Perceptual multistability predicted by search model for Bayesian decisions

Rashmi Sundareswara Department of Computer Science and Eng., University of Minnesota, MN, USA

Paul R. Schrater Departments of Computer Science and Eng and Psychology, University of Minnesota, MN, USA

Perceptual multistability refers to the phenomenon of spontaneous perceptual switching between two or more likely interpretations of an image. Although frequently explained by processes of adaptation or hysteresis, we show that perceptual switching can arise as a natural byproduct of perceptual decision-making based on probabilistic (Bayesian) inference, which interprets images by combining probabilistic models of image formation with knowledge of scene regularities. Empirically, we investigated the effect of introducing scene regularities on Necker cube bistability by flanking the Necker cube with fields of unambiguous cubes that are oriented to coincide with one of the Necker cube percepts. We show that background cubes increase the time spent in percepts most similar to the background. To characterize changes in the temporal dynamics of the perceptual alternations beyond percept durations, we introduce Markov Renewal Processes (MRPs). MRPs provide a general mathematical framework for describing probabilistic switching behavior in finite state processes. Additionally, we introduce a simple theoretical model consistent with Bayesian models of vision that involves searching for good interpretations of an image by sampling a posterior distribution coupled with a decay process that favors recent to old interpretations. The model has the same quantitative characteristics as our human data and variation in model parameters can capture between-subject variation. Because the model produces the same kind of stochastic process found in human perceptual behavior, we conclude that multistability may represent an unavoidable by-product of normal perceptual (Bayesian) decision-making with ambiguous images.

Keywords: Multistability, Bayesian inference, perceptual decision-making, Markov Renewal Process, context.

Introduction

The human visual system is remarkably good at producing crisp and consistent perceptions of the world. Underlying these crisp and consistent percepts are sophisticated methods for resolving sensory ambiguity and conflicts, many of which have been successfully described using Bayesian inference (Kersten, Mamassian & Yuille, 2004). However, we experience a breakdown of this consistency when viewing stimuli that produce phenomena like perceptual multistability and binocular rivalry. The question motivating this study is to what extent perceptual multistability can be viewed as resulting from the process of perceptual decision-making based on Bayesian inference, without recourse to special mechanisms or processing.

Many theories have emphasized the role that non-inferential explanations like neural adaptation may play in generating multistability (Attneave, 1971; Blake 1989; Lehky 1988; Long, 1992 and Taylor 1974). However, recent evidence points to both bottom-up and top-down influences on multistability (Leopold and Logothetis, 1999). Top-down influences include attention (e.g. van Ee, Noest and Brascamp and van den Berg, 2006;Rock, Hall and Davis, 1994; Pelton and Solley, 1968; Peterson and Gibson, 1994), intention (Toppino, 2003;Peterson, Harvey and Weidenbacher, 1991), recognition (Peterson and Gibson, 1994), semantic content (Davis, Schiffman and Greist-Bousquet, 1990;Walker 1978) and voluntary control (Hol, K., Koene, A. and van Ee, R.,2003, van Ee, van Dam and Brower , 2005; Toppino, 2003; Meng and Tong, 2004, van Ee, 2006; van Ee, 2005). These many and diverse influences suggest multistable perception may be better viewed as resulting from a general process of perceptual (Bayesian) decision-making than a specific low-level mechanism. In particular, it has been suggested that multistability may be an extreme case of ambiguity in normal perceptual processing (van Ee et al, 2003) where several different percepts are equally probable.

If multistability results from an extreme case of ambiguity in normal perceptual processing, then manipulations that reduce ambiguity in normal vision should increase perceptual stability when viewing multistable stimuli. One way to reduce ambiguity involves placing an ambiguous image in a global context that promotes a perceptual interpretation. Previous studies have shown that moving multistable stimuli can be stabilized by the motion of background elements (Dawson 1987;Kramer and Yantis 1997). Similar manipulations for stationary rivalrous stimuli increase dominance times (Blake and Logothetis, 2002; Alais and Blake, 1999; Sobel and Blake, 2002 and Yu and Blake, 1992).

In this paper we investigate the effects of background (context) on the dynamics of perceptual bistability while viewing the Necker cube. From an ecological or Bayesian

perspective, background context can provide information about the frequency of occurrence of attributes in an image. Because the ambiguity in the Necker cube can be thought of as viewpoint ambiguity—the dominant percepts are consistent with two viewpoints of a single cube (Mamassian, et al 1998), we introduced background context that carried viewpoint information. In particular, our context stimuli consist of fields of unambiguously oriented cubes whose orientations are consistent with one of the dominant Necker cube interpretations. These background objects flank a central Necker cube (without overlap). Similar to previous effects of context on rivalrous stimuli (Sobel and Blake, 2002) we find context lengthens the durations of the supported percepts.

To better quantify the effects of context on the dynamics of multistability, we developed a new approach to the analysis of multistability data. We developed a general non-parametric descriptive framework for analyzing multistability data based on Markov Renewal Processes. The approach describes perceptual switching in terms of a discrete state variable which undergoes probabilistic transitions after random times in a state. The approach overcomes several limitations of previous analysis methods. It can capture the temporal dynamics of the switching process, contingencies between states and durations and provides tools to relate different measures of perceptual behavior. We provide a brief exposition of the key ideas and mathematical properties of the approach in the supplementary section, "Markov Renewal Processes for Bistability."

We use our analysis to reveal the dynamics of multistability in the presence of context. We found non-trivial dynamics that were modulated by context, with strong biases toward one perceptual interpretation on initial viewing that partially dissipate in subsequent viewing, similar to those reported by Mamassian et al (2005) for rivalrous stimuli. In addition, we show how the analysis approach can be used express relationships between measures, predict co-variaton between the means and variabilities of percept durations, and to test stationarity (the property of stationarity measures whether or not the probability distributions governing perceptual state transitions vary across time).

Lastly, we show how multistability could result from the normal dynamics of perceptual Bayesian inference and decision-making. Similar to van Ee et. al, 2003 we assume that multistability arises when perceptual inference produces ambiguity that can be described by a multi-modal posterior distribution on the parameters of interest (e.g. surface slant, viewpoint, etc). The novelty of our approach was to provide a simple but specific method of implementing a perceptual decision that can account for the origin of the temporal dynamics in multistability. In particular we show that searching for good interpretations by sampling a multimodal posterior, together with memory decay which values recent samples in times more than older samples, results in multistable-like behavior similar to that exhibited by participants. We believe that treating multistablity within the context of a general approach to perception (Bayesian inference) has the potential to reconcile many of the diverse results observed for multistable perception.

Methods

Task

Participants continuously viewed a Necker cube for 10sec intervals while fixating at a point in the center of the cube (Figure 1). Perceptual states were elicited via button press at semi-random intervals by the sound of a beep. A trial defined as a single event of an arrow press at the sound of a beep to indicate the orientation of the current perception of the Necker cube (see Figure 1). We refer to the two dominant interpretations in terms of user's viewpoint with respect to the Necker cube – 'V_A' indicates the percept was consistent with the *viewpoint* of the subject was *above* the cube in 3D space (resulting in the top surface perceived as in front) and 'V_B' indicates that the percept was consistent with a *viewpoint below* the cube and hence viewing the bottom surface. At the sound of a beep, if the user was currently experiencing the V_A percept, they were instructed to press the 'up' arrow key and if they were experiencing the V_B percept, they were to press the 'down' arrow key. If any percept other than V_A or V_B appeared, they were instructed not to press any key.

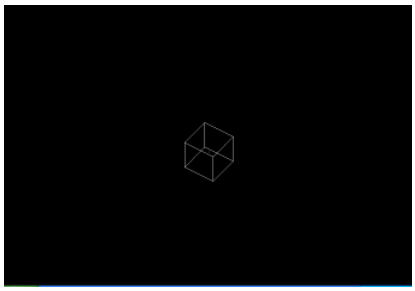


Figure 1: Necker Stimulus used in the experiment. Participants fixated the central green dot and reported their perceptual state via key-press when cued by the sound of a beep.

Stimuli and Apparatus

The Necker cube was rendered in OpenGL and oriented at 45 degrees away from the x-axis and 55 degrees away from the y-axis such that vertically and horizontally it subtended a visual angle of 3.16 degrees. Participants were directed to maintain fixation on a tiny green dot rendered in the center of the cube. The user's chin rested on a chin rest fixed in the center of the 24.2 cm high monitor at a distance of 71cm.

Training

All participants ran in the following training sessions before the main experimental data was collected. A training session consisted of the following (in the order given):

1) 40 repetitions of V_A/V_B key association practice where a single unambiguous cube would be presented in the middle of the screen and the subject was asked to indicate with a key press what percept they were experiencing.

2) 40 repetitions of a task designed to help participants discriminate between orientations around the x-axis. We refer to this as the "orientation discrimination" task. Two columns of 100 unambiguously oriented cubes appeared as shown Figure 2 and Figure 3. The two columns differed in orientation from the Necker Cube in the x-axis by random offsets with opposite (randomly-chosen) sign (e.g. +6 and -3.5 degrees). The offsets were sampled from a Gaussian distribution with a mean of 7 degrees and a standard deviation of 1. The user was asked to indicate which column appeared more fronto-parallel.

3) The Necker cube was presented for 1 minute. The user was asked to passively focus on the green dot in the middle of the screen. No beeps were sounded and no key presses were required.

4) The Necker cube was presented in the middle of the screen for 20 seconds. Beeps were generated using a Gaussian distribution with a mean of .5 seconds and a standard deviation of a 100 milliseconds after registering the previous response. The subject was asked to quickly respond with a key press at the sound of the beep, which resulted in an average inter-response interval of 1.2sec. Five 20sec blocks were collected from each subject, which constituted the *base-rate* data for the *no-context* condition.

No time limits were enforced in the training session. The session was usually completed in 6-7 minutes.

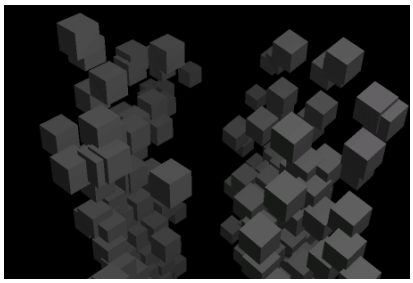


Figure 2: An example of an orientation discrimination task for unambiguously oriented cubes in V_A orientation. Subjects are asked to indicate which column was more fronto-parallel.

