A decision-theoretic model of multistability: perceptual switches as internal actions

Shervin Safavi[1,2] ✉ **and Peter Dayan**[2,3]

[1]Computational Neuroscience, Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Germany.; [2]Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Germany.; [3]Department of Computer Science, University of Tübingen, Germany.

✉ **For correspondence:**
research@shervinsafavi.org (SS)

Abstract

Perceptual multistability has been studied for centuries using a diverse collection of approaches. Insights derived from this phenomenon range from core principles of information processing, such as perceptual inference, to high-level concerns, such as visual awareness. The dominant computational explanations of perceptual multistability are based on the Helmholtzian view of perception as inverse inference. However, these approaches struggle to account for the crucial role played by value, e.g., with percepts paired with reward dominating for longer periods than unpaired ones. In this study, we formulate perceptual multistability in terms of dynamic, value-based, choice, employing the formalism of a partially observable Markov decision process (POMDP). We use binocular rivalry as an example, considering different explicit and implicit sources of reward (and punishment) for each percept. The resulting values are time-dependent and influenced by novelty as a form of exploration. The solution of the POMDP is the optimal perceptual policy, and we show that this can replicate and explain several characteristics of binocular rivalry, ranging from classic hallmarks such as apparently spontaneous random switches with approximately gamma-distributed dominance periods to more subtle aspects such as the rich temporal dynamics of perceptual switching rates. Overall, our decision-theoretic perspective on perceptual multistability not only accounts for a wealth of unexplained data, but also opens up modern conceptions of internal reinforcement learning in service of understanding perceptual phenomena, and sensory processing more generally.

Introduction

Multistable perception is the perceptual phenomenon of dynamic alternation that arises when a single sensory input has more than one interpretation or explanation. This phenomenon has been studied for decades (Leopold et al., 1999; Blake et al., 2002; Brascamp, Klink, et al., 2015; Brascamp, Sterzer, et al., 2018; Safavi and Dayan, 2022), across a variety of species, including Drosophila (Toepfer et al., 2018), fish and reptiles (Carter et al., 2020), mice (Palagina et al., 2017) primates (Blake et al., 2002), and human (Zaretskaya, 2021), and almost all sensory modalities (Schwartz et al., 2012; Alais, Keys, et al., 2021). Studying the phenomenon has provided insights into a diverse set of cognitive functions such as attention (Maier et al., 2021), visual awareness (Seth et al., 2022), perceptual inference (Brascamp, Sterzer, et al., 2018), and perceptual decision-making (Krug, 2020). Moreover, various aspects of perceptual multistability have been found to differ in a range

of psychiatric conditions (see, e. g., Ye et al., 2019; Jia et al., 2020).

The traditional account of multistability is based on a passive account of perception in which it constitutes under-controlled inference about the contents of the world (Dayan, 1998; Schrater et al., 2007; Sundareswara et al., 2008; Hohwy et al., 2008a; Reichert et al., 2011; Moreno-Bote et al., 2011; Gershman et al., 2014; Weilnhammer et al., 2017; Leptourgos et al., 2020; Cao et al., 2021; Weilnhammer, Chikermane, et al., 2021) — also see Brascamp, Sterzer, et al. (2018), for a review. However, it is difficult to reconcile this formulation with a large body of recent data concerning both behaviour and the neural substrate of perceptual multistablity (Martin et al., 2021) which collectively imply that fundamental concepts such as reward (and broadly value) and choice are missing (for a review see, Safavi and Dayan, 2022). For instance, when a percept is paired with reward, it dominates for longer periods than if it is not; and, equally, multiple brain systems, such as the prefrontal cortex (Michel, 2022; Brascamp, Sterzer, et al., 2018; Safavi, Kapoor, et al., 2014) and the anterior cingulate cortex (Lyu et al., 2022; Drew et al., 2021; Gelbard-Sagiv et al., 2018), that are generally engaged in value-based operations such as cognitive control and decision-making are also involved in multistable perception. Certainly, areas of the brain are far from singular in their computational roles; however, reconciling the neural and behavioural data in light of a decision-theoretic approach to perceptual multistability could be more plausible compared to traditional value-free accounts.

We formulate perceptual multistability as a value-based decision-making process, extending the traditional account via the formalism of a partially observable Markov decision process (POMDP). As in conventional Bayesian accounts of perception, we consider there to be a single actual state of the world at any one time that must be inferred from sensory observations; but (in the simplest case), for which there are two plausible, but incompatible candidates (Dayan, 1998; Brascamp, Sterzer, et al., 2018). Each percept is potentially associated with different sources of rewards or punishments, and switching between percepts, more precisely perceptual selection, is a form of (costly) *internal action* (Dayan, 2012) – the attentional equivalent of the external action of moving eye gaze from one object to another. This makes perception arise as a choice rather than a passive inference. The approximate solution of the POMDP is the optimal perceptual policy, which we show to replicate and explain several aspects of rivalry. These range from paradigmatic hallmarks such as spontaneous (approximately) gamma-distributed dominance periods (Blake et al., 2002), to more elusive aspects such as reward modulation (Balcetis et al., 2012; Wilbertz, van Slooten, et al., 2014; Marx et al., 2015; Wilbertz, van Kemenade, et al., 2017) and the rich temporal dynamics of perception (Suzuki et al., 2007).

