



HAL
open science

The relative contribution of noise and adaptation to competition during tri-stable motion perception

Andrew Isaac Meso, James S Rankin, Olivier S Faugeras, Pierre Kornprobst, Guillaume S Masson

► **To cite this version:**

Andrew Isaac Meso, James S Rankin, Olivier S Faugeras, Pierre Kornprobst, Guillaume S Masson. The relative contribution of noise and adaptation to competition during tri-stable motion perception. Journal of Vision, 2016. hal-01383118v2

HAL Id: hal-01383118

<https://inria.hal.science/hal-01383118v2>

Submitted on 18 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The relative contribution of noise and adaptation to competition during tri-stable motion perception

Andrew Isaac MESO^{a,d*}, James RANKIN^{b,c}, Olivier FAUGERAS^b,
Pierre KORNPORST^c, Guillaume S MASSON^a

Adresses:

^a Institut de Neurosciences de la Timone
UMR 7289 CNRS & Aix-Marseille Université
27 Bd Jean Moulin
13385 Marseille Cedex 05, France

^b Université Côte d'Azur
Inria Sophia Antipolis Méditerranée
Team Mathneuro
Routes des Lucioles – BP 93,
Sophia Antipolis 06902, France

^c Center for Neural Science,
New York University
4 Washington Place,
New York, NY 10003

^d Psychology and Interdisciplinary Neurosciences Research Group,
Faculty of Science and Technology,
Bournemouth University, Fern Barrow,
Poole BH12 5BB, United Kingdom.

^e Université Côte d'Azur
Inria Sophia Antipolis Méditerranée
Team Biovision
Routes des Lucioles – BP 93,
Sophia Antipolis 06902, France

*Corresponding Author contact
email: ameso@bournemouth.ac.uk

Keywords: Multistable motion perception, Competition model, Eye movements, Noise, Adaptation

Abstract

Animals exploit antagonistic interactions for sensory processing and these can cause oscillations between competing states. Ambiguous sensory inputs yield such perceptual multi-stability. Despite numerous empirical studies using binocular rivalry or plaid pattern motion, the driving mechanisms behind the spontaneous transitions between alternatives remain unclear. In the current work, we used a tri-stable barberpole motion stimulus combining empirical and modelling approaches to elucidate the contributions of noise and adaptation to underlying competition. We first robustly characterised the coupling between perceptual reports of transitions and continuously recorded eye direction, identifying a critical window of 480ms before button presses within which both measures were most strongly correlated. Second, we identified a novel non monotonic relationship between stimulus contrast and average perceptual switching rate with an initially rising rate before a gentle reduction at higher contrasts. A neural fields model of the underlying dynamics introduced in previous theoretical work and incorporating noise and adaptation mechanisms was adapted, extended and empirically validated. Noise and adaptation contributions were confirmed to dominate at the lower, and higher, contrasts respectively. Model simulations with two free parameters, controlling adaptation dynamics and direction thresholds, captured the measured mean transition rates for participants. We verified the shift from noise dominated towards adaptation-driven in both the eye direction distributions and inter-transition duration statistics. This work combines modelling and empirical evidence to demonstrate the signal strength dependent interplay between noise and adaptation during tri-stability. We propose that the findings generalise beyond the barberpole stimulus case to ambiguous perception in continuous feature spaces.

Introduction

Multi-stable perception occurs when a sensory input is ambiguous and can be interpreted in more than one way. This input is represented within populations of neurons in which there is a competition between the encoded alternatives, resulting in transitions in what is perceived over time (Leopold & Logothetis, 1999). A vast collection of experiments have been carried out to characterise this switching for different perceptual conditions, such as binocular rivalry (Lehky & Blake, 1991; Levelt, 1966; Matsuoka, 1985), ambiguous figure perception (Long & Toppino, 2004) and plaid motion direction (Hupe & Rubin, 2003). The fact that similar dynamics were observed in these different conditions has argued for the involvement of a few generic mechanisms known to play a critical role in dynamical systems. In particular, the role of noise and asynchronous adaptation in the transition dynamics has been investigated and modelled (Moreno-Bote, Rinzel, & Rubin, 2007; Shpiro, Curtu, Rinzel, & Rubin, 2007; Theodoni, Panagiotaropoulos, Kapoor, Logothetis, & Deco, 2011). Still, previous experimental work has remained largely inconclusive: empirical evidence for both adaptation driven and noise driven systems were found by testing for the statistical signatures of each mechanism, particularly the distribution of times between perceptual transitions (Hupe & Rubin, 2003; Lehky, 1995; Zhou, Gao, White, Merk, & Yao, 2004). The respective contribution of each factor remains unclear. Such difficulty was circumvented by theoretical work using a mathematical exploration of multi-stability to identify a parameter region in the dynamical system where there is a balance between contributions of noise and adaptation (Shpiro, Moreno-Bote, Rubin, & Rinzel, 2009). These authors proposed that perceptual transitions could occur in the neighbourhood of such balance points, where small parametric changes can shift the driving mechanisms. This theoretical proposition was recently supported using plaid patterns through the empirical analysis of switching patterns between one coherent and two transparent stimulus percept alternatives. Huguet and colleagues reported that in the presence of these three competing states, adaptation determines which of the two alternative states is next for each switch and noise drives the instant at which the switch occurs (Huguet, Rinzel, & Hupe, 2014).

2D motion integration is a well-known source of perceptual multi-stability. Because of the aperture problem, local motion signals of elongated uni-dimensional (1D) edges are highly ambiguous: a single physical translation yields a wide range of possible perceived directions (Wallach, 1935). The 2D global motion direction of a pattern can be reconstructed by integrating several non-parallel 1D motion signals, such as in plaids (Adelson & Movshon, 1982) or by combining 1D and local non ambiguous two-dimensional (2D) features such as line-endings in barber poles (Shimojo, Silverman, & Nakayama, 1989). This 2D motion computation has complex dynamics and the perceived global motion direction can vary over different time scales. On a short-time scale, the perceived direction of a barber-pole motion stimulus changes over the first hundred milliseconds from the direction orthogonal to the component grating towards the longer axis of the aperture (Castet, Charton, & Dufour, 1999; Castet & Zanker, 1999). These early, fast dynamics result from the competition between 1D (grating) and 2D (i.e. line-endings) motion signals (Shimojo et al., 1989). 2D motion signals are slightly delayed with respect to 1D signals but dominate the competition after a few hundreds of milliseconds (Castet et al., 1999; Masson, Rybarczyk, Castet, & Mestre, 2000). Longer time scales yield richer dynamics. For a square aperture such as illustrated in Figure 1, perceived direction is tri-stable, switching between diagonal (i.e. 1D-dominated) from onset and horizontal or vertical (i.e. 2D-dominated) over timescales of hundreds of milliseconds (Castet et al., 1999; Masson et al., 2000). A key aspect of 2D motion integration captured by this stimulus is that all the three possible states belong to the continuous space of global motion direction representation. Herein, perception can shift smoothly through intermediate directions. Such a continuous space departs from the discrete states that typically compete during binocular rivalry or plaid motion perception.

We previously conducted a series of experimental and theoretical studies which showed the advantage of a continuous space to better characterize the transition dynamics and detail how noise and adaptation mechanisms can change the competition between the neuronal populations tuned to different global motion directions (Meso & Masson, 2015; Rankin, Meso, Masson, Faugeras, &

Kornprobst, 2014). In the current psychophysical study, we probed this balance between noise and adaptation mechanisms by systematically varying the input strength through visual stimulus contrast changes. For binocular rivalry, it was originally proposed that the perceptual switching rate should increase monotonically with contrast, a rule known as Levelt's fourth (IV) law (Levelt, 1966). With barber-poles, we found that the relationship between input strength (i.e. grating contrast) and switching rate was non monotonic, thus deviating from Levelt IV. Our theoretical work suggested that the difference in contrast gain of 1D and 2D features, as well the balance between noise and adaptation could explain such a non-monotonic relationship, depending on each subject's modelled perceptual threshold (Rankin et al., 2014). That conjecture remained empirically unverified until the current study.

Psychophysical studies of perceptual multi-stability suffer from several limitations. First, reported directions are forced between a pre-set, limited number of possible responses, reducing the advantages of having a continuous space. Second, it is impossible to precisely estimate the time course of perceptual switches as well as the perceptual stability around the time of a switch. Third, it is difficult to ensure that noise and adaptation mechanisms impact a single, low-level representation of global motion direction. To overcome these limits, we probed the dynamics of 2D motion computation using both perceptual reports and eye movements. Ocular tracking responses are a powerful means of continuously sampling the current state of the neuronal populations representing global motion direction (see Masson & Perrinet, (2012); Lisberger, (2010) for reviews). Changes in tracking direction can thus indicate the transition dynamics from one percept to another with a much higher temporal resolution than perceptual reports (e.g. for barberpole motion: (Barthelemy, Fleuriet, & Masson, 2010; Barthelemy, Perrinet, Castet, & Masson, 2008; Masson et al., 2000)). However, gradual changes in eye direction are also intrinsically noisy. Our approach was to optimally combine the discrete, categorical but low-resolution perceptual reports with the continuous, high-resolution but noisy ocular responses. Combining eye movements and perceptual reports, we showed a robust link between reported perceived directions and instantaneous eye direction revealing the critical time window most informative of perceptual state. We note that this link, which relied on both measures, held statistically only when considering the direction of an eye movement trace over a duration in which a reported perceptual transition was known to have been made. The eye traces were too noisy to make this same link conversely, that is, by identifying the timing and direction of perceptual transitions purely from the eye movements. We were then able to carry out a detailed analysis of the patterns in the transitions between global motion direction states and compare their statistical properties with predictions made by our neural fields model of 2D motion integration incorporating noise and adaptation and extending the model of Rankin et al. (2014). We report that varying stimulus contrast results in a non-monotonic relationship between average switching rate and input strength. From comparisons of both the reported transition patterns and the eye movement traces (restricted within the established critical window) to model predictions for different sections of the transition rate-contrast curve, we find for the first time, evidence of a systematic shift from noise dominated at low contrasts, to adaptation dominated at higher contrasts.

Materials and Methods

Observers

Nine volunteer human psychophysical observers (six male and three female) with normal or corrected to normal vision were used for the experiments. Two of these were authors while the remaining seven were unaware of the hypotheses and aims of the study. Four including one of the authors were experienced psychophysical observers who had participated in numerous unrelated previous experiments (participants S2, S5, S6 and S7). They endured the constraints of limited blinking in trials, task difficulty and long blocks better than others. All participants gave their informed written consent and experiments were carried out according to ethical guidelines and approved by the Ethics Committee of Aix-Marseille Université, in accordance with the declaration of Helsinki.

Visual stimulus

A moving sinusoidal luminance grating within a square aperture was presented on a screen 57cm from the observer whose head was maintained in position on a chin and headrest. The stimulus aperture subtended a retinal image with sides of 10 degrees of visual angle. The stimulus $B(x,y,t)$ was generated according to Equation (1).

$$B(x, y, t) = L_M [1 + A_c \sin(2\pi fvt + f(ax + by))] \quad (1)$$

$$a = 2\pi \cos(\theta + \psi) \text{ and } b = 2\pi \sin(\theta + \psi) \quad (2)$$

In Equation (1) L_M is the mean luminance of the screen (25.8cd/m^2), A_c is the amplitude factor between 0-1 which scales the Michelson contrast, v is the grating speed (6deg/sec) and f is the spatial frequency (0.41cyc/deg). In Equation (2), $\theta = 45^\circ$ and ψ took values of $0, 90, 180$ and 270° to randomise presentations in the four oblique directions. Stimuli were generated on a Mac computer running Mac OS 10.6.8 and displayed on a Viewsonic p227f monitor with a 20" visible screen of resolution 1024×768 at 100Hz . Task routines were written using Matlab 7.10.0. Video routines from Psychtoolbox 3.0.9 were used to control stimulus display (Brainard, 1997; Pelli, 1997). Eye movements were recorded using an SR Eyelink 1000 video eye tracker.

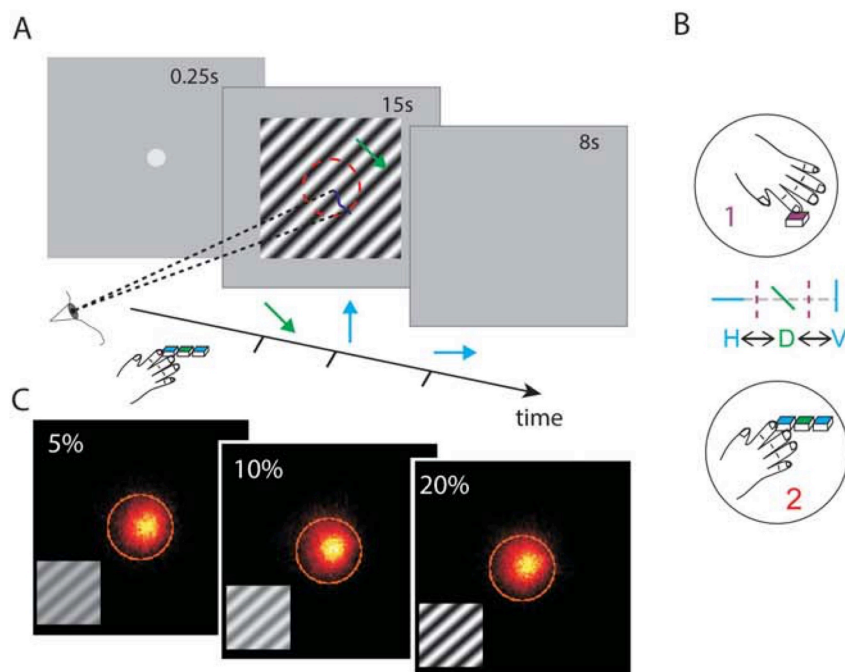


Figure 1: *Stimulus and task illustration.* (A) Panels showing trial sequence of 0.25s fixation, 15s moving grating stimulation and then a grey screen for 8s. Participants make button presses to indicate each change in perceived direction during stimulation and eye movements are recorded. (B) Button presses reporting transitions were collected either as a single report (*task 1*) to indicate when perceived direction changed without specifying direction in a simplified task; or with three directions H, D and V explicitly recorded (*task 2*), reported by participants to be a more difficult task. (C) Three panels showing eye position heat maps with combined recordings from all observers at three example contrasts of 5, 10 and 20%, inset gratings illustrate the contrasts near both ends of the tested range. Gaze remained largely within the central 3° of visual angle during the task demarcated by the orange circles.