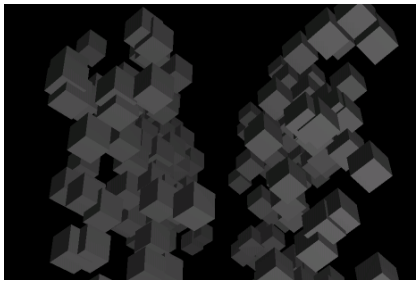


Figure 3: An example of an orientation discrimination task for unambiguously oriented cubes in V_B orientation. Subjects are asked to indicate which column was more fronto-parallel.

Context experiment

The experiment consisted of 100 *blocks* where each block consisted of 10 *trials*. During the duration of a block, the Necker cube appeared in the middle of the screen and was flanked by two columns of 100 unambiguously oriented cubes on either side of the cube as shown in Figure 4and Figure 5, in the exact conditions of the orientation discrimination task described in the training session. The orientations of the cubes in the two columns were either both V_A or both V_B. The orientation of the cubes for a block was randomly permuted such that at the end of 100 blocks, 50 of them were V_A oriented and 50 of them were V_B oriented. In a block, the beeps were generated from a Gaussian distribution with a mean of .5 seconds and a standard deviation of a 100 milliseconds (similar to the training session). It took an average of .5 seconds for the user to respond to a beep (by pressing an arrow key). Therefore, the duration of a block lasts for 10-11 seconds as 10 events of arrow presses are recorded for every block. If the subject took longer than 1.2 seconds to respond to the beep, a double beep was sounded and the block restarts. A 2-3 minute break was provided at the end of 20 blocks.

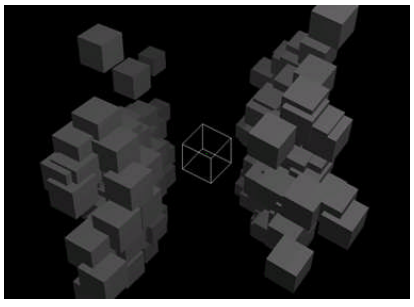


Figure 4: An instance of the V_A context stimulus.

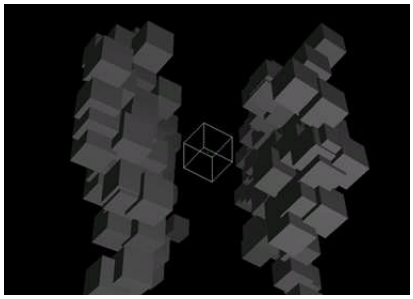


Figure 5: An instance of V_B context stimulus.

Participants

21 participants ran in three 1-hour experimental sessions scheduled over three days. All had normal or corrected-to-normal vision. Participants were either financially compensated or given extra-credits points in an introductory Psychology class.

Analysis

Data from multi-stability experiments are composed of sequences of durations spent in alternating states. A key characteristic of this data is switching between discrete perceptual states with stochastic phase durations between switch events. Estimating the distribution of phase durations is an essential part of analyzing bistability data. However, it is not sufficient to simply histogram the phase durations to model these distributions. To do so would ignore the possibilities that phase duration distributions may depend on previous perceptual states and may change across time. Our goal was to develop analysis methods that could more completely capture the temporal dynamics of perceptual switching behavior.

As pointed out by Mamassian and Goutcher (2005), to capture the temporal dynamics of perceptual switching it is not sufficient to measure only response probabilities or phase durations, because these measures are insensitive to temporal changes in the distribution of phase durations. To rectify this, Mamassian and Goutcher introduced three additional measures – transient preference, reversal probabilities and survival probabilities to capture the time-varying dynamics of perceptual switching. We also measure survival probabilities, but use a non-parametric estimation method (they fit cumulative Gaussian functions). In addition, we measure cumulative transition distribution functions (these are related to the inverse of phase duration) separately for each state. Both these measures are conditioned on particular times (rather than averaged across all occurrences in the sequence) and are conditioned on perceptual states as described below. Ultimately, these measures were chosen because of their simple relationships to a general mathematical model of probabilistic switching processes called Markov Renewal Processes, as described below in this section and in the supplementary section, "Markov Renewal Processes for Bistability." First we describe the measures, and subsequently explain how they were estimated from the data.

Preliminaries

Our data consists of sequences of response/time pairs. For the k^{th} block of responses, let $D^k=\{(r_1,t_1),(r_2,t_2),\ldots,(r_N,t_N)\}$, where r_j denotes responses, encoded as 0 for V_A and 1 for V_B, and t_j represents a recorded response time measured from the start of the response period. We also use superscripts to indicate the block number of a measurement, (e.g. t_j^k represents the j^{th} measurement from the k^{th} block. All of our measures were conditioned on *starting times*, defined for each sequence as:

$$\tau_0^k = \min_j \{t_j^k \mid r_j^k = 0\}$$

$$\tau_1^k = \min_j \{t_j^k \mid r_j^k = 1\}$$

In words, a starting time is the first time a state is observed within a response block.

In general, all measures were computed using the following steps. Responses in each block were classified according to the event type of interest (e.g. whether a transition has occurred). Interpolation is used to construct a time-continuous indicator function from the classified events separately for each block. Estimates of the desired probability functions are generated from averages of indicator functions across blocks.

Estimating Survival and Response probabilities

Survival probabilities are defined as the probability of observing the same state i after an elapsed time t:

$$P(s(\tau+t)=i \mid s(\tau)=i)$$

Note that survival probabilities do not encode the time spent continuously in a state. In the elapsed time t, there may have been any number of state transitions away from i, as long as the state returns i to by time t.

Estimating survival probabilities is complicated by the unavailability of continuous time measurements of perceptual states. To estimate survival probabilities at each time t, we interpret the response/time sequences as discrete samples from the continuous survival probability function. We classify response events for each block based on whether the initial state was V_A or V_B and linearly interpolate between response times. Let $S^k(t)$ represent the interpolated indicator function for the k^{th} block. For each t falling between two response times, $S^k(t)$ is given by (let the subscripts 0 and 1 in $S^k(t)$ represent states V_A and V_B respectively) :

$$S_0^k(t) = \left(1 - r_j^k\right)\left(\frac{t-(t_j^k-\tau_0^k)}{t_{j+1}^k-t_j^k}\right) + \left(1 - r_{j+1}^k\right)\left(\frac{(t_{j+1}^k-\tau_0^k)-t}{t_{j+1}^k-t_j^k}\right) \quad (1)$$

$$\text{for} \quad (t_j^k - \tau_0^k) \le t \le (t_{j+1}^k - \tau_0^k)$$

if the initial response is of type V_A. The interpolated indicator function for initial response of type V_B is similar.

No extrapolation is performed for times *t* greater than any of the response times. Instead, times exceeding the last response are treated as missing data for that block. To keep track of which times points are valid after τ_0^k, we compute a second indicator variable that marks valid time points:

$$J_0^k(t) = \begin{cases} 1 & 0 < t < \max(t_i^k - \tau_0) \\ 0 & \text{otherwise} \end{cases}$$

The survival probability estimate (P_S) for state *i* is computed as:

$$\hat{P}_S\left(s(\tau+t) = i \mid s(\tau) = i\right) = \frac{\sum_{k=1}^{\#blocks} S_i^k(t)}{\sum_{k=1}^{\#blocks} J_i^k(t)} \quad (3)$$

Response probabilities (P_R) are similar, but simpler to estimate. The recorded response sequences for each block were interpolated and averaged incorporating only valid times as above.

Estimating Cumulative Transition Probabilities

We also measured the cumulative transition probability from each state, conditioned on initial states with times τ_0^k and τ_1^k (where subscripts 0 and 1 represent states V_A and V_B). Note that the time spent in a state is the same as the time observed before a transition. To construct an indicator function for a transition from a state, V_A or V_B, we classify the responses subsequent to first response. in each block as preceding or following the first observed state transition. Beginning with the initial state, subsequent responses are classified as 0 until the first state-change response is observed. The remaining responses in the block are classified as 1. The classification produces an indicator variable $Z^k(t)$. Let Z_0^k denote the transition sequence described above for state V_A. Linear interpolation is used to construct indicator functions for cumulative probability until transition. For the cumulative transition probability of state V_A, the indicator function $T_0^k(t)$ is given by:

$$T_0^k(t) = Z_0^k\left(\frac{t-(t_j^k-\tau_0^k)}{t_{j+1}^k-t_j^k}\right) + Z_0^k\left(\frac{(t_{j+1}^k-\tau_0^k)-t}{t_{j+1}^k-t_j^k}\right) \quad (4)$$

$$\text{for} \quad (t_j^k - \tau_0^k) \le t \le (t_{j+1}^k - \tau_0^k)$$

We computed indicator functions for valid times as before, and generated estimates of cumulative transition probabilities (P_T) by averaging indicator functions across blocks:

$$\hat{P}_T\left(N_i(t) > 0 \mid s(\tau) = i\right) = \frac{\sum_{k=1}^{\#blocks} T_i^k(t)}{\sum_{k=1}^{\#blocks} J_i^k(t)} \quad (5a)$$

where $N_i(t)$ is the number of transitions from state *i* upto time t after the start time. Note that there is an event equivalence between this event and the time D_i, continuously spent in a state. Thus, another way of expressing Eqn. 5a is:

$$\hat{P}_T\left(D_i < t \mid s(\tau) = i\right) = \frac{\sum_{k=1}^{\#blocks} T_i^k(t)}{\sum_{k=1}^{\#blocks} J_i^k(t)} \quad (5b)$$

In Eqn. 5b, we are expressing the idea that P_T computes the cumulative probability that the duration of a percept is less than *t*. Phase duration means and variances were computed directly from numerical estimates of the moments of the cumulative transition probability curves.

Errors in estimates

Errors in probability function estimates were assessed by resampling data across blocks using standard bootstrap procedures (Efron and Tibshirani, 1993). In particular, blocks were re-sampled with replacement and 50 bootstrap estimates of the survival, response and cumulative transition probability functions were generated using the .632+ procedure (Efron and Tibshirani, 1997). Standard errors of the probability estimates were computed via standard deviations of the set of bootstrap probability function estimates.

