our single cell model. The *activity* of an LIF neuron $a_i(J)$ can be thought of as the steady state firing rate of a neuron under a constant current J and is given by

$$a_i(J) = \left[\tau_{ref} - \tau_{RC} \ln\left(1 - \frac{J_{th}}{J}\right)\right]^{-1}$$

where J_{th} is the threshold current of the neuron, τ_{ref} is the refractory period for the neuron, and τ_{RC} is the membrane time constant for the neuron. To reduce computational demands, we focus only on instantaneous firing rates, as opposed to the precise spike time information, using what are known as *rate neurons*. It has been shown, however, that the same computations can be performed with a slight increase in the number of spiking neurons (Eliasmith & Anderson, 2003). Neurons in our model are coupled by a model of synaptic dynamics to give rise to biologically realistic dynamics, and hence empirically constrained timing data.

The general modelling techniques we use for building our simulation are collectively called the *Neural Engineering Framework* (NEF). The NEF is a method for performing large scale computations using any of a variety of simulated neurons (Eliasmith & Anderson, 2003). The NEF characterizes the encoding of vector values by populations of spiking neurons, and computation of optimal decoders that allow the approximation of linear or nonlinear functions between ensembles of neurons. This allows us to perform arbitrary computations on vector or scalar quantities using simulated neurons. The following paragraphs go on to describe our computational methods and the NEF in more detail.

Vector representation

Many empirical studies of mammals have found that populations of cortical neurons can encode real-world stimuli (Hebb, 2002). In the NEF, we encode vector-valued stimuli with populations of simulated neurons, or *ensembles*.

Encoding over neural populations. Each neuron in an ensemble is tuned to receive more ionic current J when responding to a certain stimulus vector $\mathbf{e}_i$, known as that neuron's *preferred direction vector*, and receive less current the further away the stimulus vector $\mathbf{x}$ is from $\mathbf{e}_i$. So given a vector stimulus $\mathbf{x} = (x_1, x_2, ..., x_n)$, we can relate the firing rate of a single neuron in the ensemble a_i to the stimulus by

$$a_i(x) = G_i\Big[J(x)\Big] = G_i\Big[\alpha_i(\mathbf{e}_i \cdot \mathbf{x}) + J_i^{bias}\Big]$$

where G_i is the nonlinear (spiking or non-spiking) function specific to our neuron model, α_i is a gain factor, and J_i^{bias} is a background bias current.

Decoding by optimal linear estimation. In addition to being able to encode stimulus values across neural ensembles, we also would like to be able to recover the original stimulus, given an ensemble's firing pattern. Using this method, we can build a representation for arbitrary stimuli with neural ensembles (Eliasmith & Anderson, 2003). The simplest way to do this is to make the assumption that the stimulus is a linear combination of the neural activities, which turns out to be quite accurate given enough neurons in the representation (Eliasmith & Anderson, 2003). That is, we assume our stimulus vector $\hat{\mathbf{x}}$ can be represented by

$$\hat{\mathbf{x}} = \sum_{i=1}^{N} a_i \mathbf{d}_i$$

with N being the number of neurons in the ensemble and $\mathbf{d}_i$ being a vector of decoding weights for neuron i. If we know $\mathbf{x}$, it is possible to find the optimal set of linear decoders $\mathbf{d}$ that minimize the squared error between $\mathbf{x}$ and $\hat{\mathbf{x}}$. This is a common problem in linear algebra, and can be solved as follows:

$$\mathbf{d} = \Gamma^{-1}\mathbf{v}$$
$$\Gamma_{ij} = \sum_x a_i a_j$$
$$\mathbf{v}_j = \sum_x a_j \mathbf{x}.$$

Solving for the optimal linear decoders, $\mathbf{d}$, allows us to recover an estimate of the original stimulus vector given a neural ensemble's activity. As we will see, it also allows us to directly compute the neural connection weights that perform a computation between two or more ensembles.