Procedure

Participants sat in front of the screen and kept fixation on a grey spot of 0.27° diameter before each trial. Fixation disappeared and the stimulus appeared for 15s (see Figure 1A) while eye movements were recorded. This duration was a compromise between the need to maximize recorded perceptual switches up against restrictions to minimise effects of observer fatigue, blinking and adaptation of the local motion detectors resulting in motion after effects. The instruction was to respond to indicate the precise instance of each

transition between three perceived direction choices, horizontal (H), diagonal (D) and vertical (V). This response was made in two ways: with *a single button press* marking the timing of each transition and recording no distinction between actual directions (easier condition: *task. 1*, Figure 1A-B) or alternatively with *one of three button presses* at each transition corresponding to H, D or V (reported to be a more difficult condition: *task. 2*, Figure 1A-B).

Each experimental block was made up of either 36 (*task 1*) or 42 (*task 2*) trials, each containing six contrast conditions ($c = 0.03, 0.05, 0.08, 0.1, 0.15, 0.2$ i.e. 3-20%). There was an 8s wait after each trial before the observer initiated the next trial with a button press. Each block was repeated six to eight times for *task 1* and eight times for *task 2*, after a couple of initial blocks to familiarise participants with the task. Bad trials e.g. excessive blinking, containing many abrupt eye movements unlikely to be stimulus driven or self reported ‘bad blocks’ of lapsed attention were discarded. Note that *Task 2* has been explored extensively in a separate study, fully characterising the changing patterns in the relative prevalence of each of the perceptual choices (Meso & Masson, 2015). The present study builds on that work with a larger data set and a focus on the analysis of perceptual switching probed dynamically with the additional tool of smooth eye movements. The use of the simplified task (*task 1*) allowed inexperienced participants to perform the experiments more confidently than *task 2*. The choice of design omitting useful direction information was made because the complexity added by three separate button presses (H, D or V) made it too difficult for most inexperienced participants whose data we deemed critical for generalising our results.

Perceptual report analysis

We analyse the trend in the reported transition rate as a function of stimulus contrast by fitting the experimentally measured switching rates (both for individual observers and grouped averages) to two nested nonlinear functions which either rise to an asymptotic point of *saturation* (Equation 3) or rise to a *peak* (Equation 4) before gently descending with contrast.

$$F_s(c) = Amp \frac{c^n}{(c^n + Cf^n)} \quad (3)$$

$$F_p(c) = Amp \frac{c^n}{(c^n + Cf^n)} \times (1 - c^{ss}) \quad (4)$$

The basic function is the Naka Rushton used only for its convenient shape, and not intended to imply any underlying contrast response processes at this stage. The terms are the amplitude *Amp*, the exponent *n*, the C50 term *Cf* and for the peak function of Equation 4, there is an additional super saturation exponent term *ss*. The data is fitted to both these functions using an iterative least squares processes to identify the best parameters. A Kolmogorov-Smirnov Goodness of fit test for significance is carried out on the results of each of the pair of fits and the results are then further compared using the Akaike and Bayesian Information Criterion measures (AIC/BIC). These measures use likelihoods from the fits to determine which model provides a better explanation for the data, taking into account the number of parameters, thus penalising less parsimonious models (Akaike, 1981; Schwarz, 1978; Wagenmakers & Farrell, 2004).

Eye movement recordings and analysis

Eye movements have often been recorded to probe perceptual multi-stability (Hayashi & Tanifuji, 2012; Logothetis & Schall, 1990; Niemann, Ilg, & Hoffmann, 1994; Quaia, Optican, & Cumming, 2016). Barberpole motion stimuli are known to drive reflexive tracking eye movements whose dynamics reflect that of the global motion direction (Barthelemy et al., 2010; Masson et al., 2000). Thus, essential information can be extracted from the temporal dynamics of eye velocity about the instantaneous state of the motion representation and transitions between states. However, such slow eye movements can change the pattern of retinal motion. We included an intermittently re-appearing central fixation spot during the 15s stimulus presentation which observers were instructed to use to initiate a saccade back to the centre of the stimulus. The fixation interval was either randomised between 1.5-3s for each reappearance or fixed at a constant value within this range for different blocks. Controls confirmed that the fixation interval period made no difference to the perceptual switching rate. This interrupted fixation paradigm allowed brief but significant epochs of small tracking eye movements, while maintaining the visual motion input roughly constant over time, as illustrated by the eye positions in Figure 1C. Data to be analysed was collected once participants reported being comfortable with the task and instructions, and blinking and switching rates had settled down accordingly, typically after 1-3 practice blocks.

Right eye movements were recorded with the video eye tracker at a frequency of 1KHz (examples, Figure 2). The eye-position data was first low pass filtered with a Butterworth filter (6th-order, 50Hz).

Velocity traces were derived from the two-point central difference between the symmetric-weight moving averages of the position samples. Standard criteria were used to identify the start and endpoints of both saccades and blinks in these traces (Engbert & Kliegl, 2003), using specific algorithm parameters identical to those detailed in a previous work (Meso, Montagnini, Bell, & Masson, 2016). An extension of an additional 20ms before and 30ms after each of these events were identified for exclusion along with the events to conservatively restrict the traces to just the smooth sections containing gaps, e.g. Figures 2, right hand column. Combination of the x and y velocities was used to obtain the direction θ , estimated from the inverse tangent of the ratio of y and x components. Operations on eye direction were made in a circular space after aligning all stimulus directions (Berens, 2009). Processing described in this section was carried out using bespoke Visual C++ and Matlab routines.

Predicting perceived direction from eye direction

To generalise the relationship between the categorical transitions in perceived motion direction (H-D-V) and the changes in eye direction, we sought to establish whether time averaged eye direction distributions could be used to predict which state (H, D or V) had been reported during each button press in *task 2*. We designed a three parameter decoding scheme using the time averaged eye direction μ_θ estimated over a duration defined by the first decoding parameter, a variable temporal window size N ,

$$\mu_\theta = \frac{1}{N} \sum \theta_t, \left(t \in \left\{ -\frac{N}{2}, -\frac{N-1}{2}, \dots, +\frac{N}{2} \right\} \right), \quad (5)$$

and the resulting spread in the form of the standard deviation σ_θ ,

$$\sigma_\theta = \sqrt{\frac{1}{N-1} \sum_t (\theta_t - \mu_\theta)^2}, \left(t \in \left\{ -\frac{N}{2}, -\frac{N-1}{2}, \dots, +\frac{N}{2} \right\} \right), \quad (6)$$

N could be in one of three configurations along the eye trace: symmetrically centred around the instance of the button press equally extending both before and after it (symm., case 1), running from the past before the button press and stopping at the button press (pre-button, case 2) and starting from the button press and stopping some time after in the future (post-button, case 3). The generated mean, μ_θ and spread, σ_θ parameters serve as inputs into a piecewise decision operation which assigns a perceived direction using the second and third parameters, PT_H and PT_V .

$$Dir(\mu_\theta, \sigma_\theta) = \begin{cases} V : & (\mu_\theta - k\sigma_\theta) \geq PT_V \\ H : & (\mu_\theta + k\sigma_\theta) \leq PT_H \\ D : & (\mu_\theta - k\sigma_\theta) < PT_V \text{ and } (\mu_\theta + k\sigma_\theta) > PT_H \end{cases} \quad (7)$$

In this formulation, the average direction μ_θ is compared to the two threshold parameters PT_H and PT_V after the addition of the spread parameter scaled by a constant k (fixed at $k=0.25$ following initial optimisation), which captures the fact that it is the distribution rather than just the mean, whose position within the direction space is being categorised. Predictions are then made for a range of combination of values (375,000) of simulated symmetrical, pre-button and post-button temporal windows N , and parameters PT_{HV} to find the combination which optimises correct prediction for each of the participants' individual data sets. To obtain estimates of chance performance for comparison as a baseline, the shuffled set of the recorded perceptual choices made by each participant are re-assigned as decisions for each recorded transition.

Mathematical model overview

Global motion integration is often described as a two-stage process. First, local motion is sensed by a set of spatio-temporal filters with small receptive fields. Here we assumed that a set of different local motion filters extract grating and terminator motion directions (Castet & Wuerger, 1997; Lorenceau, Shiffrar, Wells, & Castet, 1993; Sun, Chubb, & Sperling, 2015) to generate a local motion representation. Information can then be integrated at a subsequent global motion stage where a population of direction-selective cells can encode the 2D motion direction of the whole pattern (C. Pack & Born, 2001; Stoner & Albright, 1996). The dynamics of global motion computation can be seen as the outcome of a diffusion process in both the direction domain and retinotopic space (Tlapale, Doshier, & Lu, 2015; Tlapale, Masson, & Kornprobst, 2010). For the sake of simplicity, we restricted the global stimulus representation to a continuous feature space of direction, a one-dimensional representation. A key feature of this space is that transitions between two perceptual states are dynamic shifts in the dominant peak of the distribution of activity within this population, providing a richer description than discrete jumps between separate states. Mutual inhibition was set between the underlying competing sub-populations of direction-selective cells as a centre-surround interaction kernel. This departs from the typical classical approach where transitions are modelled between two or more discrete neuronal populations, for instance with binocular rivalry (Lehky, 1995; Matsuoka,

1984) or tri-stable plaid patterns (Huguet et al., 2014). While reductions from continuous descriptions to discrete models have been shown to accurately capture the characteristics of dominance duration for binocular rivalry (Laing & Chow, 2002), the richness of representation afforded across continuous perceptual spaces can be exploited in modelling the tri-stable motion direction space (Rankin et al., 2014; Rankin, Tlapale, Veltz, Faugeras, & Kornprobst, 2013).

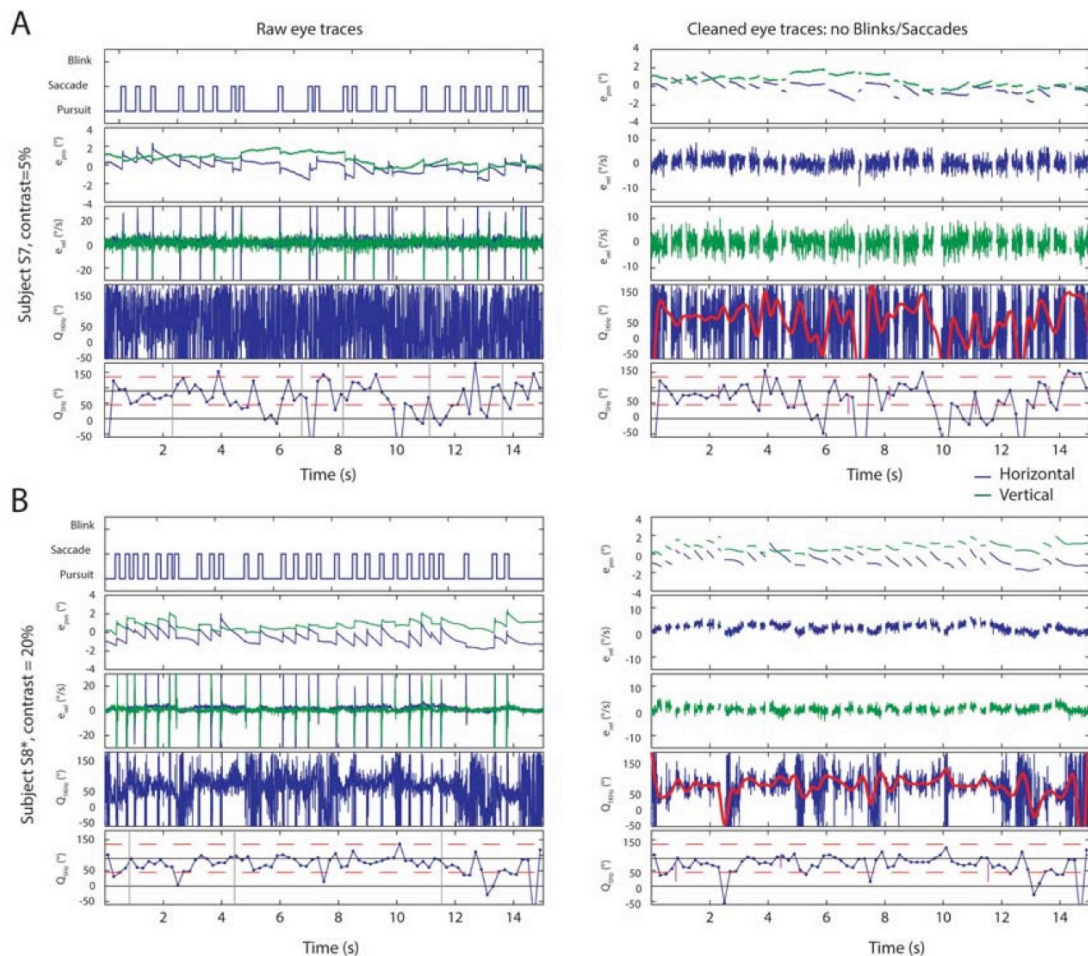


Figure 2: Raw eye movement traces and initial ‘cleaning’. (A) A single trial example of eye movement responses for *task 2*, participant S7 and contrast 5% over the 15s trial (time on the x-axes of all rows). On the left hand side column, data is in the ‘raw’ form received from the video eye tracker. First row shows validity (i.e. valid, saccade or blink based on standard criteria). Then x (blue) and y (green) position showing abrupt changes during saccades. Third row shows x and y velocities. Fourth, directions at 1000Hz and finally the same direction estimates re-binned at 5Hz for illustration of low frequency content. On the right hand side column, ‘cleaned’ data after applying a conservative removal of blinks and saccades to leave smooth responses. First a row with the x and y positions with gaps. Then x and y speeds separately with fewer abrupt changes seen than corresponding raw traces. Fourth the high resolution direction trace with gaps is not fully distinguishable from the raw trace. A smoothed trace of the underlying trend is overlaid in red for illustration. Finally the direction trace re-sampled at 5Hz to show more gradual dynamic changes in direction shown alongside the explicit perceptual reports (magenta vertical lines) made by this participant. (B) A second single trial example of raw eye movement responses for participant S8 at contrast 20% over the 15s trials, in the same format as (A). On the left hand side column, the rows shows validity with several saccades, position with the abrupt changes, x and y velocities and directions at 1000Hz and re-sampled at 5Hz. Notably, looking at the last two rows for this participant when compared with (A) demonstrates the observation we made that the variability of participant eye direction produces idiosyncratic patterns which also vary with contrast. On the right hand side column, ‘cleaned’ data is obtained with the same process described above and data is in the same format. Rows are arranged as: positions with gaps, x and y speeds separately, high-resolution direction traces (overlaid with a smoothing function in red) and then the direction trace re-sampled at 5Hz showing more gradual dynamic changes in direction and switching reports (magenta lines).