We tested several assumptions of our estimation method by assessing whether modified estimates would fall outside of the 95% bootstrap confidence intervals. In particular, we found that changing the interpolation method to cubic interpolation had no significant effect on estimates, nor did assigning transition times to the midpoint between response times. We also assessed whether selecting only those blocks where τ_0^k and τ_1^k occurred in the first 1-3 seconds would change the estimates, but also found no significant effect.

Note that the estimates are noisier for the elapsed times approaching 10 seconds after the first recorded percept time; this is because there were fewer blocks in which the appropriate initial state occurred at the end of the block.

General Framework for Analyzing Multi-stability data

Markov Renewal Processes (MRPs) (Ross, 1992) are general, well-characterized stochastic processes that model transitions among discrete states. In particular, MRPs make transitions between discrete states via a Markov chain and sample the amount of time spent in each state from a set of state-contingent temporal distribution functions.

MRPs provide greater clarity on the analysis of multistability behavior by expressing relationships between different measures and address challenging issues in data interpretation such as the temporal dynamics and state stationarity.

Some of the key properties useful for quantifying bistable data include: temporal contingencies can be measured by conditioning on past events (we focus on first order, or just the previous event); non-stationarity can be diagnosed; MRPs allow for arbitrary temporal distribution functions for perceptual durations; many of the key predictions and relationships involve moments rather than full distribution functions (so that distributional assumptions can be avoided). Finally, in previous work (Schrater and Sundareswara, 2006) we showed how MRPs can be generated by the kind of theoretical model described in section "Theoretical framework for interpretation" below. We describe the properties of MRPs more fully in the supplementary section, "Markov Renewal Processes for Bistability."

Results

Average Effects of Context

The average effects of context on Necker cube perception are reported in Figure 6, Figure 7 and Figure 8. We analyze response data compiled across 20 participants (one participant was excluded from the analysis as a clear outlier). If context partially disambiguates perception of the Necker cube, we would expect background contexts that suggest view from above to increase V_A percept durations, but may also decrease V_B percept durations. We would also expect to see an initial bias toward V_A first in the first few responses.

Figure 6 shows response probability (P_R) as a function of time for view from below responses V_B (i.e. the percentage of time participants will respond V_B at the sound of the beep). The base-rate/no-context data (blue curve), shows that participants are biased towards V_A, percepts (viewpoint from above) quite strongly for initial responses made early in a response block and asymptote at approximately 40% after 5-6 sec. The fact that the asymptotic value is not 50% shows that percepts were biased toward viewpoint from above on average without context. The red curve (V_A context) shows increased V_A response to V_A context. The green curve shows increased V_B responses to and V_B context. However, while V_B context strongly biased initial responses toward view from below, V_A context stimuli do not significantly change initial bias over baseline. Overall, V_B context produces a larger change in response behavior than V_A context. V_A context may produce a smaller shift because of the bias toward V_A responses in the no-context data.

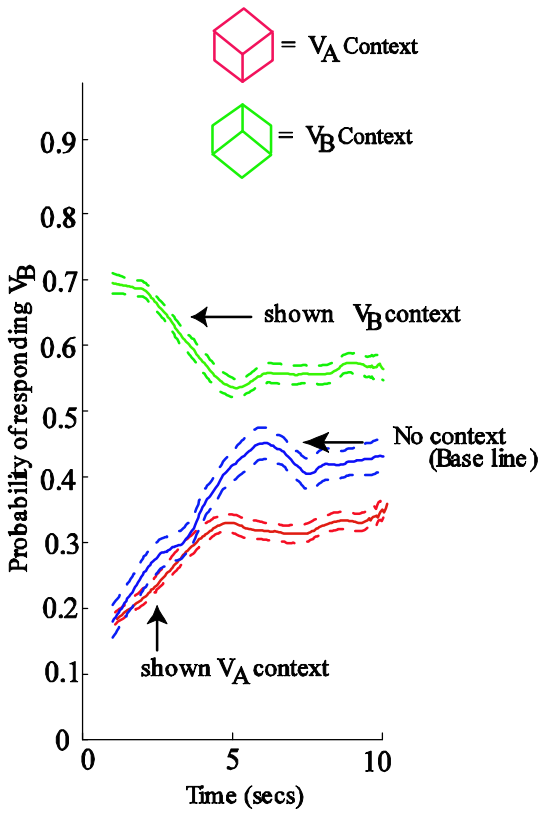
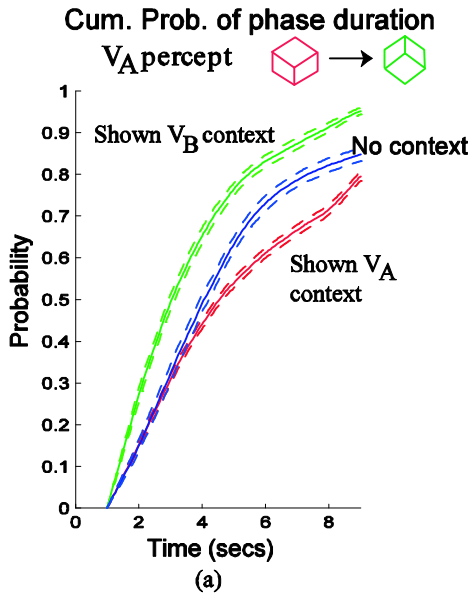


Figure 6: 'P_R' measure: Response of V_B percept as a function of time. The blue curve represents baseline data (no context), the red curve data is P_R when the context shown was V_A, and green represents P_R when the context was V_B context. The figure shows large initial response biases both with and without context. View from above context can be described as producing the same initial bias as baseline, but preserves more of the initial bias to asymptote around 0.3. V_B context produces a strong bias towards V_B responses that decreases to asymptote around 0.6 after 5 sec Dashed lines represented +/-1 bootstrap s.e. of the estimate.

While response probabilities showed overall effects of context, they do not capture state-contingent effects – for example, it does not answer whether context has differential effects on transition probabilities of similar and dissimilar percepts. We use two related measures to capture state-contingent effects, cumulative transition probability (P_T) and survival probability curves (P_S), where both are conditioned on the initial perceptual state. Based on response probability results which show increased responses to percepts that match the context, we would expect context to increase time until transition and increase survival probabilities for context-consistent perceptual states.



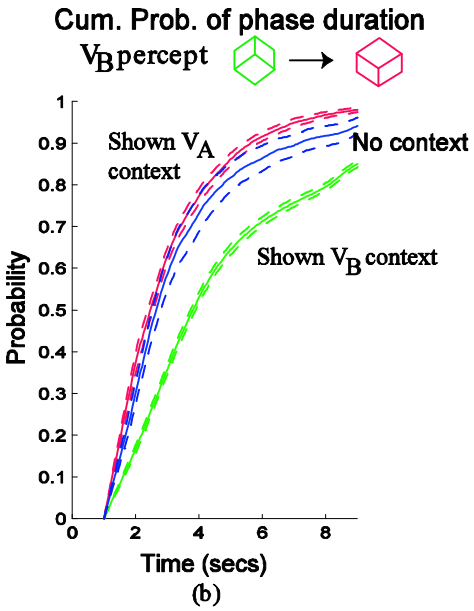
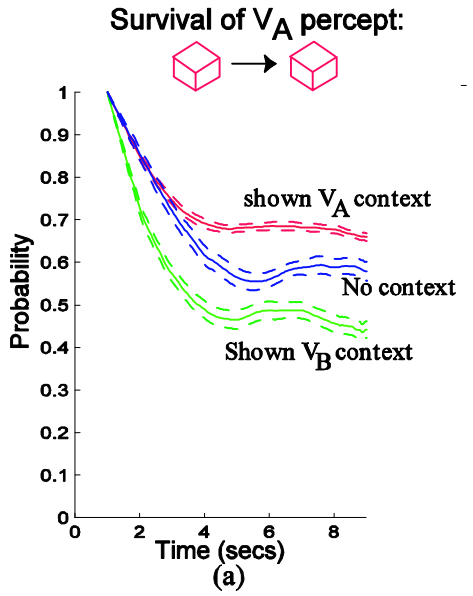


Figure 7: Average effect of context on cumulative transition probability conditioned on the initial perceptual state. The x-axis indicates time during a block and the y-axis is probability of having switched within the time indicated on the axis. Shows P_T curves given the initial percept is: (a) view from above, V_A and (b) view from below, V_B Context has both positive (positive) and negative (suppressive) effects on cumulative transition probability (red=V_A and green=V_B, dashed lines represent +/- 1 bootstrap standard error) compared to no context (blue curves). Durations (time before transition) spent in the V_A initial state (figure a) are prolonged by V_A context and shortened by V_B context. However, while initial V_B durations (figure b) are prolonged by consistent context, there is little suppression of V_B percept from view-from-above context when compared to the no-context (base-rate) condition. Note that context effects on distribution functions involve simultaneous shape and scale changes, corresponding to correlated changes in the phase duration means and variances.



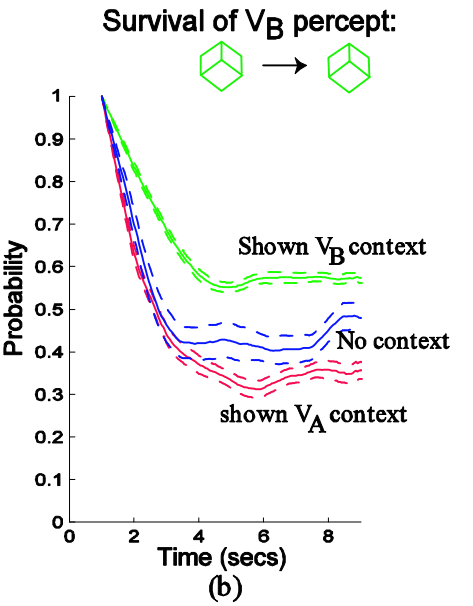


Figure 8: Effects of context on survival probabilities (P_S). **(a)** survival of V_A initial percept shown on the y-axis against time during a block shown on the x-axis. **(b)** survival of V_B initial percepts shown on the y-axis against time during a block shown on the x-axis. Context has both positive (prolonging) and negative (curtailing) effects on survival probability functions (red=V_A and green=V_B) compared to no context (blue curves). Dashed lines represent +/- 1 bootstrap standard error. Note that all survival curves have the same asymptote as the response probability functions (which they should if MRPs provide a sufficient data model) shown in Figure 6.