Our treatment of perceptual multistability is grounded in the value awarded to a pure sensory process - a form of aesthetic evaluation. Aesthetic aspects of an image can, in fact, bias perceptual preference in perceptual multistability (Huang et al., 2013; Mo et al., 2016; Shang et al., 2020; Tsikandilakis et al., 2019). Even in the frequently used case of rivalry between face and house images, faces, which occupy a special position in our aesthetic lives (Rhodes, 2006), have longer dominance times (Persike et al., 2014). Additionally, the relationship between complexity and aesthetics (Nath et al., 2024) implies complex stimuli also dominate for longer, which is experimentally confirmed (Rogers et al., 1977; Alais and Melcher, 2007; Daini et al., 2010). Furthermore, along with other cognitive factors, aesthetic value, with its evident ethological roots (Rhodes, 2006; Chittka et al., 2006; Iigaya, Yi, Wahle, Tanwisuth, et al., 2021), is a key component of the reward associated with a percept and thus the decision-theoretic reason to perceive it (Scocchia et al., 2014; Safavi and Dayan, 2022).

A particular characteristic of aesthetic value is that it changes over the course of observation, captured, for instance in modern theories (e.g., Brielmann and Dayan, 2022). We therefore examine the extent to which the temporal dynamics of multistability could be inherited from the temporal dynamics of aesthetics.

Lastly, along with these theoretical analyses, we suggest a range of empirical consequences of our account. For instance, we consider the implications for the involvement of value- and policy-

related brain circuits such as the anterior cingulate cortex (ACC, Lumer et al., 1998; Gelbard-Sagiv et al., 2018; Drew et al., 2021; Lyu et al., 2022), pulvinar (Wilke et al., 2009), locus coeruleus (LC, see, e. g., Einhauser et al., 2008), as well as (the controversial) involvement of prefrontal cortex (PFC, Zaretskaya and Narinyan, 2014; Safavi, Kapoor, et al., 2014; Tsuchiya, Wilke, et al., 2015; Odegaard et al., 2017; Block, 2019; Michel, 2022), which are more conventionally tied to value-dependent operations such as cognitive control and foraging (Montague et al., 2002; Niv, 2009; Lee et al., 2012; Monosov et al., 2020; Holroyd et al., 2021a; Rudebeck et al., 2022). Notably, this also includes the neural systems that have been suggested to compute aesthetic value, namely vlPFC (Iigaya, Yi, Wahle, Tanwisuth, et al., 2023) and ACC (Cloutier et al., 2008) which are also involved in perceptual multistability.

In sum, our decision-theoretic approach connects classical approaches to perception with modern ideas from reinforcement learning. Our approach, not only synergizes with Helmholtzian account of perceptual multistability (Dayan, 1998; Schrater et al., 2007; Sundareswara et al., 2008; Hohwy et al., 2008a; Reichert et al., 2011; Moreno-Bote et al., 2011; Gershman et al., 2014; Weilnhammer et al., 2017; Leptourgos et al., 2020; Cao et al., 2021) and offers a more comprehensive treatment of computational and algorithmic facets of perceptual multistability, but also it links new ideas about the genesis and effect of value with modern conceptions about internal reinforcement learning (Hazy et al., 2007; Dayan, 2012; Lieder et al., 2018) and thereby helps explain one of the most venerable and well-studied phenomena in sensory processing.

Results

We model the phenomenon of binocular rivalry, which is one of the most extensively studied forms of perceptual multistability (Blake et al., 2002). In the canonical form of binocular rivalry, a participant's two eyes are shown two distinct images (***Figure 1***A) over an extended period. This leads to spontaneous switches between two percepts, typically, each associated with one image. The image currently perceived is referred to as being dominant (***Figure 1***B); the periods of dominance more (Pastukhov et al., 2013), or less (Brascamp, Ee, et al., 2005), follow a gamma-like distribution (***Figure 1***C). There are also piecemeal percepts, which involve elements of both images, but, for simplicity, we do not consider them in our model.