Using neural fields equations (Amari, 1977; H.R. Wilson & Cowan, 1972), a population of direction selective cells such as found in the Middle Temporal cortical area (MT), was modelled as the time varying average membrane potential $p(v,t)$ over the continuous feature space of direction (v). The initial theoretical development is fully described in our previous computational work which details the use of bifurcation analysis to tune the early model, the basic choice of physiologically plausible model parameters for Equations (8) and (9) and the full description of the choices surrounding the model input (Rankin et al., 2014). Here, the most relevant aspects of the model developed in the current work are briefly described. The tri-stable dynamical system includes adaptation $\alpha(v,t)$ and noise $X(v,t)$ terms acting across direction, v as well as a constant input term $I(v)$ which captures the competing direction cues. The main equations describing the dynamics are:

$$\tau_p \frac{d}{dt} p(v,t) = -p(v,t) + S(\lambda[J(v) * p(v,t) - k_\alpha \alpha(v,t) + k_X X(v,t) + k_I I(v) - T]) \quad (8)$$

$$\tau_\alpha \frac{d}{dt} \alpha(v,t) = \alpha(v,t) + p(v,t) \quad (9)$$

The timescales of Equations (8) and (9) are governed by τ_p and τ_α . For Equation (8), the standard decay term is $-p$ and S is a sigmoidal function with slope parameter λ and threshold T used to constrain the firing rates. The value of λ is analogous to the gain of the contrast response, as estimated using the Naka-Rushton function (Naka & Rushton, 1966). When the maximum contrast response is fixed at 1, one single parameter determines the contrast sensitivity, the half-saturation response contrast (C_{50}). Gain coefficients for the input, adaptation and noise terms are k_I , k_α and k_X respectively. The noise $X(v,t)$ used is generated by an Ornstein-Uhlenbeck process selected to allow linear transformation of the space-time variables. Lateral interactions across the direction space are set by a centre-surround interaction kernel J that is defined by three Fourier modes and a Mexican-hat shape where a local excitatory becomes inhibitory for more distant directions.

Parameter	Value	Notes
τ_p	1	ms
τ_α	16.5	S
$\lambda(c)$	[13,25]	No units
T	-0.01	“
C_{50}	40	“
k_I	0.01	“
k_α	from data	(free)
k_X	0.004	No units
$J_{\{0,1,2\}}$	-1, 1/2, 1/6	“
PT	from data	(free)

Table 1: Parameters for the neural field model of tri-stable motion integration

Values for these parameters (summarised in Table 1) were constrained based on known properties of cortical area MT from human and non human primate neurophysiology, such as direction tuning and inhibitory interaction of units (Qian & Andersen, 1994; Qian, Andersen, & Adelson, 1994; Treue, Hol, & Rauber, 2000), normalization (Carandini & Heeger, 2012) and contrast response functions (Sclar, Maunsell, & Lennie, 1990). A complete description of the extensive parameter exploration and tuning can be found in (Rankin et al., 2014). The principle behind the model construction was to set up a dynamical system that faithfully captures the key properties of visual motion integration. To achieve that aim, it was not necessary for us to have a one to one mapping between neurophysiological structures and each of the model parameters.

Equations (8) and (9) were solved using a standard Ordinary Differential Equation solver in Matlab. We used a numerical continuation package AUTO (Doedel, 1997) to perform Bifurcation analysis to look at the positions of steady state and oscillatory solutions and study the stability of these while a pair of critical stimulus and model parameters (in this case contrast and k_α) were varied (Kuznetsov, 1998). AUTO monitors the stability of states being tracked using this information to detect Bifurcations, the points at which there is a sudden change to a qualitatively different type of solution for a non linear dynamic system as a parameter is changed. Away from bifurcation points this tracking information tells us how the relative stability is

changing with respect to parameters, a measure providing a powerful predictive tool in the current context. For simplicity, this ‘*stability*’ output can be quantified by the real part of the Eigenvalue for the non-oscillating state which corresponds here to the direction D. We call this output E . In contrast, the real part of the so called Floquet exponent we term F is similarly informative about the oscillatory or H-V (cardinal direction) states (Kuznetsov, 1998). E and F quantify the timescales of growth, or decay of perturbations towards steady and oscillatory states, respectively, so states are stable if these measures are negative and become less stable when the values increase.

The model simulations presented in the current work used a constant input $I(v)$ which is a tri-modal smooth function across the continuous direction space v with a peak centred at the diagonal ($v=0$) flanked by two peaks on either side (± 45 degrees). These peaks were described by Gaussian functions $I_{1D}(v)$ and $I_{2D}(v)$ with sigma width of 18 degrees for the 1D and 6 degrees for 2D. These 1D and 2D contributions were summed to produce the input function $I(v)$. The weighting in this summation has a contrast dependence built into the 1D term.

$$I(v) = w_{1D}I_{1D}(v) + I_{2D}(v - 45) + I_{2D}(v + 45) \quad (10)$$

$$w_{1D} = 0.5 - 0.6c \quad (11)$$

The form of Equations (10) and (11) are motivated by previous behavioural experiments on both humans and macaque monkeys demonstrating the changing role of 1D and 2D cues over the time course of visual motion integration (Barthelemy et al., 2010; Barthelemy et al., 2008; Masson et al., 2000). The desired cue relationship captured by this formulation holds for the contrast range over which the current stimulus is seen as tri-stable and continuously moving i.e. below about 40%. Above this contrast, adverse motion after effects influencing the perceived speed emerge and in Equation (11) w_{1D} is therefore set to zero above $c=5/6$. This limit marks the edge of the parameter region beyond which multistability can no longer be perceived or indeed modelled. The resulting effect of contrast on input signal to noise ratio which determines the dynamic weighting of these competing cues is consistent with previous work (Lorceau & Shiffrar, 1992).

Zero noise case: Bifurcation analysis

Bifurcation analysis is run with the noise coefficient k_x set to zero and used to identify parameter regions that conform to expectations following initial psychophysics experiments. The aim is to restrict the range of values of the adaptation gain (k_a) so that the model can best account for the non-monotonic relationship between perceptual switching rate and contrast. With strong adaptation ($k_a=0.03$) the onset of switching at a critical value of the contrast is sharp, the falling phase of the switching rate is captured in this case but not the rising phase. With no adaptation ($k_a=0$) the switching rate would increase monotonically. The tuning process seeks to identify an intermediate range of values of adaptation strength (later set near $k_a=0.01$) at which the experimentally observed rising and falling phases of the switching rate curve are both possible.

Added noise: dynamic direction simulations

Simulations are subsequently run with a non-zero noise coefficient ($k_x=0.004$). The goal is to generate a continuous dynamic direction output of high temporal resolution modelling the cortical global motion representation during the competition in each trial. This can be read out in various ways and compared both to the continuous eye traces and the perceptual decisions under the range of contrast conditions. The main readout subsequently reported is a count of the number simulated percept changes between direction states, H, D or V over each of the 1500 simulated trials (used as our standard number of simulated trials) carried out per contrast value over the tested range. To count switches, a pair of threshold values are applied to the dynamic peak of $p(v,t)$ set at a distance $\pm PT$ from the diagonal. PT captures the fact that given a forced choice decision along a continuous space, participants will make a categorical forced choice decision based on boundary criteria which will vary across individuals (see also Equations 5-7; $PT \propto PT_V - PT_H$); this parameter sets this boundary between H, D and V in direction space. The model implementation assumes symmetry between H and V in the direction space though we note that the actual data shows biases across this space. It is our first critical free parameter when bringing the model into an operating regime in which it can be tied to individual participants. The second free parameter is the adaptation strength k_a . Low-level visual adaptation dynamics show some variation across individuals that is captured by this parameter. In the perceptual competition between directions, small shifts in its value impact the transition rates, allowing us to adjust the simulated rates according to individual performance. Finally, the normalised direction distributions

of the full dynamic model output $p(v,t)$ generated for the simulations over the range of contrasts give the predicted percept probability across the direction space. This can be compared directly to the eye direction distributions over corresponding contrast ranges. We test for changes in these distributions by fitting multimodal nested functions F_D , of 1-3 peaks modelled by Lorentzian functions (typically sharper than Gaussians, providing a better fit to the current data) to quantify whether the modality of these distributions (i.e. one dominant peak or multiple peaks in eye direction) changed across the tested contrast range. The general form of the fitted function is,

$$F_D(v) = C + \sum_i \frac{Amp_n \times Sig_n^2}{Sig_n^2 + (\theta - Peak_n)^2}, i \in \{1, \dots, n\}, n \in \{1, 2, 3\} \quad (12)$$

Eye direction distribution data binned into 180 bins of 2° each are fitted to Equation 12 using a nonlinear least squares fitting to find the best parameters for three versions of the function with one mode ($n=1$), two modes ($n = [1,2]$) and three modes ($n = [1,2,3]$). The fits therefore have four, seven and ten parameters respectively, so the comparison between them is done by AIC and BIC to find which function best describes the number of peaks in the distributions. This is done both for individual data distributions and those from the grouped data.

Results

Dynamic coupling between reported perceived direction and eye direction

Previous work has demonstrated a link between perception and smooth eye movements during ambiguous perception of binocular rivalry stimuli (Hayashi & Tanifuji, 2012). We tested a similar premise in the current work, analysing the smooth portions of eye movement traces during presentation of an ambiguous barberpole stimulus. We focused on the smooth phases of eye movements, which are driven by the motion stimulus. Therefore blinks and saccades causing abrupt changes in eye direction and speeds were identified and excluded from the analysis (e.g. typical individual trials Figures 2). Eye direction was estimated from the inverse tangent of the ratio of the speeds in the x- and y-directions and this nonlinear estimate is therefore highly susceptible to noise (see fourth row, left hand column in Figures 2). Once traces were cleared of the most abrupt changes in eye position and speed, smoother traces appeared with gaps (see right hand column of Figure 2). Underlying dynamic trends in eye direction can be seen in the noisy traces in the fourth row of this right hand column of Figure 2, in the thick red lines which are the result of applying a generic smoothing, for illustration only. A similar trend can also be seen in the same data sampled at 5Hz in the last rows of panels in Figure 2, where instances of button presses are also included. These example results highlight the noisy nature of the individual traces obtained, the need for appropriate filtering (which is fully detailed in the methods section) and the inherent limitations in trying to determine instantaneous motion direction perception from this dataset without prior knowledge of the instance of button presses.

In an initial exploration of the relationship between eye direction and perceptual decisions, data from *task 2* (in which participants explicitly reported eye direction) was studied. Dynamic eye direction traces were considered by plotting the averaged traces obtained combining across participants, restricted to $\pm 0.75s$ around each button press for the different reports (H-D-V), separated by contrast (Figure 3A-C). The resulting averaged traces show that when eye direction is aligned with the button press (thick vertical black line), the reported direction is related to the eye direction for all tested contrasts. By looking at this data, it is clear that there is an optimal time window within the dynamic trace (illustrated inset, Figure 3C) over which the eye direction is most indicative of the perceptual decision. It can also be seen that averaged eye traces are largely separated across the direction space so that beyond a threshold [e.g. PT_V in Figure 3B] up from the diagonal, traces will typically correspond to a perception of V. To interrogate the spread within this data, the averaged eye direction traces extending through the same range ($\pm 0.75s$ around each button press) are then binned into histograms (50-bin width) separated by decision (H-D-V) and contrast condition. The resulting distributions are shown in Figure 3D-I. The direction densities for the cardinal directions H and V are consistently seen to flank the D density in green on either side. The differences between the means and modes (inset in the figures, black and grey lines respectively) of the H and V distributions quantify how separated the peaks are and therefore show an approximate increase in separation as contrast is increased. These distributions demonstrate intuitively how the effectiveness of a decoding process depends on the distributions within the direction space, hence relate to the PT_V and PT_H parameters. They also show that we might expect different optimal values across the contrast range.

To determine the optimum eye direction window parameters across our dataset, we use the insights gained by visualising the data as described above to devise a simple three parameters decoding scheme (see Equations 5-7 and methods). The scheme computes a forced choice decision (H-D-V) for each trace based on

the window length N (in ms) of the eye direction trace and the pair of threshold parameters PT_{HV} . When this decoding scheme is applied to the data for each of the participants who did *task 2*, the optimum prediction results can be compared for the symmetrical window of eye directions extending equally before and after the button press (Figure 4A), the pre-button window (Figure 4B) and the post-button window (Figure 4C). The top row shows the optimally fitted PT values for each participant ($\propto PT_V - PT_H$) plotted against temporal window size. The parameter distribution showing the least spread when the different window configurations are compared corresponds to the pre-button window (see Figure 4D), which notably uses approximately half the eye direction information than the longer symmetric window for the decision.