Cumulative transition probability (cumulative probability that a switch has occurred by time 't') (P_T) and survival probability (P_S) data are shown in Figure 7 and Figure 8. In both figures (a) shows results for initial V_A percepts (b) shows results for initial V_B percepts. Clear effects of context for both types of percept are observed.

When context is consistent with the initial percept, this initial percept takes longer to transition (i.e. consistent percepts are *less likely* to transition with consistent context.) and survival probabilities increase. In addition, inconsistent context has the effect of suppressing the unsupported percept, increasing the likelihood of transition. However, suppressive effects are only large for view-from-below initial perceptual states V_A. This might be related to biases toward V_A perceptual states found in the baseline conditions. For example, because V_A percepts are more frequent in baseline viewing (in terms of durations, and initial frequencies) there may be a ceiling effect for suppression. The relationship between the two measures in shown in supplementary section, "Markov Renewal Process for Bistability."

Next, we examined how the mean phase duration of both percepts was affected by context. Literature in binocular rivalry (Blake and Logothetis, 2002; Sobel and Blake, 2002) suggests that in the presence of a supporting context, durations of the matched percept are prolonged, but durations of the incongruent percept are not affected (Levelt's second proposition), suggesting that perhaps there are different neural events associated with increasing and decreasing perceptual durations.

However, in Figure 9 we show that not only were supported percept times lengthened but inconsistent percept times were shortened as well – although the amount of shortening in duration is much smaller. Thus we see no need to invoke different neural processes to explain our lengthened and shortened times in our bistability data.

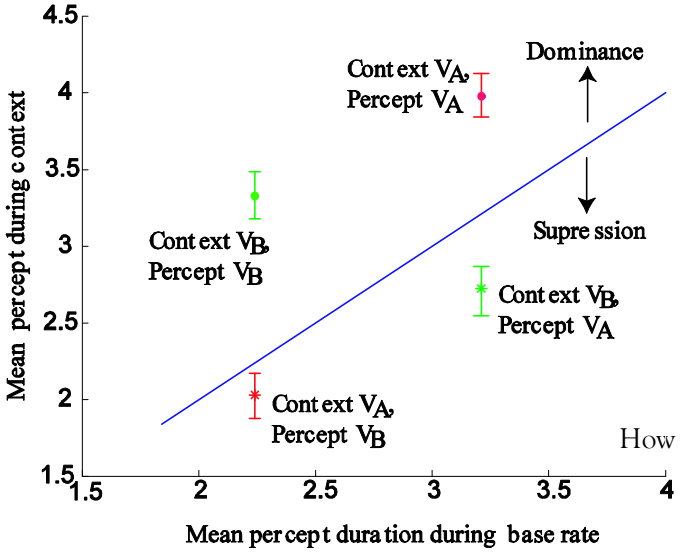


Figure 9: Effect of consistent and inconsistent context on mean percept durations. Mean percept durations with context are plotted against mean durations in the baseline condition, for each combination of percept and context type. Error bars represent 99% bootstrap confidence intervals. The blue line represents the values predicted if context had no effect. Points above the line correspond to percept durations increased by context (prolonging), while points below the line indicate context-induced shortening.

Theoretical Framework for interpretation

Probabilistic approach to Bayesian decision–making via sampling

The function of the brain's visual system is to make informed guesses or *inferences* about the world given varying levels of uncertain retinal input. Inferences resulting from Bayesian statistics have been widely employed to describe visual processing from the level of neuronal behavior (Pouget, Dayan and Zemel, 2003) to high-level visual processing (Yuille and Kersten, 2006). The purpose of this section is to show how bistable switching behavior might also arise from probabilistic models of perceptual inference and decision-making. By providing a simple method for generating spontaneous switching in a Bayesian framework we offer an alternative to the idea that explanations of perceptual bistability require specialized neural mechanisms whose job is to implement switching. At minimum, theories of bistability, should provide explanations of: 1) the existence of multiple interpretations; 2) the awareness of only one interpretation at a time; and 3) spontaneous switching between percepts.

The first two criteria are natural byproducts of using Bayesian decision theory to construct computational models of perception (Kersten and Schrater, 2002). Models of perception formulated using Bayesian decision theory treat percepts as decisions about scene (world) properties based on a combination of prior visual information (priors) and knowledge about image formulation and **scene**/image regularities (likelihood). Decisions have two parts—an inferential process that computes probabilities across interpretations and a selection process that chooses high-probability interpretations (and may incorporate costs, Kersten and Schrater, 2002).

In Bayesian decision theory, the presence of a multimodal posterior distribution function signals ambiguity in perception.. For example, part of perceptual inference for the Necker cube figure is to decide the viewpoint direction θ from which the image I of the cube was taken (we do not address the other part, which involves an inference about the shape of the wireframe object). Given that the object's shape is inferred as cuboidal, the posterior probability $P(\theta|I)$ will be bimodal (see Figure 10), with equal peaks given a uniform prior on viewpoint. Bayesian inference can also incorporate viewpoint information (gathered from experience or background context) to bias perception toward more probable views of the cube. Previous work

by Mamassian and Landy (1998) show that human observers are biased towards a viewpoint from above (our "no context" condition data suggests this as well). In Bayesian models, percepts result from selecting a good interpretation, typically by choosing the viewpoint with maximal posterior probability. When the maximum is not unique, Bayesian decision theory provides no mechanism for choosing between equally good interpretations. Thus, while criteria 1) and 2) are natural byproducts of using Bayesian decision theory, criterion 3) (the spontaneous switching of percepts) is not.

However, we found that spontaneous switching can arise when Bayesian decisions are implemented by a process that generates a set of samples from the posterior distribution across time, and selects a percept from the largest probability sample in the set. We describe this process in detail below. The key idea of the approach is to implement decision-making by a sampling procedure, in accord with some recent theories of how Bayesian decision theory may be *neurally implemented*. In particular, Lee and Mumford (Lee and Mumford, 2003), and Pouget (Pouget, Dayan and Zemel, 2003), point out that probabilistic population coding models of cortex can be thought of as representing probability distributions in terms of samples from the distribution. In this approach, receptive fields of neurons in the population encode regions of the interpretation parameter space, and firing rates encode the probability of a sample. Whether or not this interpretation of neural coding is an accurate reflection of biology, sampling constitutes a fundamental method for implementing Bayesian decisions, and the brain's computations may perform something equivalent. In the next section we explain how switching can spontaneously arise in Bayesian decisions based on sampling.

Bistability from a sampling model for perceptual decision-making

We assume the brain performs Bayesian inference on the Necker cube figure and computes samples from the (bimodal) posterior distribution, representing the two dominant viewpoint interpretations. A decision process operates on this posterior, and brings the best interpretation into awareness and memory. More precisely we make the following assumptions:

1) **Posterior probabilities** are represented by a set of samples, updated across time. In particular, at a sequence of times t_i (on a neural scale), the brain explores the posterior distribution by collecting a new set of samples and discarding the oldest samples. Samples, consist of both the parameter values $\theta(t_i)$ and the associated posterior probabilities, $w(t_i)$, i.e. $\{(\theta(t_1),w(t_1)),\ldots,(\theta(t_N),w(t_N))\}$, where $w(t_1)=P(\theta(t_1)\,|\,I)$ are *weights* that represent the posterior probabilities when first sampled and t_i are consecutive time points. Weights represent the quality of the interpretation associated with a sample and they are discounted by a memory decay process.

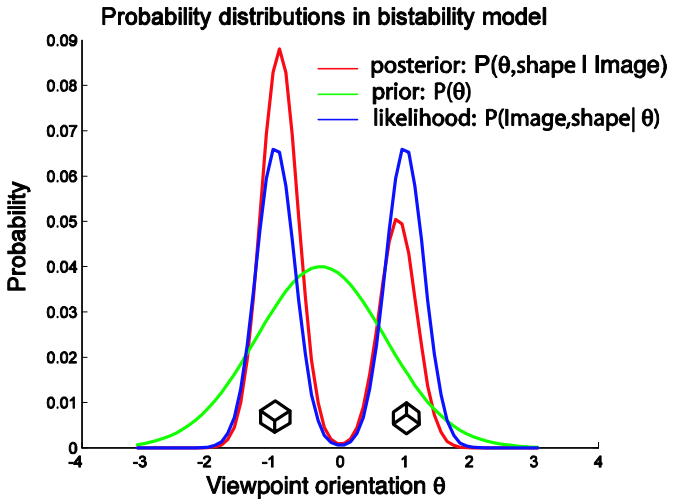


Figure 10: An illustration of the probability densities used in the model. The interpretation ambiguity is represented by a bi-modal density function on cube orientation θ (blue curve). The prior $P(\theta)$ encodes biases toward one of the interpretations, including the effects of context (green curve). The product of the likelihood and the prior produces the (normalized) posterior probability (red curve), from which we sample in our simulations.

2) **Memory Decay** expresses the idea that the quality of an old interpretation decreases with time. We assume simple exponential discounting of a memory sample's weight by the age of the sample.
3) **Perceptual Decisions** result from choosing the sample with the highest discounted weight. This sample's parameter value is chosen as the current interpretation and brought into memory. The memory sample's discounted posterior probability $w(t_i)$ represents the quality of the interpretation; this is also stored in memory. The interpretation process is illustrated in
4) Figure 11. These basic assumptions produce spontaneous perceptual switching similar to human behavior.

In (Schrater and Sundareswara, 2006), we analyze the probabilistic properties of this kind of sampling model in a more general setting and show that they generate Markov Renewal processes. Note that memory update events (where samples in memory are exchanged) can be thought of as perceptual state transitions. These transitions can either be same-state or between state transitions, but we assume that same-state transitions are not perceptually salient. Additionally, we showed that MRP models can be good descriptions of our human bistability data. Here we illustrate the explanatory power

of the theory described above by accounting for the main elements of our data with a simple implementation of a sampling-based decision model.