Vector transformation

Now that we have defined a way of encoding and decoding stimulus values, we can perform computations between neural ensembles using our encoding and decoding vectors. Suppose we want to have an ensemble $\mathbf{y}$ encode some function of the value another ensemble is encoding, $\mathbf{x}$. i.e. $\mathbf{y} = f(\mathbf{x})$. We simply compute the decoders for $\mathbf{x}$ as above, only substituting $f(\mathbf{x})$ for $\mathbf{x}$ when computing $\mathbf{v}_j$. Then in order to encode our desired function, we multiply our new functional decoding weights $\mathbf{d}$ by our encoding weights for population $\mathbf{y}$, yielding a new set of weights between the populations that generate the desired transformation.

$$\omega_{ij} = \alpha_j(\mathbf{d}_i \cdot \mathbf{e}_j)$$

where α_j is a gain term associated with neuron j. Note that this technique works well for nonlinear functions as well as linear ones, as we are in effect projecting into a higher dimensional space than our representation, effectively turning a nonlinear function into one that is linear in the weight space.

Population dynamics

The NEF also defines a way of computing functions defined over time, or *dynamic* functions. Incorporating time-dependance is important in understanding and modelling neural responses, since in the real world, neural activity is dependant on time. In general, we describe a linear dynamic function by $d\mathbf{x}/dt \equiv \dot{\mathbf{x}} = A(\mathbf{x}) + B(\mathbf{u})$, where $\mathbf{x}$ is the value currently being represented, and $\mathbf{u}$ is an input value from another ensemble.

One useful example of such a function is a two-dimensional oscillator, defined by $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. To have an ensemble exhibit this behavior, we define a *recurrent* connection from this population to itself as described in Eliasmith and Anderson (2003). As shown there, it is possible to solve for the connection weights that allow the ensemble to exhibit the desired behavior, allowing for the implementation of arbitrary dynamical systems.

Visual motion processing and perceptual decision making

The circuit we propose has three main information processing stages: 1) a velocity filter that uses the principle of oscillator interference to determine the direction and speed of pattern motion using networks of V1 simple cells; 2) a retinotopic representation of motion evidence in MT; and 3) competition-less integration of sensory evidence over time by an *n*-dimensional vector integrator in LIP. A schematic circuit diagram is depicted in Figure 1.

Velocity selection using oscillating networks of V1 simple cells

The extraction of direction of motion employs the oscillator interference (OI) mechanism, depicted in Figure 2. The initial translational motion of an edge in a local region of the visual field is encoded in a burst signal at $t = t_0$ ($\phi = 0$) to simulate LGN output. The signal is filtered through an input filter to control the initial phase of the oscillator. The input drives the rotation of the neurally represented state, $\mathbf{x}(t) = (r(t), \phi(t))$, through a progression of Gabor phase angles in the counter-clockwise direction, with a rotation period intrinsic to the oscillator. Damping effects cause the neural representation of $\mathbf{x}(t)$ to return quickly to zero without further input. Subsequent input bursts at times t_i add vectorially to, and thus *interfere* with, $\mathbf{x}(t)$. Constructive interference increases $||\mathbf{x}(t)||$ while destructive interference decreases it. Thus, if the direction and speed of the edge transiting the input gate of the neural oscillator are sufficiently close to the magnitude and phase of $\mathbf{x}(t)$, a resonance response occurs and $\mathbf{x}(t)$ sustains its magnitude and rotation. High responses from neurons tuned to states later in the period indicate strong velocity (direction and speed) correlation for all earlier phase times after t_0. Summation of the activities of the late-phase neurons from