For this pre-button window, the optimal duration values are seen to be distributed between 400-500ms across participants and PT_V/PT_H show a range of values which all reveal the asymmetry in the direction space for the empirical data (H-bias, see also Figures 3D-I). The complete results for optimal fitting and subsequent decoding including a comparison of the three window configurations are given by Table 2 and Figure 4H. These include the baseline performance of around 36% shown with standard deviation in horizontal blue lines of Figure 4H obtained by shuffling the response data and using them to re-assign randomised decisions. The average prediction performance of the selected pre-button window is 65% and therefore almost 30% higher than the baseline, a performance slightly higher than the alternative window configurations. We note that we were indeed able to achieve even higher prediction performance (>80%) by optimising separate classification parameters for each contrast. The resulting parameters from those simulations would be less general and such extensions are therefore beyond the aim and scope of the current work. The present results reliably indicate that changes in time averaged eye direction precede the button press by several hundred milliseconds reflecting the button press choice and therefore help elucidate the constrained window in which these two measures are most related. This window is about 484 ± 77 ms before the button press. We acknowledge that in this decoding result, we were unable to achieve the more difficult goal of reliably determining when within an eye movement trace a perceptual transition occurred. Though further work towards that goal continues, this may well also be limited by the variability of individual eye movement traces. We instead rely critically on the complementary nature of our two behavioural measures.

	<i>Symmetrical window (1)</i>				<i>Pre-button (2)</i>				<i>Post-button (3)</i>				Chance
	Duration	PT_H	PT_V	Predicted	Duration	PT_H	PT_V	Predicted	Duration	PT_H	PT_V	Predicted	
S8*	1232	28	52	69.9%	560	33	54	69.0%	672	27	57	67.4%	36.9%
S7*	642	36	62	56.9%	405	35	61	60.7%	153	41	58	54.3%	35.1%
S9*	586	35	46	59.8%	405	33	46	63.7%	363	42	49	54.5%	35.3%
S2*	1288	32	46	64.5%	560	29	46	64.3%	602	33	46	59.8%	33.0%
S6*	867	29	46	64.3%	490	30	46	64.7%	391	29	46	61.8%	37.9%
<i>M</i>	923	31,9	50,5	63.1%	484	32,1	50,4	64.5%	436	34,6	51,0	59.6%	35.6%
<i>Std</i>	325	3,4	7,1	5.0%	77	2,4	7,0	3.0%	207	6,8	6,1	5.5%	1.9%

Table 2: Best fitting parameters from the application of the prediction scheme described by Equations 5-7. Predictions are divided into the three window types and shown separately for the five participants who did *task 2*. Based on this data, 2 is chosen as the best window of prediction.

The effect of grating contrast on reported perceptual switching rates

Varying the contrast of the grating is one way of modulating the strength of the stimulus. For each value within the range of contrasts tested (3, 5, 8, 10, 15 and 20%), we plotted the average number of transitions reported per 15s trial for each one of the participants who performed the task. These are shown for both *task 1* and *task 2* in Figures 5A and 5B respectively. The increase in reported transitions as contrast rises from 3% can be seen for almost all participants and, in some of them a gentle reduction at higher contrasts can also be seen. The results from the two tasks are similar, though *task 2* was more difficult and thus showed more variability in the trends. We ask whether these plots are better fitted by a rising and saturating function or whether the gentle reduction in switching rates at higher contrasts is indeed significant. We do this by fitting all the data for the two tasks combined with either a saturating function in Equation 3, or the same function with a supersaturating term added to make it a peak function, given in Equation 4. The resulting fits are shown in Figure 5C, with the actual data shown by the points and standard error bars overlaid with the saturating (blue line) and peak (red line) function fits. We apply the model comparison BIC and AIC tests

described in the methods section on the pair of fits and find that with both measures, the peak function has a lower score for both AIC (-0.045 vs -0.025) and BIC (-0.2 vs -0.15) indicating that it is indeed a better overall description of the trends observed. This trend was obtained for the whole data set and we ask the same question for the 12 individual participant's data which show some variation even between *task 1* and *task 2* (e.g. Figure 5D), and which can also be fitted by peak and saturating functions individually (e.g. Figure 6E). We repeat the procedure, obtaining fits for both function types and then calculating the AIC and BIC comparison results for individual fits (see Figure 5F). Most traces show a lower score for the peak than saturating functions and when we consider the individual scores and directly compare them for the pair of functions, we find a significantly lower value for the peak scores in a one tailed t-test for both the AIC (means -2.48 and -1.50; $t(11) = -1.99$, $p = 0.036$) and BIC (means -2.65 and -1.64; $t(11) = -2.05$, $p = 0.032$). Therefore our results demonstrate (both at group and individual levels of analysis) that the switching rate rises fast at low contrasts, peaks at mid contrasts around 8-10% and then decreases gently – though at different rates for different participants.

Dynamic Neural Field Model construction and simulations

We now seek to extend and adapt our previously proposed model of tri-stable motion perception (Rankin et al., 2014) to the experimental observations to enable us to run simulations with perceptual switches. The cardinal and diagonal inputs, with respect to the competing directional cues within the input grating stimulus, are defined as Gaussian functions of 6° and 18° widths respectively (Figure 6A). We first used bifurcation analysis (Kuznetsov, 1998) to study the relationship between the dynamical behaviour of the underlying tri-stable neural system and two parameters: the adaptation strength (k_x) and the contrast (c). The model represents the continuous perceived direction space with the dynamic function $p(v,t)$ which peaks at just one 'winning' direction following the application of mutual inhibition across direction space, resulting in a peak which drifts across the direction space over time. The purpose of bifurcation analysis was to identify qualitatively different regions of interest in this parameter space and work within regions with dynamic properties comparable to the experimental data. The bifurcation curves plotted in Figure 6B were computed without noise ($k_x = 0$). The curves bound three parameter regions with different dynamics: in white, a low contrast regime where the system is below threshold (the input is not detected); in light grey, regular oscillations (switches) are driven by adaptation, and in dark grey, no switching is predicted in the deterministic (no noise) case. When noise is added ($k_x \neq 0$), switching in the light grey region of Figure 6B remains driven by adaptation and is therefore labelled **a**. In the dark grey region of the same figure with zero or very low adaptation strength, switches can only be driven by noise so we label this region **n**. In a parameter tuning explained in the Methods section, we identify an operating regime for the model where the switching rate relationship found in the experiments (Figure 5) is best accounted for. This zone lies near the transition between the regions **n** and **a** (thick black curve in Figure 6B, within the dotted rectangle labelled **n/a**). Within this rectangle, there is a shift in the dominant mechanisms driving the switching behaviour from noise to adaptation as contrast is increased, shown by the gradation of shading from dark to light grey.

Within the parameter region which generates the required model behaviour, we select two free parameters for the simulations – the adaptation strength k_x and the perceptual threshold PT which fine tune the switching rate functions to allow them to vary with the range of trends observed in the experiments (Figure 5A-B). PT demarcates the direction space separating what is considered D, from H and V. It defines the symmetrical distance from the diagonal at $D=45^\circ$: when the value of the peak of the dynamic function $p(v,t)$ falls below $45-PT$, direction is H (i.e. PT_H in the eye movements), and above $45+PT$, the direction is V (i.e. PT_V in the eye movements). We assume symmetry in the model even though this is not the case in the eye traces, where an H-bias can be seen. The effect this parameter has on simulated switching rates is shown in Figure 6C, where it is seen to shift the function up or down and change the steepness of the low contrast rise. The k_x parameter determines the strength of adaptation controlling its depth of modulation across the direction space. When it is finely controlled, restricting it to within the rectangle identified in the bifurcation analysis, this parameter is seen to shift the position of the switching rate peak and the extent to which there is a reduction in switching rate at higher contrasts (See Figure 6D). Finally, the stability parameters obtained during the bifurcation analysis and described in the methods section, the Floquet exponent, F (red trace), corresponding to the adaptation driven transitions and the Eigenvalue of the steady solution corresponding to the noise driven transitions, E (black trace), are plotted for the contrast range in Figure 6E. The change in stability predicted with increasing contrast can be visualised through potential wells within which gravity acts on a particle, illustrated in Figure 6F. Dominance in the D directions shifts systematically towards the cardinal directions as stimulus contrast is increased.

Using model simulations, we find the best fitting parameters (PT and k_c) for each of the 12 data sets so that we can compare the empirical and simulated switching rate functions by plotting them together (see Figure 7, compare simulated grey squares with data in black circles). We find that through variation in the two parameters, the model is able to closely capture the full range of trends reported by the participants across both tasks. Note that the range of fitted PT values in the simulations is as broad as that seen in the eye movement data of task 2, in Figure 4A-C.

Further testing of model predictions

Using the best-fitted PT and k_c parameters for all the data combined ($PT=13.125$, $k_c=0.01125$), simulations are then run across the contrast range to extract additional statistical properties of the perceptual switching. The first of these is the coefficient of variation (CV) of the time between switches for each contrast, which is calculated as the variance of percept durations divided by the average of percept durations. The results are shown in Figure 8. The model predicts a gradual reduction in CV (grey dashed trace) from 0.85 as contrast is increased from 3 to about 10% before the value plateaus at around 0.65, which is maintained for the rest of the contrast range. A similar trend is observed for the experimental data (black circles) with the standard error of CV across participants given by the error bars. This trend is driven by a shift from highly variable percept durations for the low contrast noise driven transitions, towards slightly less variable durations during the adaptation dominated transitions at higher contrasts. The model predicts a slightly higher variability than the experiments in the low contrast range where noise dominates, but model and data converge at the mid and higher contrasts.

The consequences of the shift in stability predicted in Figures 6E-F as contrast is increased were tested. The rationale is that the most stable parts of the continuous direction space illustrated as the bottoms of potential wells would act as attractors in the system and as a result peaks would occur in the dynamic representation of perceived direction. This is first shown with the simulations in Figure 9A, where a distribution of the dynamic value of the peak of $p(v,t)$ obtained from 1500 simulations of 15s each are plotted for three contrasts: 3, 8 and 15%. It can be seen in this prediction that there is a clear transition from a unimodal distribution centred on the diagonal direction (black dashed trace) towards bimodal distributions which increase in separation as contrast is increased (grey dashed lines). To test this prediction with the experimental data, we use eye movements recorded from *task 2* and separate out eye directions obtained for H, D and V button presses restricted a critical 450ms time window before the button press and plot the peak direction (and inter quartile range as error bars) for the contrast range tested (see Figure 9B). What we see is a progressive increase in the separation of peaks as contrast is increased, consistent with the predicted increase in stability, the same trend as the simulations but asymmetrically skewed towards the H direction in the empirical data. The use of the peak of the $p(v,t)$ function was necessary because when the full continuous function is considered unrestricted to the 450ms 'decision window', the resulting distribution is broader and does not clearly show the separate underlying peaks (see Figure 9C, obtained for the same simulations as 9A). Because eye direction in the current experiment was a continuous empirical measure, these simulated distributions were analogous to those produced when all the eye direction data from both tasks were plotted combining all the H-D-V button presses indiscriminately, in Figure 9D. The eye data distributions no longer show clearly discernible peaks, but get progressively wider with increasing contrast.

The last challenge was to quantify the shift in these distributions of eye directions with no clearly discernible peaks (see individual examples in Figures 10A). This is done by fitting the distributions with a Lorentzian function in Equation (12) which can have one, two or three superimposed peaks (see methods for details). We compare the modality of the group fits in Figure 10B using AIC and BIC tests, which give a lower score for the better model of the distribution. The AIC scores favour the one-peak fit (Figure 10C) at lower contrasts but shift towards the 2 and 3 peak fits (3-peak shown in Figure 10D) at higher contrast, seen in the AIC scores at 3, 8 and 20% contrasts which are -6.74 , -5.64 and -5.33 for one peak and -6.10 , -6.17 and -7.73 for the three peak function. Very similar numbers are seen for the BIC test, confirming that low contrasts (<8%) are better modelled as unimodal distributions dominated by the D direction and higher contrasts (>10%) tend towards a multimodal distribution with dominant cardinals but consistently present diagonal directions.

We ran the same test on individual participant data to confirm that the trend is consistent across individual participants, and not driven entirely by single individuals. The participant eye direction distributions are much noisier than the grouped distributions (compared Figures in 10A with 10B) and so some data cases across the six contrast range could not be reliably fitted by Equation 12. For this reason, participant data from S1, S3 and S2* were excluded from the individual data Table 3. The table shows AIC

and BIC values for each participant, separated across different rows for one, two and three mode fits for the six contrasts along the columns. The best fit from each group of three is highlighted in grey (light for unimodal and dark for multi-modal fits). At the bottom of the data tables, the count of the cases under each modality for low (3 & 5%), mid (8 & 10%) and high (15 and 20%) contrasts cases are shown. These counts show a general shift from unimodal, towards multimodal (both bimodal and trimodal) distributions systematically as contrast is increased. The individual data supports the conclusions from the group data that contrast drives a shift towards multistability, with both bi-modal and trimodal distributions increasing with contrast.