Simulation Methods:

Posterior: A 2D image I of a Necker cube is theoretically is compatible with the projection of an infinite number of polyhedra, depending on the depth of its 8 vertices. However, human perception is dominated by two interpretations, both cuboidal in shape, that involve two different poses or viewpoints on the object. Therefore we simplified the 8-dimensional Necker Cube vertex-depth space into a 1D viewpoint (elevation) parameter space θ of the observer (we will henceforth refer to the viewpoint elevation parameter as just viewpoint parameter for brevity.) The viewpoint likelihood defined by $P(I|\theta)$ is modeled as the sum of two Gaussians centered on each of the two Necker cube interpretation modes, θ_{VA} and θ_{VB}. Assuming a Gaussian prior on viewpoint centered on θ_M, the posterior, $P(\theta|I)$ is proportional to $P(I|\theta)P(\theta)$.

Memory Decay: Memory decay is implemented by exponential function $e^{-t/\tau}$ that discounts the weights attached to samples as a function of their age t. The rate of decay is determined by the parameter τ. The sampling rate was arbitrarily set to 2 samples/sec, however, this choice did not matter because changes in sampling rate can be compensated by changes in τ.

Sampling and Perceptual Decisions: The sampling is carried out as follows. At the start of the simulations, 'n' independent samples are drawn i.i.d. from the posterior distribution. The number 'n' is set to 2τ. The order in which we process the samples is indicative of the temporal nature of the sampling. The memory decay function ($e^{-t/\tau}$) is multiplied by the posterior probability of the samples, where 't' indexes the samples. The sample with the largest weight is selected as the current percept–the orientation of this sample represents the perceived orientation of the Necker cube. Subsequently, new samples are drawn in the same way and added to the current set of samples, and samples older than 2τ seconds are discarded.

Simulations: To illustrate model properties we varied two parameters in the simulations, the ratio of posterior heights at the peaks of the two likelihoods and the memory decay rate. These two parameters were varied because they corresponded to hypothesized differences between participants (this is confirmed by Figure 13).

We varied the ratio of posterior heights to account for potential differences in response probabilities and cumulative transition probabilities between subjects. There are many ways to vary the posterior heights, but we were particularly interested in assessing the effect of varying the width of the prior distribution because the width of the prior distribution encodes how strongly viewpoint is biased by the context manipulation. Leaving the prior mean θ_M fixed to reflect either context V_A or V_B, the relative heights of the two posterior peaks can be controlled by varying the prior variance σ^2, which adjusts the *ratio* of the prior probability values at the two peaks of the likelihood:

$$\omega = \frac{P(\theta_{VA})}{P(\theta_{VB})} = \exp\left(\frac{(\theta_{VB}-\theta_M)^2-(\theta_{VA}-\theta_M)^2}{2\sigma^2}\right) \quad (6)$$

where θ_{VA} is the mean parameter value that represents the "view from above" percept and θ_{VB} is the mean that represents the "view from below" percept.

We varied ω between 0.1 and 10 for each context condition. For simplicity, we rescaled the viewpoint axis so that the two Necker cube percepts corresponded to -1 (V_A percept at mode θ_{VA}) and +1 (V_B percept at mode θ_{VB}) and the likelihood standard deviations were set to 0.25. The mean of the prior, θ_M, was fixed at 0.1 or at -0.1, depending on which context condition we were simulating.

To model potential differences in memory discounting between observers, we varied a second simulation parameter, the exponential memory decay constant τ which determines how quickly a sample's posterior probability decays. Assuming that 2 samples are generated every second, we varied the memory decay constant (τ) from 20 seconds to .0198.

Properties of the Model

The sampling model generates a number of interesting properties that can be described by Markov Renewal Processes (MRP). MRPs are generated by distributions $Q_{ij}(t)=\pi_{ij}F_{ij}(t)$ which represent the probability that the process makes a transition to state j in time less than or equal to t, given the process just entered state i. MRPs transition between states via a Markov chain generated by the transition matrix $\pi_{ij} = P(s_N = j | s_{N-1} = i)$. The times between state transitions are described by temporal distribution functions $F_{ij}(t)$ that depend on the current state and the next state entered.

Sampling from this generating process involves picking an initial state, using the transition matrix to draw subsequent states, then sampling a duration from the appropriate temporal distribution. In (Schrater and Sundareswara, 2006), we derive the correspondence between the sampling model and MRPs. There we show the transition matrix parameters in π_{ij} are a function of the relative probability mass under the posterior peaks. Because the sampling we used was independent (i.e. the transition to a new state is not dependent on the previous state), the rows in the transition matrix of the corresponding MRP model are equal to each other. The MRP temporal distribution functions F_{ij} correspond to how long a maximally weighted sample in the Bayesian model remains the maximum before replacement.

In summary, in the Bayesian model, the ratio

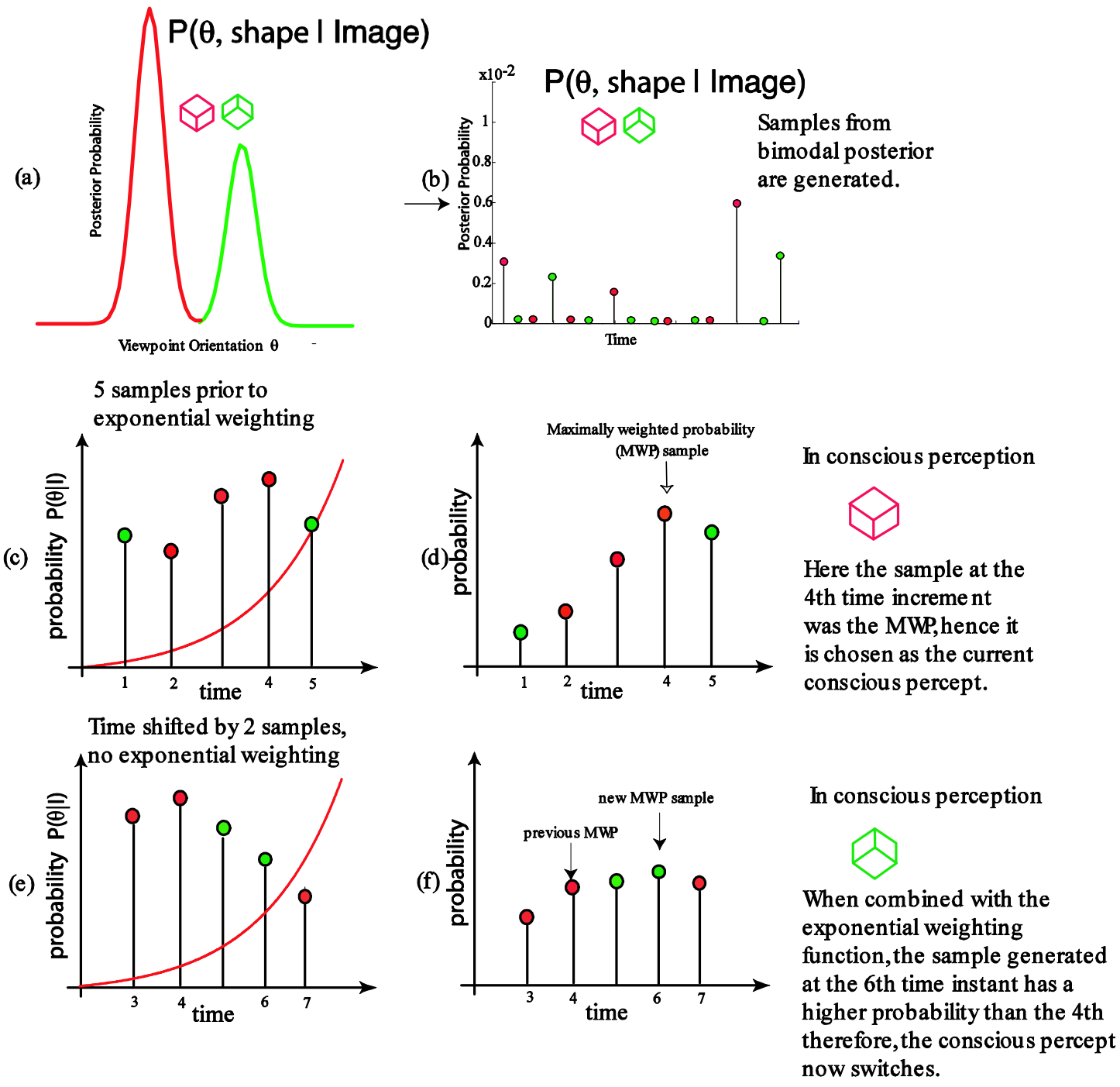


Figure 11: An illustration of the sampling procedure. (a) A bimodal posterior over viewpoint parameter (x-axis). The two modes correspond to the two interpretations: V_A percept and V_B percept (b) The probability values for samples chosen earlier in (a) shown: red = V_A percept, green = V_B percept (c) 5 samples in window are shown. Each sample is generated at the point in time as indicated by time (x-axis) axis. The red corresponds to the V_A percept and the green corresponds to the V_B percept. The exponential decay factor is applied to the samples (i.e. each sample's probability is weighted by the exponential decay factor's value corresponding to that point in time). (d) The sample with the maximum value after the weighting is chosen as the current percept. (e) A new sample at the 6^{th} time instant is now in the window – the exponential decay factor is applied. (f) The new maximum probability sample is now the sample that arrived at the 6^{th} second; this causes a switch in percepts. The last percept in memory was V_A, now it has been changed to V_B.

of the mass under the posterior peaks and the memory decay determine the amount of time spent in each percept; in the MRP model, this translates to the entries of the transition matrix (π_{ij}) and the probability of duration in each state given by entries of the temporal distribution matrix F_{ij}.

In the next section we show that the model-generated MRP has temporal distributions F_{ij} that are approximately exponential and the same for all four state transitions. In addition, because we used independent sampling, the transition probability are independent of the previous state: i.e. $\pi_{ij} = \pi_{jj,}$. Thus, the MRP process for bistability reduces to just 2 parameters (the probability of transition to V_A and mean of the exponential temporal function).