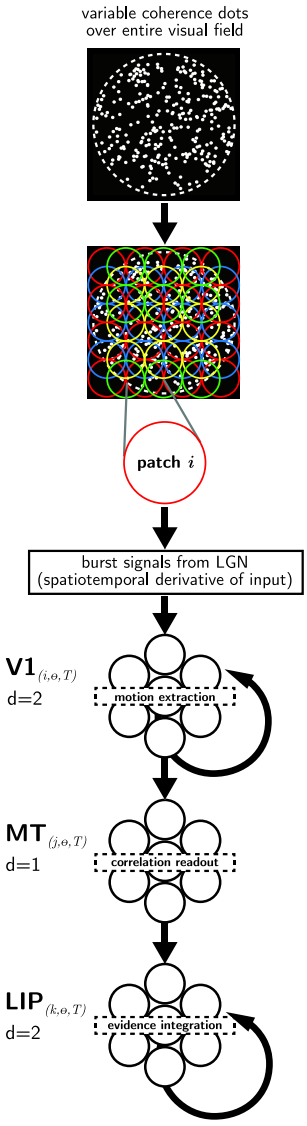

Figure 1: **Unit circuit schematic for perceptual decision circuit.** This figure details the circuit associated with each small patch of the visual field indexed by i. These units are repeated for each preferred direction, θ. Each cluster of circles shown is a neural ensemble with N LIF neurons. Index d is the dimensionality of the decoded quantity encoded by the ensemble. T is the period of the natural (undamped) frequency of the oscillator. Each MT ensemble pools the activities of several V1 ensembles with the same θ and T; likewise for LIP pooling of MT. The LIP ensemble is an n-dimensional integrator whose activity represents the direction of motion *vector* that emerges as motion evidence accumulates from all directions. In these simulations, $n = 2$ as we are testing for the perceived direction of motion in a plane.

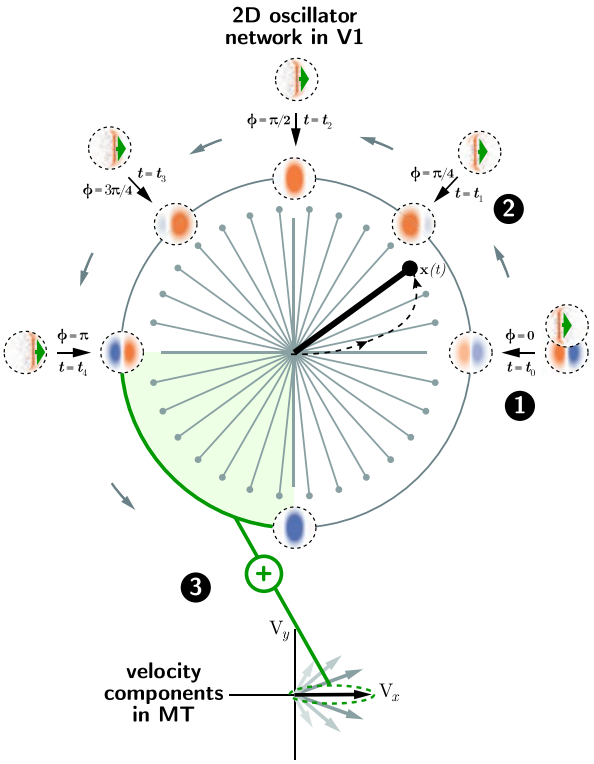

Figure 2: **Velocity selection mechanism based on oscillator interference (OI).** The velocity filter is an array of recurrently connected ensembles of direction selective V1 simple cells. The connection weights are determined using the NEF to endow the ensemble with oscillatory phase sensitivity and thus *speed* selectivity. The system *state* has components of magnitude and phase, $\mathbf{x}(t) = (r(t), \phi(t))$. The initial (rest) state is $\mathbf{x}(t) = (0,0)$. ① An initial burst signal from the LGN is triggered by the translational motion of an edge in the receptive field, shown as a bar moving to the right inside the dotted circle, overlapping the input filter. $\mathbf{x}(t)$ begins to increase in magnitude and rotate through the phase angles. ② Further input bursts at times t_1 to t_4 *interfere* constructively with the system state only if $\mathbf{x}(t) \approx \mathbf{x}(t_i)$. ③ The activity of neurons tuned to phases late in the period will be high only if correlation with visual input is similar earlier in the cycle. The late-phase activities drive an associated direction vector representation in MT. Other V1 oscillators associated with the same patch but tuned to different directions contribute a weight proportional to the component of motion velocity in their preferred direction (bottom, grey arrows).

the oscillator produce a scalar weight of an associated vector represented in a retinotopic field of motion evidence in area MT. This is a generic mechanism that captures motion information from any visual input.