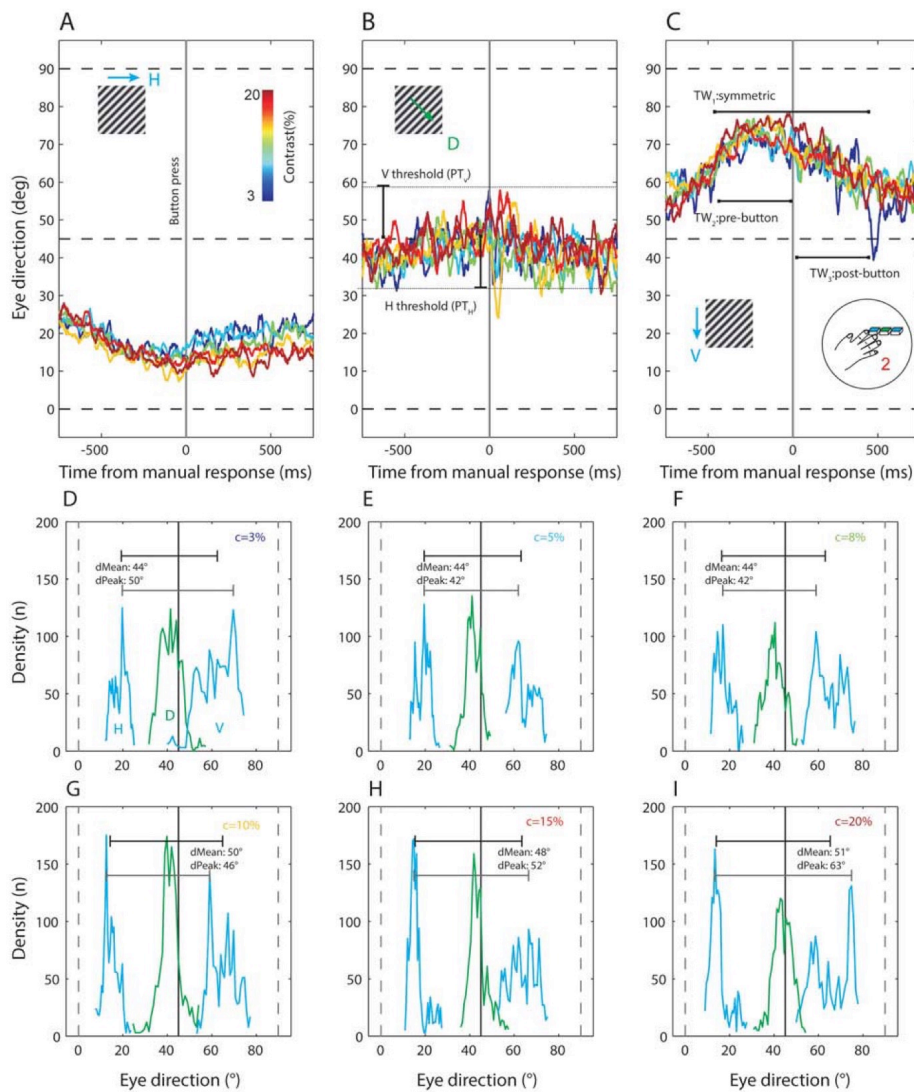


Figure 3: *Dynamic averaged eye direction traces and distributions from task 2 restricted to ± 750 ms from each button press:* (A) Six traces averaged across all participants and colour coded for the range of contrast (see colour bar inset) for the cases where H-direction was selected, with eye direction ($^{\circ}$) on the y-axis and time in ms on the x-axis. The vertical black line at $t=0$ is the button press. We see that there is a maximum downward deviation from the stimulus diagonal (45°) at the button press. (B) Cases where D was selected in the same format as H above. The traces lie near the dashed line, more below than above it, indicating a slight H-bias consistently seen in our recorded eye directions. The horizontal and vertical thresholds are indicated: these are proposed as a constraint for a scheme predicting perception from eye direction. (C) Cases where V was selected, in the same format as the two above. Traces show maximum upward deviation from the diagonal direction at $t=0$ for all traces. Inset, three temporal windows of consideration spanning symmetrically before and after [1], only before [2] and only after [3] the button press are illustrated as additional constraints for predicting perception from eye direction. (D) These are also plotted as distributions across 50 histogram bins in the direction space. The density traces for cases where reported directions were in the cardinal direction i.e. button presses H and V, are in blue while the diagonal is in green. The separation between the means and the peaks of the distributions are shown above each trace. For the lowest contrast of 3%, a broad diagonal peak overlapping with the vertical is seen which has the lowest distance between the means of H and V distributions. (E) Distributions for 5% contrast. (F) Distributions for 8% contrast. (G) Distributions for 10% contrast, where the H and D peaks appear to be sharper and further apart. (I) Distributions for 15% contrast, similar to those for 10% contrast. (J) Distributions for 20% contrast, the highest contrast. The peaks of the cardinal components are prominent and furthest away from each other for this contrast. This data is indicative of a systematic underlying relationship between eye direction and perception, though it cannot be characterised fully here as these distributions are based on averaged traces (1500 samples) around the button presses H, D or V.

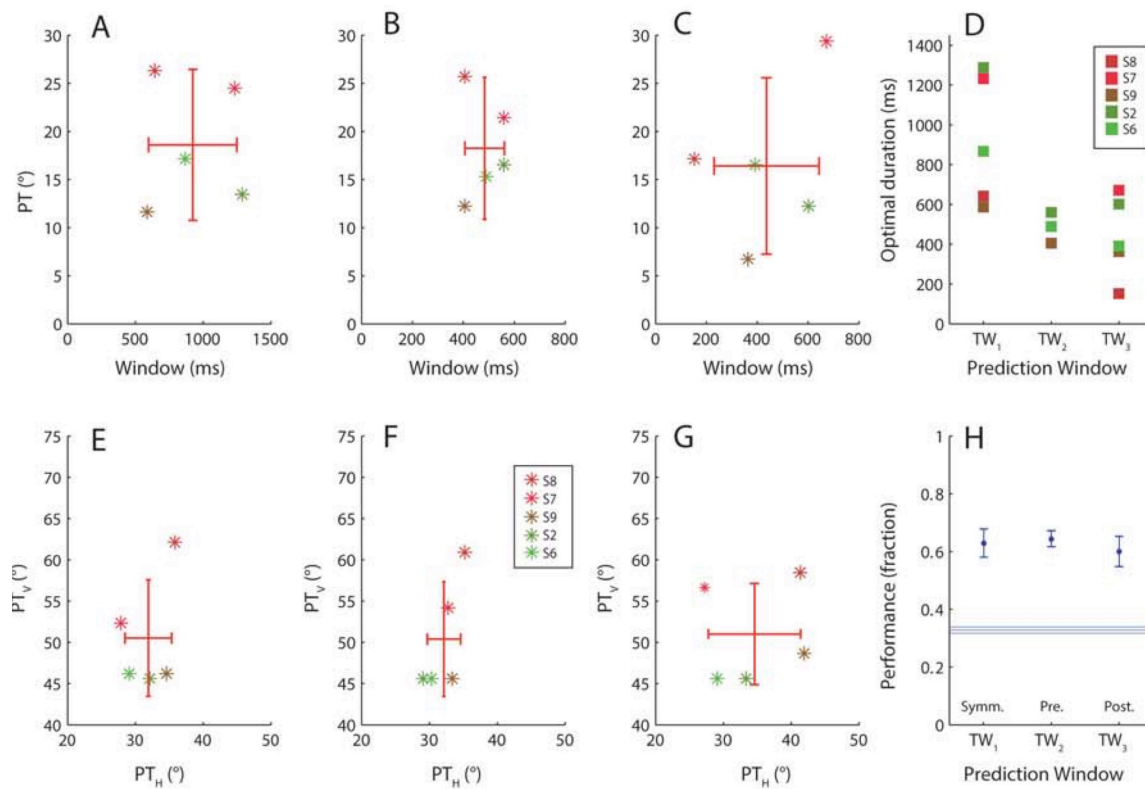


Figure 4: *Optimal prediction parameters linking perceived direction to eye direction in task 2 and prediction performance.* (A) The Perceptual Threshold (PT) obtained by combining the best PT_H and PT_V is plotted against window duration, first for the symmetrical window for the five participants. The optimal window distribution on the x-axis is broad for this case. (B) Similar to A, PT against window duration for the pre-button window. The spread along both axes is narrowest for this case. (C) The post-button case shows the largest distribution of optimal PT parameters. (D) Re-plotting all three cases to compare the optimal durations shows that for the pre-button window (TW_2) optimal windows seem to fall consistently between 400-500ms. (E) PT_V against PT_H first for the symmetrical window then, (F) the pre-button window and (G) the post button window. Most cases have best fitted thresholds biased towards the horizontal direction, considering 45° as the diagonal. (H) The fraction of correct predictions of button presses for all participants based on their individual optimal parameters is plotted for each of the three prediction window types. The baseline chance performance based on choice shuffling is shown in the horizontal blue lines which show the mean and standard deviation across participants. The pre-window in the centre of the plot returns the best performance.

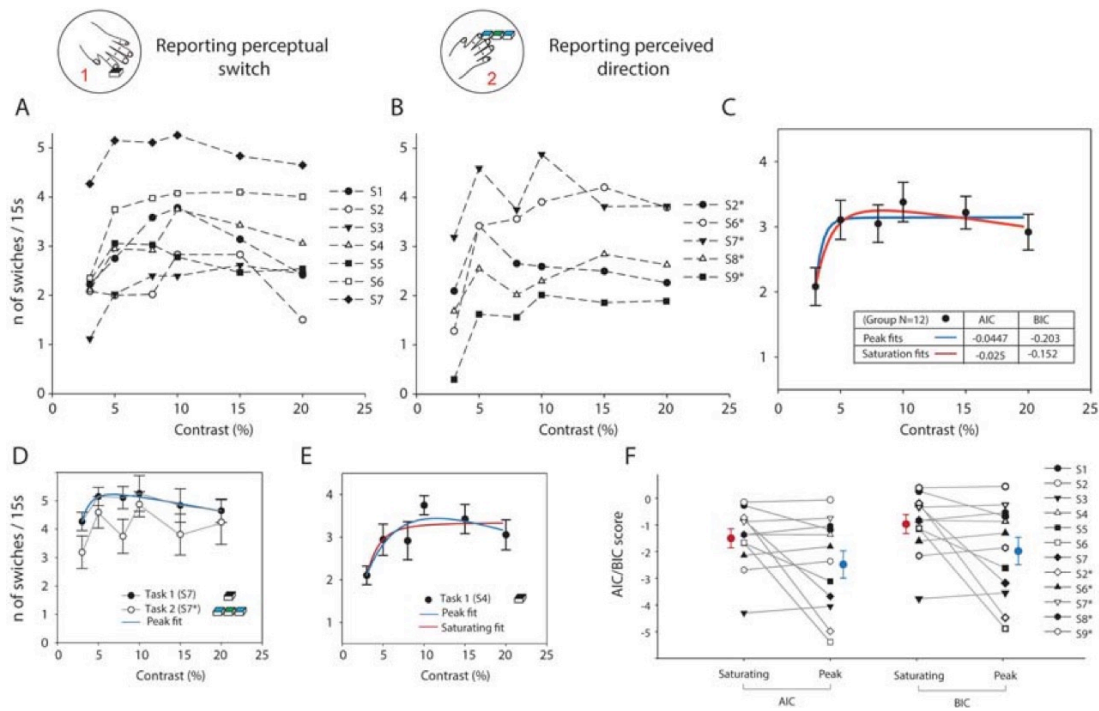


Figure 5: Transition rates from perceptual reports plotted against contrast. (A) Average number of reported transitions per trial for the 7 participants performing the simplified *task 1* plotted with different markers showing the variation in individual patterns of perceptual transitions. (B) For the more difficult *task 2*, average reported transitions per trial for the 5 participants, three of whom also performed the simplified task, are plotted with different markers showing trends generally similar to (A). (C) The data from (A) and (B) are combined and fitted with a saturating (blue line) and a peak (red line) function for comparison. Both functions fit the data. A model comparison using both the Akaike and Bayesian Information criteria (numbers inset) find that the peak function provides a better fit to the data even when corrected for the extra parameter. Scores are shown inset: more negative scores indicating a better model. (D) Examples of data for the two tasks for the same participant S7, *task 1* (black circles) and *task 2* (white circles). The solid black line shows the best fitting peak function for *task 1*. (E) Perceptual reports for participant S4 doing *task 1*, showing both the peak (black line) and saturating (grey line) fits. (F) A comparison of the AIC (left pair of points) and BIC (right pair) scores for individual participants' switching rate curves. The peak function value is compared to the saturating function and points include data from the 9 participants and 12 sessions. The larger points show the mean and standard error of the scores, returning with better performance for the peak function.