We can use this correspondence to elucidate key properties of the sampling-based decision model. First, a percept's duration is the sum of 'k' sample times from the exponential distribution, where 'k' is the number of self-state transitions that occur before a state-change transition. Because means and variances of i.i.d. random variables sum, we predict correlation between means and variances of percept durations. Moreover, the probability of self-state transitions plays the critical role in the determining perceptual duration distributions. Each self-state transition contributes an exponentially-distributed random time—if the number of self-state transitions were fixed rather than stochastic, percept durations would be gamma-distributed (the sum of 'k' exponentials is gamma-distributed with parameter λ has mean and k/λ and variance k/λ^2). Since k is random, the duration distributions are mixtures of exponentials, and the probability of self-state transition determines this mixture. Because the relative number of self-state transitions is determined by the posterior probabilities, the effects of changing the posterior probabilities will cause specific, predictable changes in the shape of the percept duration distributions, the survival probability functions and the mean and variances of the percept durations. We test these predictions in the results section below.

Simulation Results

In this section, we compare the results of the simulations with our experimental data. To better compare human and model behavior, we simulated the experimental data collection procedure to collect model responses and analyzed the simulated response data exactly the same way as human data.

First we show that the simulated data exhibits switching distributions similar to humans. Figure 12 shows simulated cumulative transition and survival probability curves for the maximum-likelihood parameters that best matched human data. Mean and confidence intervals were obtained by obtaining maximum likelihood estimates for 50 bootstrapped samples of human data. The mean maximum likelihood parameters for context V_A are posterior peak ratio, ω= 1.1015 +/-.0165 (1 std) and τ= .0846 seconds = +/-.02 (1 std). The mean estimate obtained is the .632+ estimate (Efron, 1997). The mean maximum likelihood parameters for context V_B are posterior peak ω = 0.9645, +/- 0.0127 (1 std) and τ = .0652 +/- .01 (1 std). Since the memory decay parameter, τ is similar within 1 standard deviation between contexts, it is not a parameter that explains the difference between context behavior. Instead the single parameter ω (which determines the heights of the posterior peak ratio) explains the difference in cumulative transition and survival probability functions induced by context. We also fit the mean durations of both subject and the simulated data durations with gamma curves; all the fits returned could not be rejected.

Next we addressed whether variations in simulation parameter can capture much of the between-subjects variation. In Figure 13 below, we plot phase duration means against variances for each of the 20 participant's data in the context condition. Plotted on the same graph are simulated means and variances, where each point corresponds to a different pair of posterior ratio (at the two peaks of the likelihood) and memory decay values. Despite the simplicity of the theoretical assumptions and simulation choices, the model produced bistable behavior similar to human data and the variation in simulation parameters captures most of the spread of human data. Note that the simulated data follows the same "arc-like" trend of the human observers for the context consistent condition (the arc is a consequence of the data collection method—since we only collected responses for up to 10 seconds after the start of the block, this caused variability to decrease for mean durations greater than half the data collection period). Additionally, variation in simulation parameters generates correlations between means and variances for all four conditions that are comparable with human data.

To better characterize the statistical properties of the simulation model, we computed the MRP generators that corresponded to the best-match simulation parameters for context V_A. From the best-matching simulations, we computed π from the percentage of memory update events that were of each transition type. The resulting transition for V_A context was:

$$P_{ij} = \begin{pmatrix} .6415 & .3585 \\ .6415 & .3585 \end{pmatrix}$$

The first row of the matrix shows that $0 \rightarrow 0$ transitions (π_{00}) occurred at almost twice the rate of $0 \rightarrow 1$ (π_{01}), and similarly the second row shows $1 \rightarrow 0$ (π_{10}) occurred at almost twice the rate of $1 \rightarrow 1$ transitions (π_{11}).

The resulting transition matrix for V_B context was :

$$P_{ij} = \begin{pmatrix} 0.4448 & 0.5552 \\ 0.4448 & 0.5552 \end{pmatrix}$$

The difference in transition probability between contexts captures all the effects of context on switching behavior in the model.

As previously mentioned, sampling from this model, generates temporal distribution functions that approximately exponential and the same for all transitions.

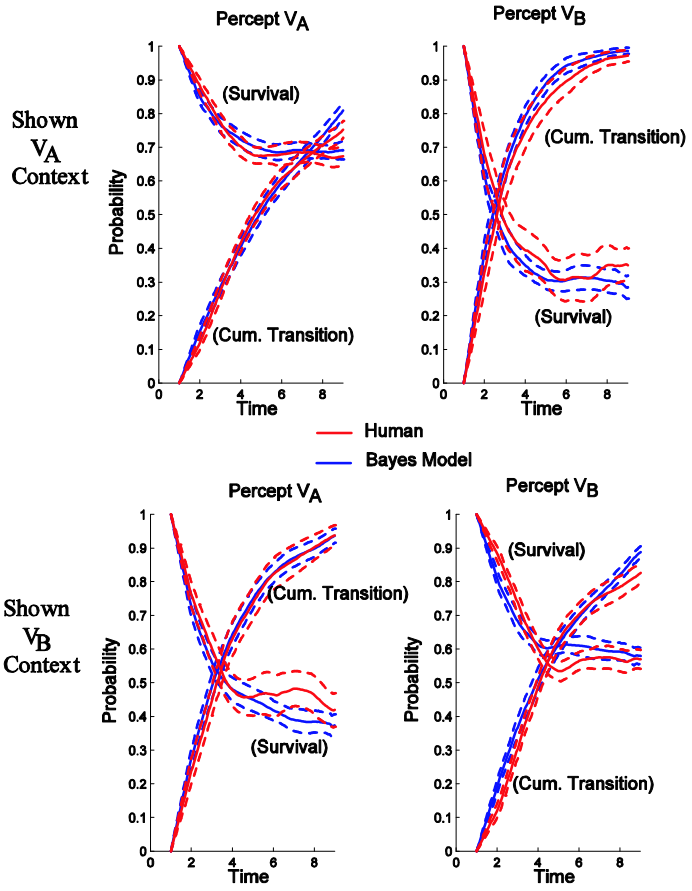


Figure 12: Comparison of pooled human (red) and model (blue) survival probability curves, for the best-matched simulation parameters. Survival probability curves for: (a) Percept V_A in context V_A. (b) V_B in context V_A. Simulations for context V_A run with best-matched parameters: $\omega = 1.1015 +/- .0165$ (c) V_A in context V_B. (d) V_B in context V_B. Best matched simulation parameters $1/\omega = 1.037 +/- .0127$. Dashed lines represent +/- 1 bootstrap SE.

The MRP generator temporal distribution functions $F_{ij}(t)$ for context V_B are shown in Figure 14 These curves were obtained by computing histograms of the amount of time between memory updates. Notice that they are similar to exponential functions and to each other, implying that they share the same temporal distribution mean. The temporal distribution functions for V_A context show similar behavior.

In summary, a prior bias, which determines the posterior heights in the model, affects the probability of transitioning into V_A and V_B percepts, effectively by changing the probability of self-state transitions. The simplicity of this model is surprising given the alternatives. For example, adaptation or fatigue models of bistability make no clear predictions about the effect of context—the distribution of percept durations could be almost anything, and the distributions could change in many ways.

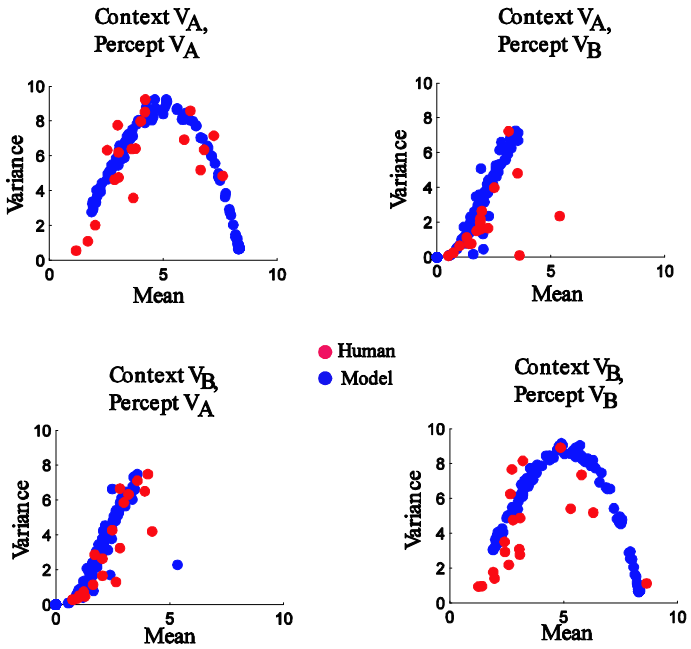


Figure 13: Comparisons of phase duration means vs variance in context-consistent and context-inconsistent phases. In all figures a-d, red points indicate subject data and blue points indicate simulated data for one pair of (ω,τ) parameter values. (a) Context-consistent view-from-above data (V_A context/V_A percept). There is an arc-like pattern (which is a result of the data collection) in the human data that is mirrored by the simulated data. (b) Context-inconsistent V_B percept data (V_B percept durations in V_A context). (c) Context-inconsistent V_A percept data. (d) Context-consistent V_B percept data (V_B percept survival in V_B context).

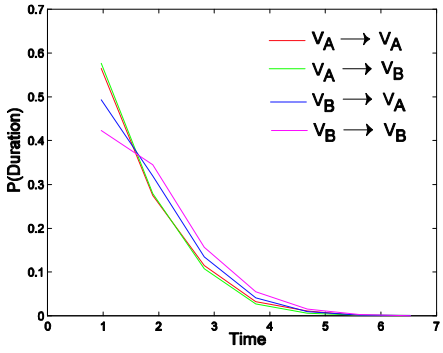


Figure 14: Temporal distribution functions for all transitions recorded while simulating bistable data (Best matched human parameters for V_B context were used for the simulation). The simulation gives rise to temporal distribution functions that are approximately exponential and are similar for all transitions.