Motion evidence map in MT

Figure 3 shows time snapshots of sample velocity maps represented in MT. These are depictions of the stimulus motion

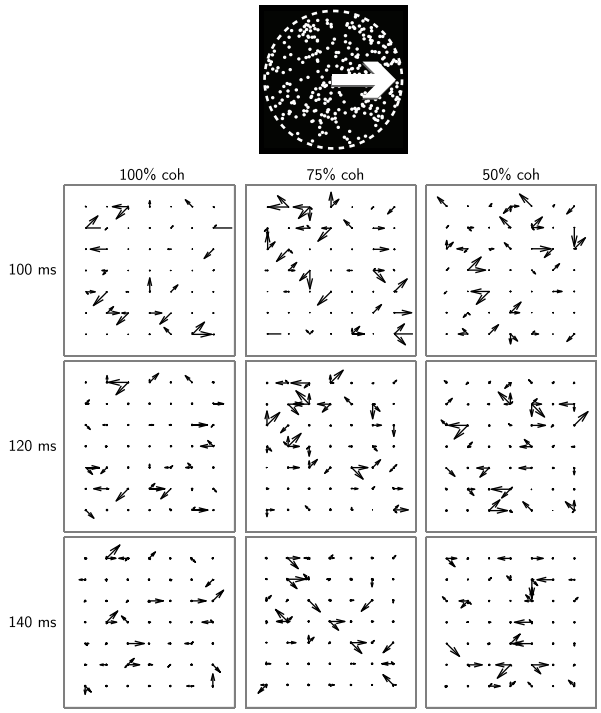

Figure 3: **Retinotopic velocity maps in MT.** Samples of vector read-out (optical flow) maps in MT for a 7×7 array of receptive fields for times $t = 100, 120, 140$ ms after stimulus input. The response latency was 50-65 ms. Stimulus coherence levels are categorized by column. For all coherence levels, the stimulus produces a distribution of motion responses. The target direction is not obvious from inspection and requires temporal integration.

in the visual field for any number of directions (for clarity we depict eight directions) at the given times. Each point in the 7x7 array represents the centre of a patch that is the domain of visual signal input to each unit circuit. The scalar output of each V1 oscillator provides the weight of an associated velocity for a given patch in the field. It should be stressed here that no task-dependent categorization of the motion field is imposed.

For complex pattern motion like variable coherence dots, even at high coherence levels (50-100%), the wide distribution of velocity response maps provides an indication as to why temporal integration is required for the biased direction to emerge.

Higher dimensional vector integration in LIP

An important contribution of the model is its employment of a higher-dimensional vector integrator. The linear dynamical equation is

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}(t)$$

where $A = 0$, $B = I$ (the identity matrix), and $\mathbf{u}(t)$ is the input evidence. Using the NEF we can determine that the recurrent matrix for neurons to implement this dynamical system is

$$\omega_{ij} = \alpha_j \mathbf{d}_i (A + I) \mathbf{e}_j = \alpha_j \mathbf{d}_i \mathbf{e}_j$$

where i and j index the same population of neurons. Because the NEF is defined for vector representations, these weights will result in a neural state that represents the integration of information in the dimensionality of $\mathbf{x}$ (in this case $D = 2$). Multi-dimensional integrators of this sort have been previously employed in neural models of working memory (Singh & Eliasmith, 2006), but not for decision making.

Experiment

Model implementation

The neural system simulation package used to implement the circuit was *Nengo*, (*http://nengo.ca*). Table 1 provides the neurophysiological parameters used. A total of 2.9×10^5 spiking LIF neurons were used. The random-dot motion movies were generated using the *Psychtoolbox-3* extensions for Matlab® (Kleiner, Brainard, & Pelli, 2007; Pelli, 1997; Brainard, 1997). The visual input signal was in the preferred directions of the associated V1 oscillators. To simulate thalamic bursting (Butts et al., 2010), temporal derivatives of spatial overlap between the stimuli and oscillator input filter were taken at 2-ms pulse widths.