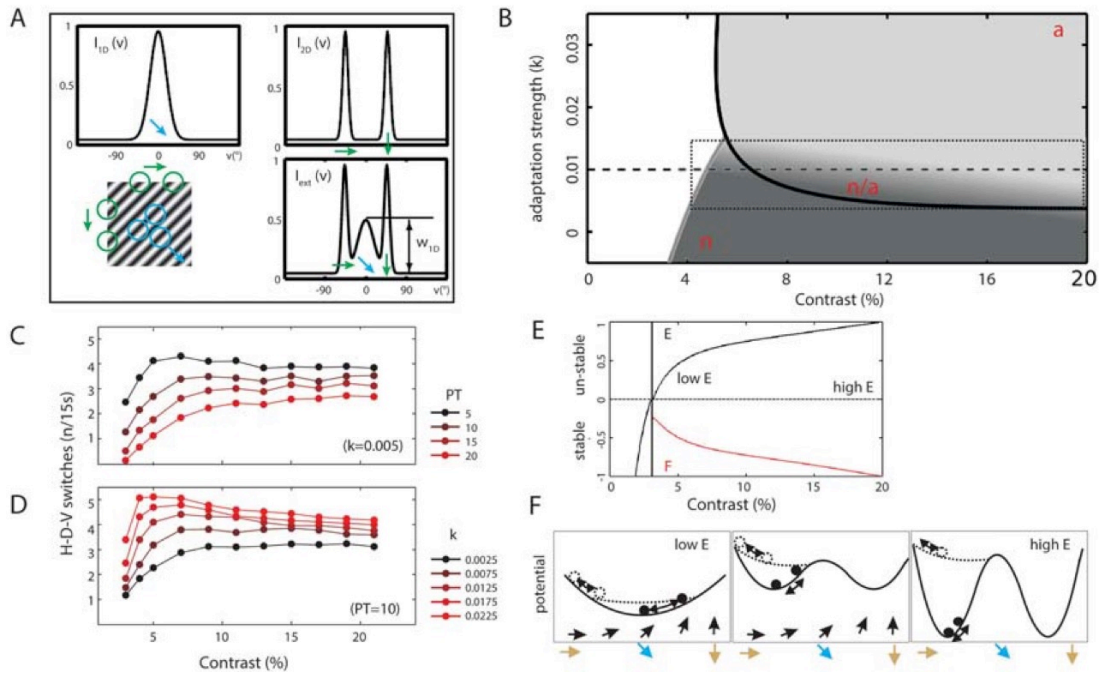


Figure 6: *Dynamic neural fields model structure and characteristics.* (A) 1D motion cue model input (top left) which is a Gaussian distribution centred on the diagonal in the direction space. 2D motion cue input (top right) with two narrower Gaussians in the cardinal directions. Inputs are the result of stimulation with the stimulus illustrated (bottom left) with gold apertures indicating the edge locations with 2D cardinal direction signals and blue apertures indicating interior 1D motion signals subject to the aperture problem. These 1D and 2D apertures have a resultant shown by their linear combination (bottom right) used as the model input. (B) A bifurcation diagram exploring the changing dynamics as adaptation (y-axis) and contrast (x-axis) are varied in the model. The thick continuous bifurcation lines constrain regions with different dynamics, with the white region on the left being below threshold, the lighter grey on the top right marked **a** showing adaptation driven oscillatory dynamics and the darker grey region on the bottom right marked **n** showing above threshold but steady-state responses which can only show transitions in the presence of noise. The model is operated in the region demarcated by the dotted rectangle in which noise and adaptation drive transitions. (C) Simulations of number of switches per 15s period with 1500 trials per trace generated by the model with added noise for a range of parameter values of PT which is the threshold between different directions. The parameter mainly shifts the whole curve vertically. (D) Simulations generated as in (C) for a range of 5 values of k_{α} . This parameter characterises the strength of adaptation, thus determining the transition from noise dominance to adaptation dominance and the prominence of the peak of the switching rate function. (E) The Eigenvalue E (black trace) and the Floquet exponent F of the oscillating state (H-V, red trace) are both normalised to a magnitude of 1 and plotted against contrast. The stability of the perceptual states driven by noise ($E \approx 0, F \approx 0$) and adaptation ($E \approx 1, F \approx -1$) in the model can be compared by the changes in the values of these terms. For the given system, increasing contrast makes changes driven by adaptation more stable, whilst reducing the effect of noise. (F) Changing stability of states illustrated through oscillations of elements moving under gravity-like acceleration determined by the local gradient within the dynamic potential wells. The contrast relationship given in (E) means that under this model, there will be primarily noise driven transitions due to local fluctuations at low contrasts (left) with systematic shifts towards adaptation driven transitions at higher contrasts (mid- and then right, with adapted wells in the dashed lines).

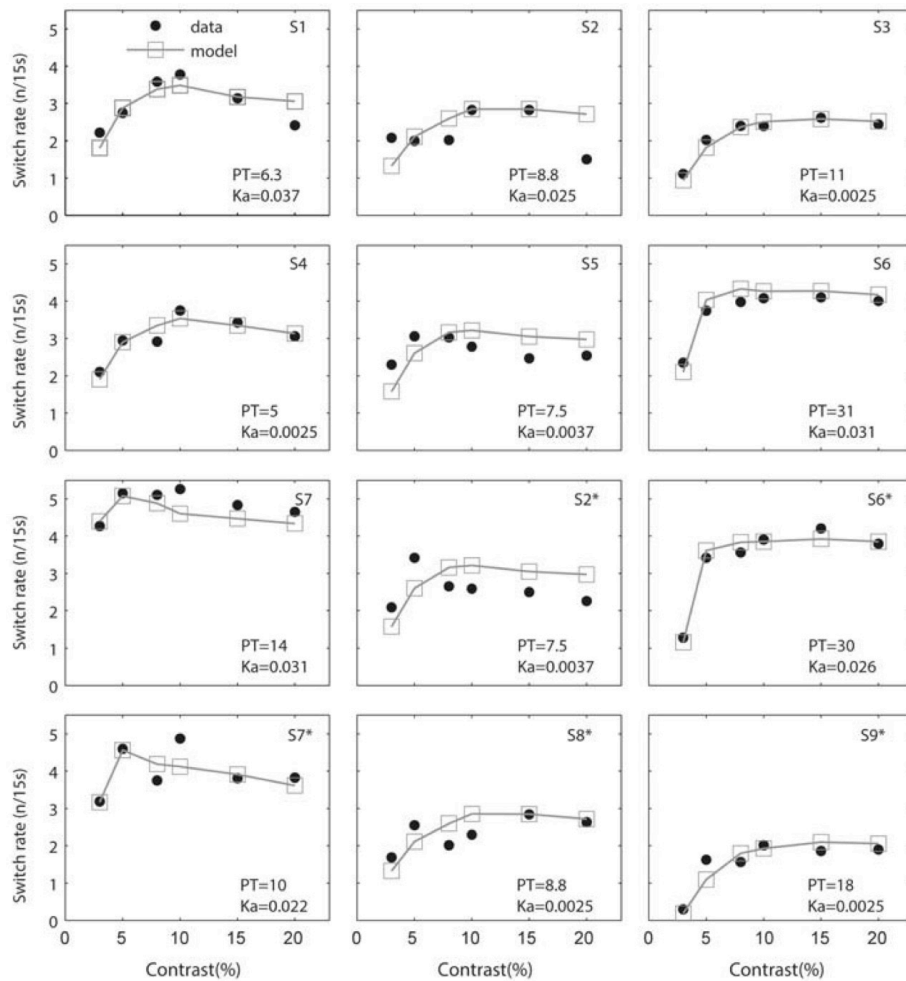


Figure 7: Comparing experimental and model non-monotonic curves of switching rates. Average switching rates per 15s trial (y-axes) are plotted against stimulus contrasts for all the data collected (black circles) alongside corresponding model simulations (open grey squares). Model simulations are run for the best estimates of parameters PT and k_alpha for each participant, which are shown inset. (A)-(G) Show fits for the data from seven participants obtained under *task 1*. The model makes good fits except where the data is not smoothly varying, e.g. B and E. (H)-(L) Shows data collected from the five participants under *task 2*. These trends show a little more variability than those under *task 1*. Notably, for participants S2 (compare B,H), S6 (compare F,I) and S7 (compare G,J) who performed both tasks, parameter estimates of PT and k_alpha remain within comparable values under both tasks, suggesting that they capture idiosyncratic observer characteristics.

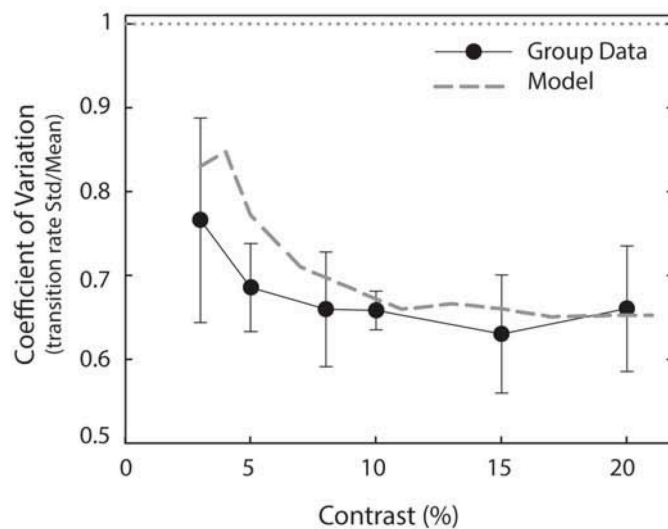


Figure 8: *Model prediction of switching duration variability compared to experiments:* The coefficient of variation (CV, y-axis) calculated from the standard deviation of the durations divided by the mean duration and calculated and then plotted for each contrast (x-axis). The mean and standard deviation from all experimental data, where percept duration could be reliably determined (black circles), shows a gentle reduction in this value consistent with the trend of model predictions (grey dashed line). While model predictions are for slightly higher values in the 3-8% contrast range, the trends over this range are consistent.

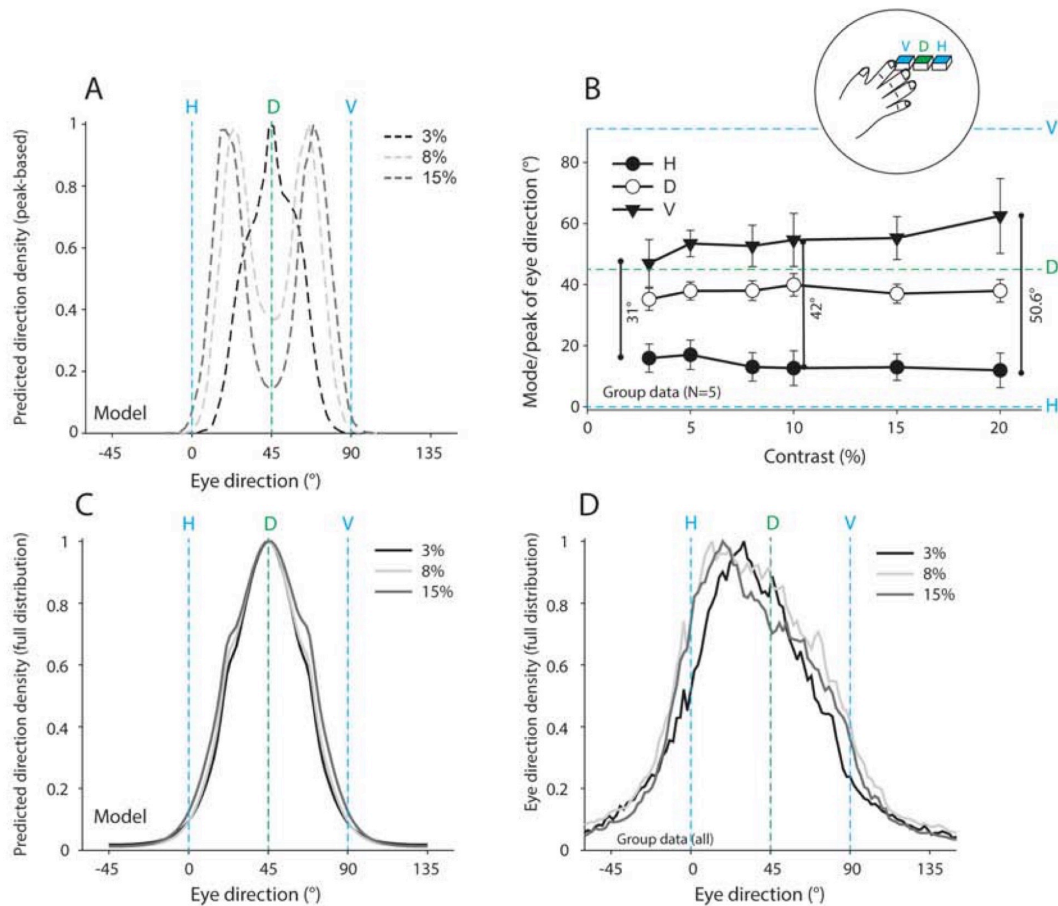


Figure 9: Changes in model and experimental direction distribution with contrast: Normalised distributions of dynamic responses obtained from model simulations of the output $p(v,t)$ accumulated over 1500 trials are compared to distributions obtained from eye movement samples during the experiments. (A) Model direction distributions obtained based on the peak of the dynamic output $p(v,t)$ at three contrasts (3, 8 and 15%). Simulations over 1500 trials capture the shift from uni-modal through tri-modal towards bi-modal distributions over the contrast range. The peak positions systematically shift away from the diagonal. (B) To compare simulated direction distributions based on the peak of the dynamic functions $p(v,t)$ to experimental eye movements, we use eye movement data from each reported transitions combined to obtain averaged traces of 1.5s, symmetrically spanning before and after button presses. A histogram of this average serves as a measure of the peak of the perceptual distribution of directions. These distributions of eye direction from *task 2*, are separated into H, D or V reported cases, and the mode and inter-quartile ranges from the eye directions are plotted for H (filled circles), D (unfilled circles) and V (triangle) directions. There is a contrast dependent shift in the mode consistent with the simulations: cardinal states become more prominent at higher contrasts. (C) Broader distributions are obtained when the full neural fields model output $p(v,t)$ is used to produced direction distributions resulting in less prominent separate peaks for three simulated contrasts. There are only subtle changes in the width of the distributions as contrast is increased, with broader less Gaussian-like distributions at higher contrasts. Therefore by shifting from the peaks (A) to the entire simulated distributions, we obtain predictions that might be expected with noisy neural representations of direction. (D) Eye direction distributions from all participant data in both tasks, after cleaning and basic filtering (see text for details). Included samples are restricted to a 450ms temporal window before the button press, a window found to be closely linked to the reported perceived direction (see text for details). For three contrasts 3% (black lines), 8% (light grey) and 15% (dark grey), the lowest contrast trace looks unimodal while higher contrasts are broader, asymmetric and with a less well defined shape. This is consistent with the simulations of (C), additional modes within the possibly composite distributions cannot be easily identified without further analysis.