Discussion

We presented a sampling-based decision-making model that modeled the effect of context on bistability. Using exponentially weighted samples from the posterior distribution, we were able to produce switching behavior similar to human observers. From the standpoint of MRP generators, the simulations generated a very simple model, and that model captures much of human switching dynamics. However, the sampling idea presented above can generate much more complicated behavior. In particular, in Schrater and Sundareswara (2006) we show that the independent sampling assumption can be relaxed, and that relaxing that assumption generates a richer class of MRP models. The simulations also rely on a fixed posterior distribution, which is unreasonable. A better model of the perceptual inference should incorporate the effects of time, eye movements and memory that cause the posterior distribution to vary with time.

As an example of the simplifications in the probability model, we used a likelihood model on viewpoint (pose) $P(I|\theta)$. A more realistic representation would incorporate likelihood functions that implement perspective projection constraints in the inference of 3D objects from 2D, the effects of eye movements on the image data, and priors embodying the typical biases found in interpreting line drawings, like preference for symmetry between 3D edges, or compactness. Additionally, the current system could be embedded in a hierarchical recurrent inference system similar to the one proposed by Lee and Mumford (Lee and Mumford, 2003). However, the simple simulations were sufficient to illustrate the idea that implementing Bayesian decision-making via sampling can create human-like switching behavior.

In the basic implementation presented here, we assume that switching behavior is generated by a stationary process. However, a non-stationary switching process is easy to simulate, although harder to characterize mathematically. By allowing posterior probabilities to vary with time (e.g. due to eye fixations or previous stimuli), we believe many reported bistability effects could be accounted for by models that compute Bayesian decisions via sampling (see overall Discussion below).

Our approach is related to a model proposed by Gepshtein and Kubovy (Gepshtein and Kubovy, 2005) to account for data from experiments with bistable dot-lattices. The authors posit that bistability in successive presentation of dot lattices results from an intrinsic bias that stochastically drifts in orientation space, causing a change in perception. Their "intrinsic bias" and "stimulus support" are similar to prior and likelihood respectively. Percepts are probabilistically generated at transition times, where the probability of a percept depends on the ratio of its intrinsic bias and stimulus support. Their model can be viewed as an instance of a sampling model, but with no memory and a non-stationary prior distribution (the drifting intrinsic bias). In fact the key to generating switching behavior in their model is the drift in the "intrinsic bias". In contrast, the most important aspects of the model used in our simulation are self-state transitions. Both models show the potential explanatory power of sampling within an inference framework to account for bistability data.

Discussion

When a bistable image was embedded in a context that favored one of the interpretations, perceptual ambiguity was reduced; the interpretation supported by the context persists longer and occurs more frequently. Additionally, the duration of the percept that is not supported by the context is shortened relative to baseline (without any context). These empirical results suggest context is modulating the perceptual ambiguity of the Necker cube in a manner consonant with Bayesian models of perceptual inference. Our model's successful prediction of both the time course of perceptual switching and the impact of context suggest that perceptual bistability may occur as a by-product of normal perceptual decision-making: postulating special mechanisms for switching may be unnecessary. In particular, the same processes may mediate lengthening and shortening of duration of bistable percepts (a view also suggested by Gepshtein and Kubovy, 2005). This view is not supported by binocular rivalry results, where supporting context might prolong the consistent percept, but does not affect the inconsistent percept (Sobel and Blake, 2002).

In this paper we introduced a perceptual decision-making model based on Bayesian inference as an approach to modeling bistability. The model is based on searching for the high probability interpretation of the Necker cube by sampling a posterior probability distribution on interpretation parameters over time, together with exponential discounting of past samples. We showed that this model is capable of producing bistable-like behavior when the posterior distribution is bimodal. Much of the explanatory power of this model stems from its reliance on Bayesian models of perceptual inference. Bayesian models use likelihood terms to encode knowledge about image rendering by expressing the scene parameters consistent with the image. For the Necker cube, a likelihood would express that 3D vertex depths must be perspectively consistent with the image lines and vertices. When combined with a prior that promotes 3D interpretations with equal angles between edges, the posterior distribution is bimodal on pose or viewpoint. Within this framework, the effects of context can be simply described as influencing prior probabilities of scene interpretations.

Can other effects of perceptual multi-stability be explained within our theoretical framework? We believe the theory provides a natural interface to interpret both high-level and low-level effects on bistability, because any event that has an impact on the relative probability of

perceptual interpretations can influence decision-making in the model. In particular, manipulations of variables that produce changes in a Bayesian interpretation of the ambiguous image should impact perceptual bistability. For example, variables that influence perceptual stability like eye position (Einhauser, Martin and Konig, 2004; Ellis and Stark, 1978; Kawabata, 1978) and attention (van Ee, Noest, Brascamp and van den Berg, 2006; Hol, Koene and van Ee, 2003; van Ee, van Dam and Brower , 2005; Toppino, 2003; Meng and Tong, 2004) are known to affect perceptual decision-making as well. To explain these effects with our model we would need to show that these variables mediate their influence by changing the relative probability of perceptual interpretations.

The idea that perceptual multi-stability originates at a stage of visual processing involving perceptual decision-making may help reconcile perceptual multi- stability and binocular rivalry. We showed that context could be modeled as influencing a prior distribution on interpretations for bistability. The strong similarities between rivalry and bistability effects raises the possibility that rivalry may result from a similar kind of perceptual decision process as bistability, except that image interpretations choices in rivalry are coupled with choices about data sources.

Quantifying perceptual bistability

The measurement framework introduced in this paper can be used to address several complex issues that arise in the interpretation of bistability data, including stationarity, covariation between phase duration measures such as mean and variance, and relationships between state-transition models and percept duration distributions.

Intuitively, stationarity measures whether the probability distributions governing perceptual state transitions vary across time. A stochastic process with stationary dynamics will converge to a steady-state distribution after a burn-in period, and maintain that distribution for all subsequent time. Convergence occurs when the probability of observing a state is independent of previous states after sufficient elapsed time T. Moreover, the rate of convergence conveys important information about the strength of dependence between subsequent states. By definition, survival probability functions capture exactly the information needed to assess both stationarity and the rate of convergence. For percept V_B, stationarity implies there is an elapsed time T such that the probability of the present state is independent of the past:

$$P(V_B(t+T)|V_B(t))=P(V_B(t+T)|V_A(t)) \quad (7)$$

Over a finite data collection interval, a test of this condition will indicate whether the process is non-stationary (or failed to converge in the data collection period if the required T is longer). This condition is the same as requiring that survival probability asymptotes agree, which follows by expressing the right hand side of equation 7 in terms of a survival probability function. By laws of conditional probability,

$$P\left(V_B(T+\tau)\,|\,V_A(\tau)\right)=1-P\left(V_A(T+\tau)\,|\,V_A(\tau)\right)$$

Plugging in equation 7 yields a testable condition for stationarity: there must exist a time T such that survival probabilities agree according to the equation:

$$P\left(V_B(T+\tau)\,|\,V_B(\tau)\right)=1-P\left(V_A(T+\tau)\,|\,V_A(\tau)\right)$$

For times greater than T the survival probability functions should converge to the response probabilities, *if the process is stationary*. In contrast, there exist non-stationary processes (e.g. periodic processes) that can express long-range dependence between states and will not pass this test.

We present the survival probability curves from our baseline and context experiment re-plotted in Figure 15. In both figures, the dark curves shows $P(V_A(t+\tau)|\ V_A(\tau))$ and the light curves shows $1 - P(V_B(t+\tau)|\ V_B(\tau))$. Both context and baseline data appear stationary and converge within the data collection period, within measurement error. The rapid convergence observed in our Necker cube data contrasts with the results of Mamassian and Goutcher (2005) for binocular rivalry data. Using survival probability functions, they found rivalry states were strongly dependent on the initial state, and this dependence persisted for at least an order of magnitude longer than we found in our data.

The covariation between means and variances we observed across participants has been found in several other studies (Mamassian and Goutcher, 2005; Borsellino, 1972). We show that our theoretical model generates an MRP with simple properties and covariation that strongly resembles our participant's data. Thus, our approach provides a compelling explanation for one of the recurring patterns in bistability data.

Traditionally, bistablity data has been characterized by measuring phase duration distributions, frequently using gamma distributions. Markov Renewal Processes employing same-state transitions and exponentially distributed times between state transition events produce phase durations that can be well-approximated by Gamma distributions, even though not exactly Gamma. Rather than providing a better fitting family of distributions, the introduction of Markov Renewal Processes provides a mathematical framework for expressing properties and relationships between model properties and measures of the dynamics of bistable behavior of the visual system.

Conclusions

In this paper, we showed how perceptual multi-stability could be interpreted as arising from a general process of probabilistic decision-making with ambiguous images,

without requiring special mechanisms to implement switching via adaptation or hysteresis.

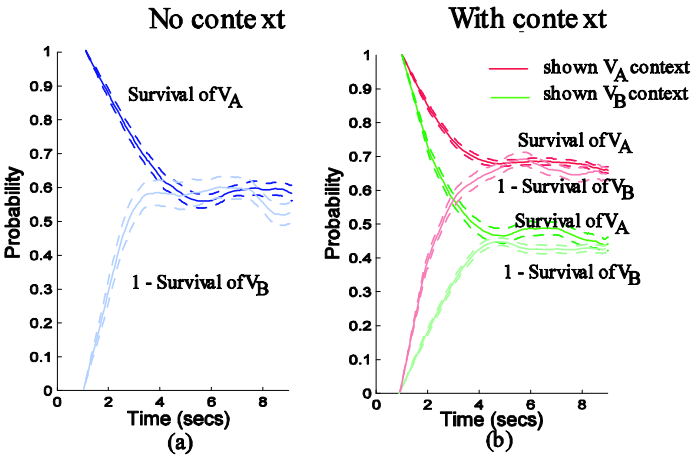


Figure 15: Assessing stationarity using survival probability curves. (a) No- context condition: Because the dark-blue (V_A survival) has the same asymptote as the light-blue curve (1- V_B survival), the survival of a percept at asymptote is independent of initial state—indicating a stationary process (b) Similar behavior is observed in the case of context. The two red survival curves correspond to V_A context and the two green curves to V_B context. 1.s.e. bars are shown.