Decision test description

We performed a two-alternative, forced-choice, fixed duration test of 1-second duration, using variable coherence random-dot motion movies for a single patch. The decision threshold value was held fixed and was the only parameter adjusted to fit behavioural data. The length of the percept state vector, when the average success rate of the circuit was 80%, was used as the decision radius (analogous to the decision threshold use by Gold and Shadlen for the same test in monkey trials (Gold & Shadlen, 2007). The coherence level (motion strength) was lowered progressively, decreasing motion information and stressing the signal-to-noise ratio resolving capability of the circuit. For each coherence level 10 tests were run.

Results

The model was able to determine direction of motion in the majority of cases down to about 5% coherency (Figure 5), and showed similar characteristics to data collected from monkeys in (Gold & Shadlen, 2007). Particularly, as shown in Figure 4, neuron responses in area MT stayed relatively constant over time, with certain neurons showing stronger firing rates when given stronger motion evidence (higher coherency). At the same time, neuron responses in area LIP got stronger over time, particularly when nearing the decision threshold under medium to high coherency. Additionally, as shown in Figure 5, the experimental results relating to the percentage of correct decisions and time taken to make a decision over varying coherency levels were in accordance with experimental data.

Conclusion

In the TAFC visual decision task we have used to test our model, we have shown the validity the OI velocity selection

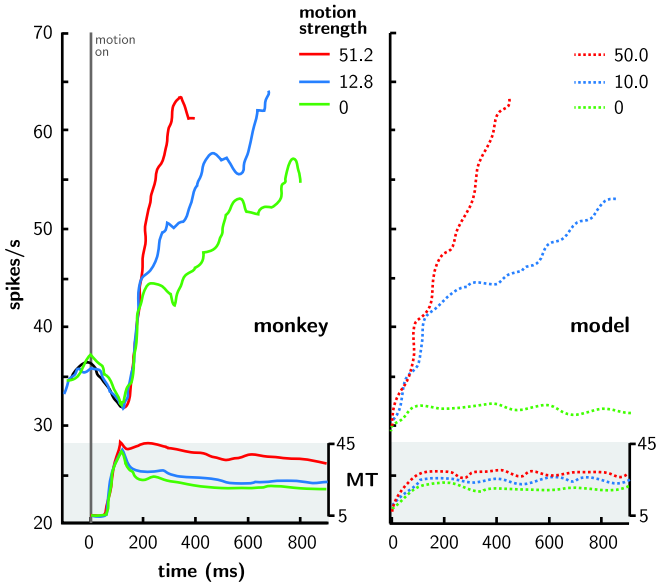

Figure 4: **Electrophysiology of MT and LIP neurons during the decision task.** Recreated from (Gold & Shadlen, 2007).

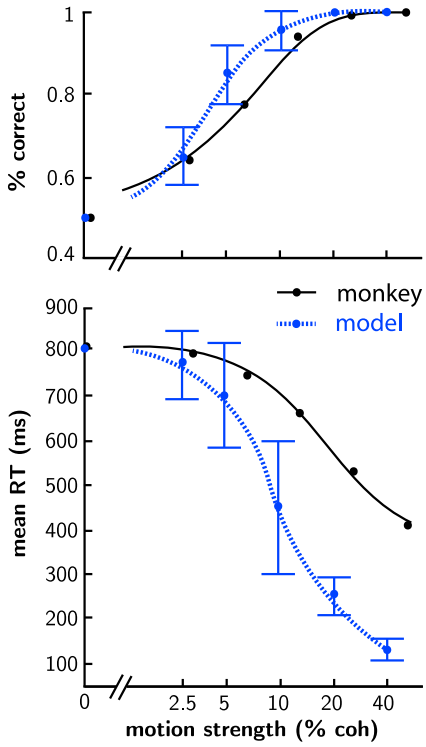

Figure 5: **Psychometric performance.** The circuit can discern motion direction reliably for coherence levels down to 10%, below which it drops to 50% success (random guess) as motion strength approaches 0. The disparities in reaction time between our model and the experimental data may be attributable to motor reaction time and other behavioural factors for which we do not account. Monkey data plots recreated from (Gold & Shadlen, 2007).