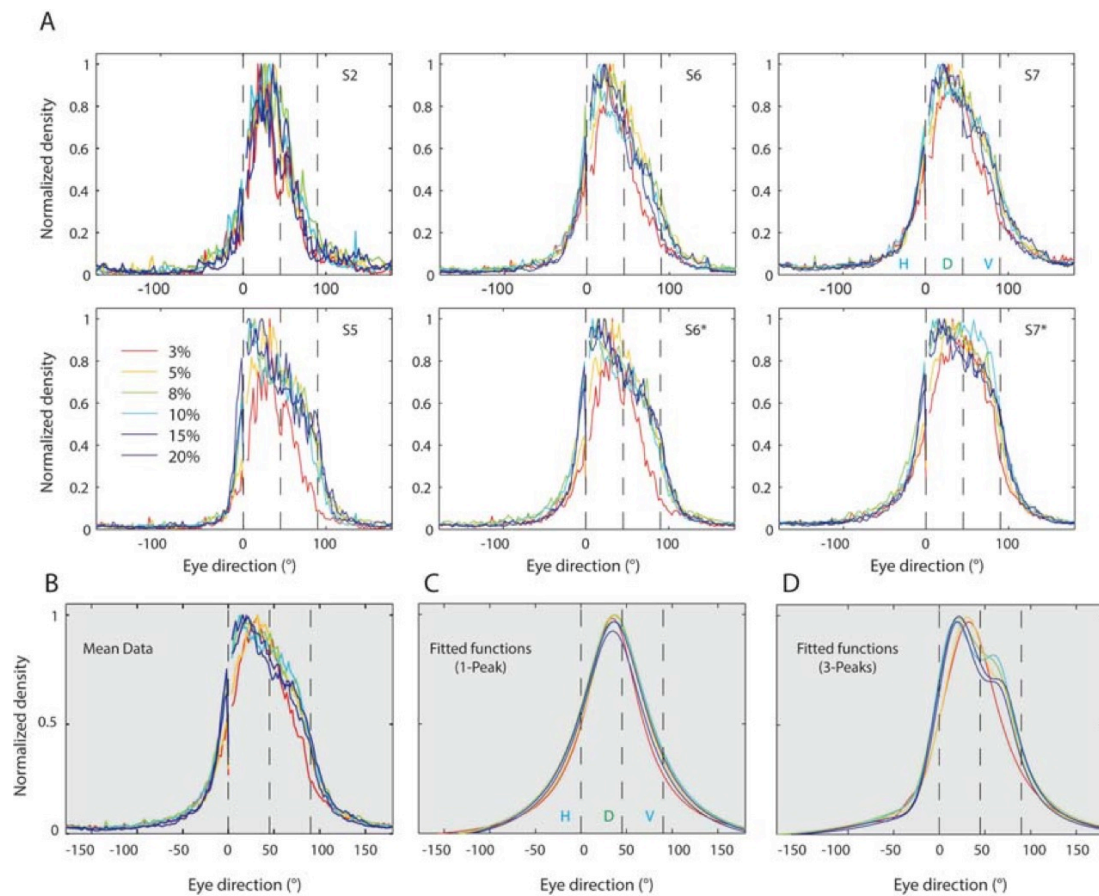


Figure 10: *Characteristics of individual and grouped eye movement distributions:* Normalised eye movement distributions across the direction space (x-axes) for the six contrasts tested, each plotted with the range of colours shown in the key inset within the first panel. Eye direction samples are extracted exclusively within the 500ms temporal window preceding each button press this is the window most informative of perceptual state. (A) to (D) show distributions obtained from *task 1*, all consistently showing narrower distributions with clearer defined peaks for the lowest contrasts 3 and 5% (red and yellow traces) than for traces at higher contrasts. The higher contrast cases (i.e. darker blue/purple) appear broader and typically shifted towards the H-direction indicative of a bias observed during the experiments. (E) to (F) Show distributions obtained under *task 2*, for the same two participants S6 and S7 as (B) and (C) for comparison. The trends are similar for all traces, though there is some idiosyncratic patterns of variability. (G) The averaged, normalised distributions including contributions from all data collected over all experiments and all six contrasts are plotted. Resulting traces have smoother trends than individual participant data which allows us to fit nested multi-peak functions and evaluate whether they tend towards multi-modality. (H) Least squares-fitted single peak functions based on the Lorentzian function (see text for details), showing little difference between traces in terms of peaks and widths. (I) Least squares-fitted triple peak Lorentzian function, showing a stronger role for the additional peaks at higher contrasts than at lower contrasts. This shift towards multiple modes is quantified using AIC and BIC criteria in the text.

Discussion

With most of the well known multi-stable displays like binocular rivalry and structure-from-motion it remains difficult to disentangle the role of intrinsic dynamics of low and middle level hierarchical visual processes (H.R. Wilson, 2003). The aperture problem in motion perception is a constraint that results in ambiguous visual input due to the properties of low-level visual motion detection mechanisms (Nakayama & Silverman, 1988; Wallach, 1935). Recent computational models have proposed competing mechanisms between motion energy and feature mechanisms to explain perception in a barber pole within moving edges (Sun, Chubb, & Sperling, 2014; Sun et al., 2015). Since neurons detecting motion locally have small receptive fields, the perceived direction of elongated internal edges (the 1D signals) is ambiguous. The direction of local 2D features affords more precision but takes slightly longer (Masson et al., 2000). As a consequence, it takes time to compute the perceived motion direction of such a barber-pole stimulus (Masson et al., 2000) and dynamic competition between three directions, the oblique (D), the horizontal (H) and vertical (V) persists and yields perceptual multi-stability (Castet et al., 1999; Castet & Zanker, 1999; Meso & Masson, 2015; C. C. Pack, Gartland, & Born, 2004).

Here, we studied the dynamics of switches in perceived direction while varying the input strength of the tri-stable barberpole motion stimulus presented over a 15 second duration. Participants reported each switch in perceived direction during the trial, extending beyond typically studied motion integration timescales (<1s). Simultaneously recorded eye movements provided a second behavioural probe. We sought to tease apart the contributing roles of noise and adaptation previously proposed to drive multi-stability for binocular rivalry (Shapiro et al., 2009; van Ee, 2009). The change from bi-stable to tri-stable stimuli in empirical work has previously elucidated history effects which emerge within the three choices (Naber, Gruenhage, & Einhauser, 2010) and over longer presentation epochs of over a minute, progressive shifts between tri-stable and bi-stable states have been suggested (Hupe & Pressnitzer, 2012; Wallis & Ringelhan, 2013). The presence of three alternatives has also importantly enabled a dissociation between alternative driving mechanisms (Huguet et al., 2014). We exploited all these advantages in the current work, particularly seeking to dissociate contributions of noise and adaptation.

Subj ID	Modes	(AIC)						(BIC)					
		CONTRAST						CONTRAST					
		3	5	8	10	15	20	3	5	8	10	15	20
S2	One	7.15	6.81	7.27	6.74	6.68	10.21	8.07	8.00	7.18	6.92	6.52	10.30
	Two	8.82	7.95	6.81	6.55	6.29	10.24	8.04	6.94	6.07	6.93	6.69	10.38
	Three	9.76	8.85	7.04	7.42	7.14	10.25	11.79	8.78	7.54	7.90	7.03	10.44
S4	One	7.26	8.30	8.66	8.51	9.14	10.06	6.84	8.33	8.80	8.73	9.07	10.15
	Two	6.92	8.52	9.44	8.95	7.82	8.65	6.85	8.30	8.52	9.25	8.01	8.79
	Three	8.77	8.92	10.68	8.47	7.79	7.79	9.68	9.33	10.44	9.00	7.82	7.98
S6	One	12.59	9.92	12.33	9.74	10.11	10.83	11.85	9.09	14.01	9.68	10.24	10.91
	Two	11.68	9.20	10.13	9.46	10.15	9.19	12.72	9.83	10.27	10.33	10.59	9.33
	Three	9.39	9.61	11.54	10.70	10.89	9.22	11.05	10.14	11.94	10.63	11.35	9.42
S6*	One	7.01	8.97	10.81	11.03	11.19	15.27	9.05	9.12	9.71	10.66	11.29	15.36
	Two	7.51	10.29	13.07	10.16	10.95	11.83	8.55	9.41	14.34	9.64	11.37	11.97
	Three	8.63	10.64	10.92	11.15	10.61	11.86	8.83	11.06	11.05	11.22	10.84	12.05
S7	One	8.93	9.81	11.48	11.64	15.14	12.14	8.61	9.72	10.81	11.72	15.22	12.22
	Two	7.93	13.22	9.74	11.08	11.65	12.17	7.27	13.08	10.87	11.53	11.95	12.31
	Three	11.12	11.63	12.08	12.15	11.87	9.77	10.12	11.54	12.57	12.06	11.95	9.96
S7*	One	9.22	15.71	11.29	11.70	12.04	12.45	9.21	13.89	11.05	11.90	12.24	12.54

	<i>Two</i>	8.63	9.46	10.63	11.94	12.15	12.48	7.37	11.84	10.11	11.50	12.06	12.62
	<i>Three</i>	11.48	18.63	11.97	12.92	12.74	9.73	10.02	19.14	12.02	13.19	12.67	9.92
<i>S8*</i>	<i>One</i>	8.63	9.82	12.06	12.29	12.05	12.82	9.25	11.49	11.71	11.81	12.50	12.91
	<i>Two</i>	9.08	10.19	11.61	14.99	12.30	11.87	9.51	9.48	11.70	14.66	11.99	12.01
	<i>Three</i>	13.40	13.17	13.06	13.41	12.71	12.88	11.63	13.32	13.58	13.72	12.80	13.08
<i>S5*</i>	<i>One</i>	9.59	9.99	11.67	12.70	16.45	13.29	9.86	10.37	12.79	12.28	16.94	13.38
	<i>Two</i>	7.82	10.99	11.65	11.72	16.60	13.16	9.49	9.97	11.55	11.88	16.41	13.31
	<i>Three</i>	11.87	12.36	13.57	13.77	13.78	12.40	11.16	13.35	13.94	13.34	13.77	12.59
<i>S9*</i>	<i>One</i>	9.05	15.34	11.85	16.45	12.92	13.71	9.68	15.45	11.76	17.16	13.22	13.80
	<i>Two</i>	8.66	10.22	11.06	13.03	13.22	12.87	9.48	10.67	11.08	12.20	13.20	13.01
	<i>Three</i>	11.46	19.06	11.50	14.42	13.87	12.90	11.78	18.91	11.74	13.73	14.02	13.10
			low	mid	high			low	mid	high			
<i>Group</i>	<i>One</i>		9	4	5			6	5	3			
<i>totals</i>	<i>Two</i>		8	13	6			11	12	6			
	<i>Three</i>		1	1	7			1	1	7			

Table 3: Individual values of the AIC and BIC model comparisons applied to the eye direction distributions from the experiments (see text for details). Each group of three points compares fits with one, two and three modes, and the lowest value in each 3-value column is highlighted in grey. Light grey highlighting indicates better performance for unimodal fits while darker grey indicates better performance for multimodal fits. The last three lines summarise the data by counting the number of instances of each type of preferred fit for comparison across the contrast range separated into low (3 & 5%), mid (8 & 10%) and high (15 & 20 %). Participants S1, S3 and S2* are excluded because reliable fitting of direction distributions across the contrast range was not possible. We note that model comparison on the individual data was generally noisy given the model complexity of 4-10 parameters across the modality comparison.

The temporal resolution of forced-choice reported perceptual decisions of several seconds enabled us to characterise the switching dynamics at a coarse scale. For finer grain consideration, we combined both empirical measures analysing the discrete, perceptual responses aligned to the continuous, high-resolution ocular responses centred on each button press. We did not explicitly instruct participants to ‘follow’ the stimulus but rather, we designed an interrupted fixation task in which short epochs of small, seemingly automatic tracking responses were elicited. From the instantaneous eye direction, we were able to study both transitions between perceived global motion states and the fluctuations within each possible state to probe the underlying stability of the perceptual states. Using the subset of data where perceived directions were explicitly reported as H, D or V, we initially demonstrated that the smooth parts of recorded eye movements and perceptual reports of transitions were tightly coupled and informative of the perceptual states during tri-stable motion perception. This coupling was strongest within a time window of 480ms before the button press, extending previous findings seeking to link eye movements to motion perception (Hayashi & Tanifuji, 2012; Price & Blum, 2014; Quaia et al., 2016; Spering & Montagnini, 2011). Our findings provide evidence that changes in ocular behaviour precede the perceptual report. In the context of 2D motion integration, within such a temporal window of optimal coupling between the behavioural measures, we could further test the statistical properties of the eye movements recorded during the tasks. Unlike some of the previous work, our demonstration of a coupling between the measures did not allow us to determine when within a given eye movement trace a perceptual switch occurred. The noise within the recorded individual traces made that a particularly difficult challenge, which remains to be tackled in future.

The key experimental result was that as grating contrast was increased from a near threshold value, we obtained a non-monotonic function of switching rates against contrast. At low contrast, there was an initially rising limb – similar to that proposed by Levelt IV in binocular rivalry experiments (Levelt, 1966) and consistent with previous studies (van Ee, 2009). As the contrasts further increased however, the same

function reached a maximum switching rate and then, surprisingly a small gradual reduction for increasing contrasts, less consistent with Levelt's IV was observed. We verified this key result with several fitted model comparison tests. When testing nested saturating and peak functions as descriptors of the trend, we consistently found that the peak function with a gentle reduction at higher contrasts provided better fits for the experimental data. While non-monotonic responses with changing input strength have been suggested in theoretical work describing binocular rivalry as a two state dynamical system (Seely & Chow, 2011; Shpiro et al., 2007), these have been in the opposite direction to the current findings with an *inverted u-shape* when percept durations are considered, which results in a *u-shaped* trend of switching rates against contrast. Three percept options therefore add to transition complexity, creating different classes of competing transitions (cardinal-cardinal and cardinal-oblique), which cannot be fully described by Levelt's IV. In a previous consideration of a subset of the current psychophysics data in which directions are explicitly reported, we showed that a relative decrease in oblique direction prevalence and conversely an increase in reports of cardinal directions occurred as contrast was increased (Meso & Masson, 2015). The results presented here suggest that the non-monotonic function emerges from that interaction of oblique and cardinal directions.

To directly study the underlying mechanisms, a neural fields model of the dynamic system was used. The characteristics of similar models have previously been studied computationally by either injecting noise (Freeman, 2005; Salinas, 2003) or using adaptation driven oscillatory mechanisms (Lago-Fernandez & Deco, 2002; Lehky, 1988) both of which had a limited scope for suitable empirical verification. The current stimulus was defined as having competition along a continuous direction space which could be configured with variable contributions of noise and adaptation to the tri-stability (Rankin et al., 2014). A cortical population based on Middle temporal motion area (MT) was assumed to encode the perceived direction with a set of basic properties such as directional tuning, lateral inhibition across the neural population encoding the direction space, nonlinear contrast responses and normalization. Using bifurcation analysis, the model was tuned such that parameters were identified which brought it into an operating regime at a boundary region between two types of dynamics, noise and adaptation-driven. This critical operating regime has been identified in previous theoretical work (Shpiro et al., 2009; Theodoni et al., 2011).