In the model, a decision is made as to which percept is currently experienced by sampling from a probability distribution across possible interpretations, and percepts are selected as samples with the largest time-discounted probability. We show that the model can reproduce the detailed dynamics of bistability, including covariation between means and variances of percept durations and the inverse relationship between the durations of the two percepts. We find that introducing regularity in the orientations of background objects influences the dynamics of bistability in ways quantitatively fitted by the model. To characterize the temporal dynamics effectively, we used Markov Renewal Processes; this quantification both allows the process to be described mathematically with its measures, and allows us to introduce the same-state transitions that underlie the success of our modeling approach.

References

Alais, D., and Blake, R. (1999) Grouping visual features during binocular rivalry. *Vision Research.* 34 4341-4353.

Alais, D., O'Shea, R.P., Mesana-Alais, C. and Wilson, G. (2000) On binocular alternation. *Perception* 29, 1437-1445.

Attneave, F. (1971) Multistability in Perception. *Scientific. American*. 225, 63–71.

Barlow, HB (1990) A theory about the function role and synaptic mechanism of visual after-effects. In Blakemore C (editor) *Vision: cosing and efficiency*. Academic Press, New York, pp 363-375.

Bialek, W., DeWeese, M. (1995) Random Switching and Optimal Processing in the Perception of Ambiguous Signals. *Physics Review Letters.* 74(15) 3077-80.

Blake, R.R. (1989) A neural theory of binocular rivalry *Psychol. Rev*. 96, 145–167.

Blake, R. and Logothetis, N.(2002) Visual Competition. *Nature Reviews Neuroscience,* 3, 1-11.

Brascamp, J. W., van Ee, R., Pestman W. R., and Van den Berg, A.V. (2005). Distributions of alternation rates in various forms of bistable perception. *J. of Vision* 5(4), 287-298.

Brewster, D. (1847) On the conversion of relief by inverted vision. *Edinburgh Philosoph. Trans.*, 30, 432–437.

Borsellino, A., De Marco, A., Allazetta, A., Rinesi, S., & Bartolini, B. (1972). Reversal time distribution in the perception of visual ambiguous stimuli. *Kybernetik*, 10, 139-144

Brower, G.J. and van Ee, R.(2006) Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Vision Research*, 46, 3393-3402.

Davis, J., Schiffman, H.R., and Greist-Bousquet, S. (1990) Semantic context and figure-ground organization. *Psychological Research* 52, 306-309.

De Marco, A., Penengo, P., Trabucco, A., Borsellino, A., Carlini, F., Riani, M., et al. (1977). Stochastic models and fluctuation in reversal time of ambiguous figures. *Perception*, 6, 645-656.

Diaz-Caneja, E. (1928) Sur l'alternance binoculaire. *Ann. Ocul.(Paris).* October 721-731. (Translation provided in Alais, D., O'Shea, R.P., Mesana-Alais, C. and Wilson, G. (2000)).

Eckstein, M.P., Pham B.T., Shimozaki, S.S. (2004). The footprints of visual attention during search with 100% valid and 100% invalid cues, Vision *Research*, 40, 1193-207

Eckstein, M.P., Drescher B., Shimozaki, S.S. (2006) Attentional cues in real scenes, saccadic targeting and Bayesian priors, *Psychological Science*, 17, 973-80.

Efron, B. and Tibshirani, R.J. (1993), *An introduction to the bootstrap*, Chapman & Hall.

Efron, B. and Tibshirani, R.J. (1997) Improvements on Cross Validation: The .632+ Bootstrap Method. *Jounral of the American Statistical Association.* Vol 92, No. 438, p.548-560.

Einhauser, W., Martin, K. A., and Konig, P. (2004). Are switches in perception of the Necker cube related to eye position? *Eur J Neuroscience.* 20(10), 2811-2818.

Ellis, S.R. & Stark, L. (1978) Eye movements during the viewing of Necker cubes. *Perception*, 7, 575–581.

Freeman, W.T (1994) The generic viewpoint assumption in a framework for visual perception, *Nature*, vol. 368, April 1994.

Fox R, (1991) "Binocular rivalry", in *BinocularVision Ed.D M Regan, volume 9 of Vision and Visual Dysfunction* Ed. J R Cronly-Dillon (London: Macmillan), 93-110.

Gepshtein, S. and Kubovy, M. (2005) Stability and change in perception: spatial organization in temporal context. *Exp. Brain Res* 160: 487-495.

von Grunau, M. W., Wiggin, S. and Reed, M. (1984). The local character of perspective organization. *Perception and Psychophysics*. 35(4), 319-324.

Moran, J. and Desimone, R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science,* 229, 782-784.

Hol, K., Koene, A. and van Ee, R. (2003). Attention-biased multi-stable surface perception in three-dimensional structure-from-motion. Journal of Vision, 3, 486–498.

Kawabata, N., Yamagami, K. & Noaki, M. (1978) Visual fixation points and depth perception. *Vision Res.*, 18, 853–854.

Kersten, D., Mamassian, P. and Yuille, A. (2004) Object Perception as Bayesian Inference. *Annual Review of Psychology*. Vol. 55, 271-304.

Kersten, D. & Schrater, P. R., (2002). Pattern Inference Theory: A Probabilistic Approach to Vision. In R. Mausfeld, & D. Heyer (Ed.), Perception and the Physical World. Chichester: John Wiley & Sons, Ltd

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. Trends in Neuroscience. 27(12), 712 - 719.

Kovács, I., Papathomas, V., Yang, M. and Ákos Fehér (1996) When the brain changes its mind: Interocular grouping during binocular rivalry. *Proc. Natl. Acad. Sci.* 9, 15508-15511.

Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America*. Vol. 20, No. 7.

Lehky, S.R (1988) An astable multivibrator model of binocular rivalry. *Perception*. 17, 215-229.

Leopold, D. and Logothetis, N.(1999) Multistable phenomena: Changing views in Perception. *Trends in Cognitive Sciences*. Vol.3, No.7, 254-264.

Leopold, D., Wilke, M., Maier, A. and Logothetis, N. (2002) Stable perception of visually ambiguous patterns. *Nature Neuroscience*. Vol. 5, No. 6.

Levelt, W. (1965) *On Binocular rivalry.* Institute for Perception RVO-TNO. Soesterberg, The Netherlands.

Long, G., Toppino, T. and Mondin, G. (1992) Prime Time: Fatigue and set effects in the perception of reversible figures. *Perception and Psychophysics.* Vol.52, No.6, 609-616.

Lu, Z.L and Dosher, B.A (1998) External Noise distinguishes Attention Mechanisms. Vision Research. Vol. 38, No. 9. 1183-1198.

Pouget, A., Dayan, P. and Zemel, A. (2003) Inference and Computation with Population Codes. *Ann. Rev. Neuroscience.* 26: 381-410.

Maier, A., Logothetis, N. and Leopold, D. (2005) Global competition dictates local suppression in pattern rivalry. *Journal of Vision.* No. 5, 668-677.

Mamassian, P. and Goutcher, R. (2005) Temporal dynamincs in Bistable Perception. *Journal of Vision.* No. 5, 361-375.

Mamassian, P. and Landy, M. (1998) Observer Biases in 3D Interpretation of Line Drawings. *Vision Research* (38) 2817-2832.

Mather G (1997) The use of image blur as a depth cue. Perception 26, 1147-1158.

Meng, M. and Tong, F. (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision.* 4, 539-551

Moran, J. and Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715) 782-784.

Necker, L.A. (1832) Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinbugh Philosophical Magazine and Journal of Science* 1. 329-337.

Pelton, L.H. and Solley, C.M. (1968) Acceleration of the reversals of the Necker Cube. *Am. J. Psychology* 81, 585-588.

Peterson, M.A and Gibson, B.S. (1994) Object recognition contributions to figure-ground organizations: operations on outlines and subjective contours. *Percept Psychophysics* 56, 551-564.

Peterson, M.A., Harvey, E.M. and Weidenbacher, H.J (1991) Shape recognition contributions to figure-ground reversals: which route counts? *J. Exp. Psychol Hum. Percept Perform.* 17, 1075-1089.

Rock , I. (1975) Introduction to Perception. Macmillian, New York.

Rock, I., Hall, S. and Davis J. (1994) Why do ambiguous figures reverse? *Acta Psychologica*, 87, 33-59.

Rock, I. and Mitchener, K.(1992) Further evidence of the failure of reversal of ambiguous figures by uninformed Participants. *Perception.* 21, 39-45.

Ross, Sheldon (1992) Applied Probability Models with Optimization Applications, Mineola, N.Y., Dover Press.

Schrater, P.R. and Sundareswara, R. (2006) Theory and Dynamics of Perceptual Bistability. *Neural Information Processing Systems*, 2006 (in press).

Sobel, K. and Blake, R.(2002) How context affects influences predominance during binocular rivalry. *Perception.* 31, 824-831.

Stocker, A. and Simoncelli, E. (2006) Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience.* vol.9, no.4, 578-585.

Suzuki, S., & Peterson, M. A. (2000). Multiplicative effects of intention on the perception of bistable apparent motion. *Psychological Science*, 11, 202–209.

Taylor, M.M. and Aldridge, K.D. (1974) Stochastic processes in reversing figure perception. *Perception and Psychopysics*, 16, 9–27

Toppino, T. C. and Long, G. M. (1987). Selective adaptation with reversible figures: don't change that channel. *Perception and Psychophysics*. 42(1), 37-48

Toppino, T. C. (2003). Reversible-figure perception: mechanisms of intentional control. *Perception and Psychophysics.* 65(8), 1285-1295.

van Ee, R., Adams, W. and Mamassian, P. (2003) Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J. Opt. Soc. Am. A/* Vol. 20, No. 7

van Ee, R., Noest, A.J., Brascamp, van den Berg, A.V. (2006) Attentional control over either of the two competing percepts of ambiguous stimuli revealed by a two-parameter analysis: Means do not make the difference. *Vision Research*, 46, 3129-3141.

van Ee, R., van Dam, L. C. J., & Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bi-stability. *Vision Research*, 41, 41–55.

Walker, P. (1978). Binocular rivalry: central or peripheral selective processes? Psychol. Bull. 85, 376–389.

Yu, K. and Blake, R. (1992) Do recognizable figures enjoy an advantage in binocular rivalry? *J. Exp. Psychology and Human Perception Performance*. 18, 1158-1173.

Yuille, A. and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci*, 10(7), 301-30