	Ensemble parameters	Model Value	Biological Value	Reference
V1	RC constant (τ_{RC})	20	10-20	(Shadlen & Newsome, 1994)
	Post-synaptic constant (τ_{psc})	5.0	$\sim$6.6	(Faber & Korn, 1980)
	Abs refractory period (τ_{ref})	2	$1^{\dagger}$	(Friedman-Hill, Maldonado, & Gray, 2000)
	Max firing rate	100-250	$\sim$ 70-100	(Carandini & Ferster, 2000)
MT	RC constant (τ_{RC})	20	10-20	(McCormick, Barry W. Connors, & Prince, 1985)
	Post-synaptic constant (τ_{psc})	5.0	$\sim$6.6	(Faber & Korn, 1980)
	Abs refractory period (τ_{ref})	5	$1^{\dagger\dagger}$	–
	Max firing rate	100	100-200	(Felleman & Kaas, 1984)
LIP	RC constant (τ_{RC})	20	–	–
	Post-synaptic constant (τ_{psc})	5.0	–	–
	Abs refractory period (τ_{ref})	5	–	–
	Max firing rate	70	70	(Gold & Shadlen, 2007)

Table 1: **Neurophysiological parameters used.** † = value based on a model estimate. †† = using V1 value. (–) = not available.

mechanism and the effectiveness of integrating a percept vector over time, without any consideration of the number of choice alternatives. The percept vector evolved over time, toward the left or right direction in two dimensions, producing a binary decision. This was due to the nature of the input, the sensory processing and integration mechanisms, and not any imposed task structure. Since the OI mechanism is isometric in the visual plane, identical results would result from forced choice tasks in any direction. We have tested the same model with additional forced-choice options (e.g. 4 and 8), and it performs similarly well (results not shown). Predictably, fewer choice alternatives lead to faster decisions, since the minimum detectable difference in signal level between two alternatives is greater than if that same magnitude were distributed among 8 alternatives.

It is natural for us to consider the percept vector and its temporal integration to a DS in much higher dimensions. The approach we have presented here can likely be applied to higher order sensory or non-sensory decision making that requires integration of evidence over time.

References

Adelson, E., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2).

Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.

Butts, D., Desbordes, G., Weng, C., Jin, J., Alonso, J., & Stanley, G. (2010). The episodic nature of spike trains in the early visual pathway. *Journal of neurophysiology*, *104*(6), 3371–3387.

Carandini, M., & Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience*, *20*(1), 470-484.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT Press.

Faber, D. S., & Korn, H. (1980). Single-shot channel activation accounts for duration of inhibitory postsynaptic potentials in a central neuron. *Science*, *208*(4444), 612-615.

Felleman, D., & Kaas, J. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology*, *52*, 488-513.

Friedman-Hill, S., Maldonado, P. E., & Gray, C. M. (2000). Dynamics of striate cortical activity in the alert macaque: I. Incidence and stimulus-dependence of gamma-band neuronal oscillations. *Cerebral Cortex*, *10*, 1105-1116.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535-74.

Hebb, D. O. (2002). The organization of behavior: A neuropsychological theory. In (new ed.). Psychology Press.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*. (ECVP Abstract Supplement)

McCormick, D. A., Barry W. Connors, J. W. L., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, *54*(4).

M. E. Mazurek, J. D., J. D. Roitman, & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *3*(11).

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11).

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*, 569-579.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in area MT. *Vision Research*, *38*(5).

Singh, R., & Eliasmith, C. (2006). Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *The journal of neuroscience*, *26*(14), 3667–3678.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron*, *60*.