Our model simulations were able to reproduce these empirical non-monotonic functions with just two free parameters corresponding to a constant controlling adaptation strength and the individual participant decision thresholds separating out the direction space. The first parameter finely controls the adaptation-driven average duration of persistence. The second sets the thresholds for switching away from the diagonal direction within the continuous space. The effect of both was to finely tune the balance of noise and adaptation to each participant. Linking the key modelling results to the previously noted systematic shift in proportions of participant reports from dominant oblique directions towards cardinal directions with contrast increments (Meso & Masson, 2015), we now directly associate noise dominance with 1D motion cues and adaptation dominance with 2D (cardinal direction) motion cues in the stimuli.

The model was able to make further predictions on stability and switching duration statistics, going beyond the classical switching rate analysis. These predictions were tested against the experimental results. An increasing stability of cardinal direction states was predicted during the shift from noise-driven to adaptation driven regimes. This could be seen through corresponding changes in the mode of the eye direction distributions. Indeed, there was a systematic shift from a predominantly uni-modal (D centred) towards a bimodal (cardinal centred) or sometimes tri-modal distributions in the histograms of the eye directions from smooth eye movements during the tasks. In a second prediction, the coefficient of variation from the inter-transition period in simulations showed a reduction to a plateau of this measure of variability as contrast was increased and this trend was verified experimentally. Overall, our neural fields model provided a very good description of the underlying competition mechanisms and characterised the respective role of noise and adaptation in both perceptual stability and transition dynamics for different levels of input strength.

Conclusion

Our novel results empirically demonstrate that perceptual transitions are governed by both the signal strength (which determines the Signal to Noise Ratio, SNR) and adaptation. Their relative contribution depends upon SNR so that when the contrast is high, the driving effect of adaptation is increased and the noise has relatively reduced effect. Conversely, with a weaker input the noise is relatively large and therefore makes a bigger contribution to driving the transitions between alternative percepts. In the present barberpole stimulus, the noise regime was dominated by the diagonal percept at the centre of the continuous direction space, which is also the vector average of the alternative directions. In contrast, binocular rivalry involves

transitions between two entirely discrete percepts. Here we have shown that contrast increments drive an increase in the stability of the cardinal states over this continuous direction space at the cost of the diagonal, resulting in a non-monotonic switching rate-contrast relationship, contravening propositions made for the less complex competition space of binocular rivalry (Levelt, 1966). Thus tri-stability within the continuous direction space is not fully analogous to the well-studied area of binocular rivalry owing to the two different classes of stable states, oblique and cardinal directions. Continuous low-level variables are however ubiquitous in visual space e.g. orientation, depth and speed, and ambiguities in these spaces could benefit from dynamic modelling with similar schemes. Eye movements were exploited here to explore the dynamics of mechanisms underlying transitions within the small critical window 480ms before button presses, and such a probe could be exploited in other stimulus contexts. These findings complement the recent suggestion that during tri-stable perception noise and adaptation are responsible for the timing and actual choice, respectively, of each transition (Huguet et al., 2014). Indeed the stochastically variable timing/duration of each percept was determined by the noise, but the shift towards slightly more regular switching observed at higher contrasts was consistent with our proposal of a shifting balance between noise and adaptation. Tri-stable motion perception was well described as a dynamical system through which simulations with a high number of trials served as invaluable tools to establish signatures of underlying dynamical mechanisms. With antagonistic and opponent interactions ubiquitous in neural processing, identifying such signatures particularly the stability of states could reveal underlying mechanisms in other ambiguous perception contexts.

Acknowledgments

We would like to thank the volunteer participants for their time. We thank Manuel Vidal, Jean-Michel Hupé and the anonymous reviewers for useful comments and constructive criticisms of earlier versions of the manuscript, and Fred Barthélémy and Anna Montagnini for discussions during experiments and eye data analysis. AIM was supported by the Aix-Marseille Université Foreign Post-doctoral Fellowship 2011, and ANR grants (Visafix, ANR-1D-blanc-1432 and SPEED ANR-13-SHS2-0006) to GM. JR, OF and PK were partially funded by the ERC grant no 227747 (NERVI) and the EC IP project FP7-ICT-2011-8 no. 318723 (MatheMACS). GM, OF, PK and JR acknowledge funding by the EU FP7 grant no 269921 (BRAINSCALES).

References

- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523-525.
- Akaike, H. (1981). Likelihood of a Model and Information Criteria. *Journal of Econometrics*, *16*(1), 3-14.
- Amari, S. (1977). Dynamics of pattern formation in lateral inhibition type neural fields. *Biological Cybernetics*, *27*, 77-87.
- Barthelemy, F. V., Fleuriet, J., & Masson, G. S. (2010). Temporal dynamics of 2D motion integration for ocular following in macaque monkeys. *J Neurophysiol*, *103*(3), 1275-1282.
- Barthelemy, F. V., Perrinet, L. U., Castet, E., & Masson, G. S. (2008). Dynamics of distributed 1D and 2D motion representations for short-latency ocular following. *Vision Res*, *48*(4), 501-522.
- Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10), 1-21.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.
- Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51-62.
- Castet, E., Charton, V., & Dufour, A. (1999). The extrinsic/intrinsic classification of two-dimensional motion signals with barber-pole stimuli. *Vision Research*, *39*, 915-932.
- Castet, E., & Wuerger, S. M. (1997). Perception of Moving Lines: Interactions Between Local Perpendicular Signals and 2D Motion Signals. *Vision Research*, *37*, 705-720.
- Castet, E., & Zanker, J. M. (1999). Long-range interactions in the spatial integration of motion signals. *Spatial Vision*, *12*, 287-307.
- Doedel, E. J. (1997). Nonlinear numerics. *Journal of the Franklin Institute-Engineering and Applied Mathematics*, *334B*(5-6), 1049-1073.
- Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035-1045.

- Freeman, A. W. (2005). Multistage model for binocular rivalry. *Journal of Neurophysiology*, *94*(6), 4412-4420.
- Hayashi, R., & Tanifuji, M. (2012). Which image is in awareness during binocular rivalry? Reading perceptual status from eye movements. *Journal of Vision*, *12*(3).
- Huguet, G., Rinzel, J., & Hupe, J. M. (2014). Noise and adaptation in multistable perception: Noise drives when to switch, adaptation determines percept choice. *J Vis*, *14*(3).
- Hupe, J. M., & Pressnitzer, D. (2012). The initial phase of auditory and visual scene analysis. *Philos Trans R Soc Lond B Biol Sci*, *367*(1591), 942-953.
- Hupe, J. M., & Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res*, *43*(5), 531-548.
- Kuznetsov, Y. (1998). *Elements of Applied Bifurcation Theory* (Second Edition ed. Vol. 112): Springer.
- Lago-Fernandez, L. F., & Deco, G. (2002). A model of binocular rivalry based on competition in IT. *Neurocomputing*, *44*, 503-507.
- Laing, C. R., & Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *Journal of Computational Neuroscience*, *12*(1), 39-53.
- Lehky, S. R. (1988). An Astable-Multivibrator Model of Binocular-Rivalry. *Perception*, *17*(2), 215-228.
- Lehky, S. R. (1995). Binocular rivalry is not chaotic. *Proc Biol Sci*, *259*(1354), 71-76.
- Lehky, S. R., & Blake, R. (1991). Organization of binocular pathways: modeling and data related to rivalry. *Neural Computation*, *3*, 44-53.
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn Sci*, *3*(7), 254-264.
- Levelt, W. J. M. (1966). The alternation process in binocular rivalry. *British Journal of Psychology*, *57*(3/4), 225-238.
- Lisberger, S. G. (2010). Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron*, *66*(4), 477-491.
- Logothetis, N. K., & Schall, J. D. (1990). Binocular motion rivalry in Macaque monkeys: Eye dominance and tracking eye movements. *Vision Research*, *30*, 1409-1419.
- Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin*, *130*(5), 748-768.
- Lorenceanu, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, *32*, 263-273.
- Lorenceanu, J., Shiffrar, M., Wells, N., & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, *33*, 1207-1217.
- Masson, G. S., & Perrinet, L. U. (2012). The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neurosci Biobehav Rev*, *36*(1), 1-25.
- Masson, G. S., Rybarczyk, Y., Castet, E., & Mestre, D. R. (2000). Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Vis Neurosci*, *17*(5), 753-767.
- Matsuoka, K. (1984). The Dynamic Model of Binocular Rivalry. *Biological Cybernetics*, *49*, 201-208.
- Matsuoka, K. (1985). Sustained Oscillations Generated by Mutually Inhibiting Neurons with Adaptation. *Biological Cybernetics*, *52*(6), 367-376.
- Meso, A. I., & Masson, G. S. (2015). Dynamic resolution of ambiguity during tri-stable motion perception. *Vision Research*, *107*, 113-123.
- Meso, A. I., Montagnini, A., Bell, J., & Masson, G. S. (2016). Looking for symmetry: fixational eye movements are biased by image mirror symmetry. *Journal of Neurophysiology*, jn 01152 02015.
- Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol*, *98*(3), 1125-1139.
- Naber, M., Gruenhage, G., & Einhauser, W. (2010). Tri-stable stimuli reveal interactions among subsequent percepts: Rivalry is biased by perceptual history. *Vision Research*, *50*(8), 818-828.
- Naka, K. I., & Rushton, W. A. (1966). S-potentials from luminosity units in the retina of fish (Cyprinidae). *J Physiol*, *185*(3), 587-599.
- Nakayama, K., & Silverman, G. H. (1988). The aperture problem - II- Spatial integration of velocity

- information along contours. *Vision Research*, 28, 747-753.
- Niemann, T., Ilg, U. J., & Hoffmann, K. P. (1994). Eye-Movements Elicited by Transparent Stimuli. *Experimental Brain Research*, 98(2), 314-322.
- Pack, C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409, 1040-1042.
- Pack, C. C., Gartland, A. J., & Born, R. T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *Journal of Neuroscience*, 24(13), 3268-3280.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437-442.
- Price, N. S. C., & Blum, J. (2014). Motion Perception Correlates with Volitional but Not Reflexive Eye Movements. *Neuroscience*, 277, 435-445.
- Qian, N., & Andersen, R. A. (1994). Transparent Motion Perception as Detection of Unbalanced Motion Signals. II. Physiology. *Journal of Neuroscience*, 14, 7367-7380.
- Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent Motion Perception as Detection of Unbalanced Motion Signals. I. Psychophysics. *Journal of Neuroscience*, 14, 7357-7366.
- Quaia, C., Optican, L. M., & Cumming, B. G. (2016). A Motion-from-Form Mechanism Contributes to Extracting Pattern Motion from Plaids. *The Journal of neuroscience*, 36(14), 3903-3918.
- Rankin, J., Meso, A. I., Masson, G. S., Faugeras, O., & Kornprobst, P. (2014). Bifurcation study of a neural field competition model with an application to perceptual switching in motion integration. *J Comput Neurosci*, 36(2), 193-213.
- Rankin, J., Tlapale, E., Veltz, R., Faugeras, O., & Kornprobst, P. (2013). Bifurcation analysis applied to a model of motion integration with a multistable stimulus. *Journal of Computational Neuroscience*, 34(1), 103-124.
- Salinas, E. (2003). Background synaptic activity as a switch between dynamical states in a network. *Neural Computation*, 15(7), 1439-1475.
- Schwarz, G. (1978). Estimating Dimension of a Model. *Annals of Statistics*, 6(2), 461-464.
- Sclar, G., Maunsell, J. H. R., & Lennie, P. (1990). Coding of Image-Contrast in Central Visual Pathways of the Macaque Monkey. *Vision Research*, 30(1), 1-10.
- Seely, J., & Chow, C. C. (2011). Role of mutual inhibition in binocular rivalry. *Wallace, J. M.*, 106(5), 2136-2150.
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619-626.
- Shapiro, A., Curtu, R., Rinzel, J., & Rubin, N. (2007). Dynamical characteristics common to neuronal competition models. *J Neurophysiol*, 97(1), 462-473.
- Shapiro, A., Moreno-Bote, R., Rubin, N., & Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *J Comput Neurosci*, 27(1), 37-54.
- Spering, M., & Montagnini, A. (2011). Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: a review. *Vision Res*, 51(8), 836-852.
- Stoner, G. R., & Albright, T. D. (1996). The Interpretation of Visual Motion: Evidence for Surface Segmentation Mechanisms. *Vision Research*, 36, 1291-1310.
- Sun, P., Chubb, C., & Sperling, G. (2014). A moving-barber-pole illusion. *Journal of vision*, 14(5), 1.
- Sun, P., Chubb, C., & Sperling, G. (2015). Two mechanisms that determine the Barber-Pole Illusion. *Vision research*, 111(Pt A), 43-54.
- Theodoni, P., Panagiotaropoulos, T. I., Kapoor, V., Logothetis, N. K., & Deco, G. (2011). Cortical microcircuit dynamics mediating binocular rivalry: the role of adaptation in inhibition. *Front Hum Neurosci*, 5, 145.
- Tlapale, E., Doshier, B. A., & Lu, Z. L. (2015). Construction and evaluation of an integrated dynamical model of visual motion perception. *Neural Networks*, 67, 110-120.
- Tlapale, E., Masson, G. S., & Kornprobst, P. (2010). Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, 50(17), 1676-1692.
- Treue, S., Hol, K., & Rauber, H.-J. (2000). Seeing multiple directions of motion - physiology and psychophysics. *Nature Neuroscience*, 3, 270-276.
- van Ee, R. (2009). Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: evidence from serial correlations in perceptual bistability. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 26(12), 2612-2622.

- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192-196.
- Wallach, H. (1935). Ueber visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung*, *20*, 325-380.
- Wallis, G., & Ringelhan, S. (2013). The dynamics of perceptual rivalry in bistable and tristable perception. *Journal of Vision*, *13*(2).
- Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences USA*, *100*(24), 14499-14503.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1-24.
- Zhou, Y. H., Gao, J. B., White, K. D., Merk, I., & Yao, K. (2004). Perceptual dominance time distributions in multistable visual perception. *Biol Cybern*, *90*(4), 256-263.