Perceptual multistability has rich temporal dynamics. Suzuki et al. (2007) reported that the rate of perceptual switches changes on three distinct timescales: naive participants experience an initial increase in the rate of switches over the course of dozens of trials (***Figure 1***D; a similar observation was made by Brown, 1955). Then, over the course of a single trial (tens of seconds), participants experience a gradual decrease in the rate of perceptual switches (***Figure 1***E; van Ee, 2005, also reported a somewhat similar observation), Finally, the switching rate slowly increases across several sessions (***Figure 1***F; Weilnhammer, Fritsch, et al., 2021a, also reported a similar slow dynamics). Using our model, we replicate and explain these rich dynamics along with several other characteristics of perceptual multistablity.

Modelling binocular rivalry with a POMDP

We approach perceptual selection as an internal action, and dominance sequences as decision sequences. Thus, we model binocular rivalry (see ***Figure 2*** for a schematic) based on a well-established model of decision-making — the partially observable Markov decision process (POMDP, Littman, 2009, and see Methods section, for more details). Based on the conventional Bayesian accounts of perception (Yuille et al., 2006; Dayan, 1998; Brascamp, Sterzer, et al., 2018), we consider there to be a single actual state of the world at any one time that must be inferred from sensory observations. Given that, there are two input images, we assume that there are two plausible but incompatible candidates for this inference. However, Bayesian perception ends at the estimation of the posterior distribution over these two candidates; we take the further Bayesian decision-theoretic step *of considering an explicit choice between the candidates, treating this as an internal decision or action of*

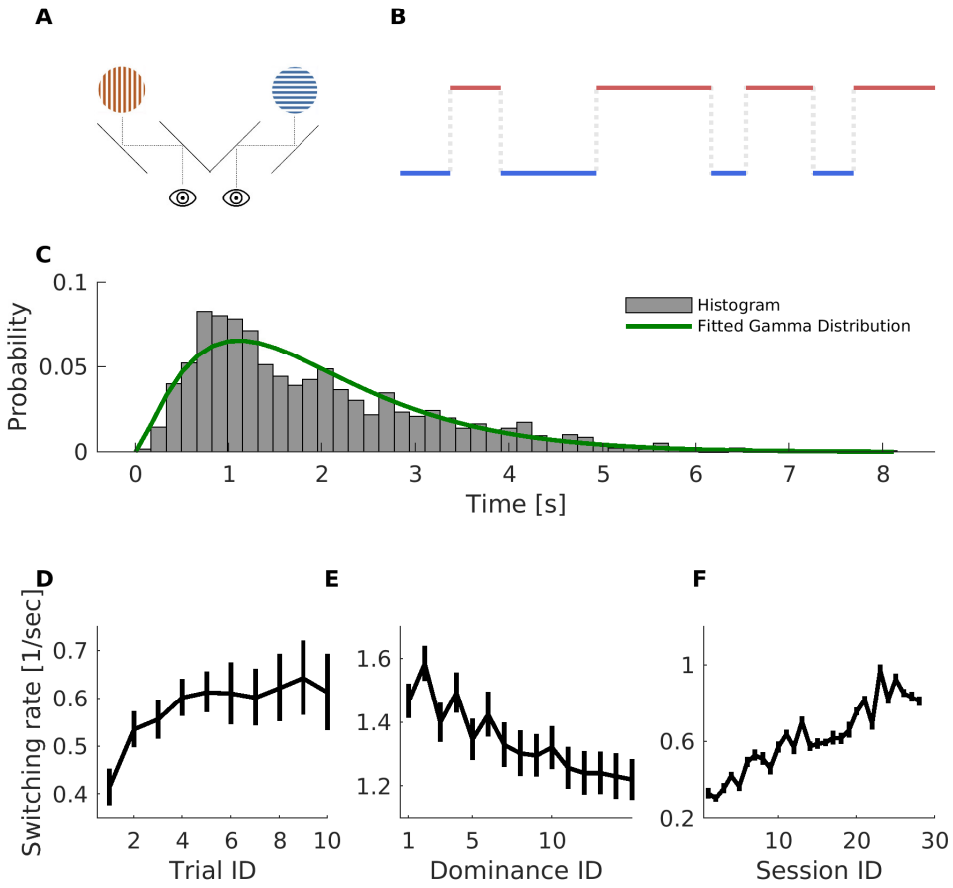


Figure 1. Basics of binocular rivalry (A) Binocular rivalry involves a participant's two eyes seeing images that differ by more than just depth disparity. **(B)** Participants typically perceive one of the presented images, vertical (left, red) or horizontal (right, blue) gratings with *dominance periods* (when either the left or right image is exclusively perceived), and *perceptual switches* (transition from one percept to the other). **(C)** Dominance durations in binocular rivalry share the common statistical property of being approximately gamma-like distributed. The example distribution is based on data presented in Kapoor et al. (2022). **(D)** The initial increase of switching rate over 10 trials, (including all dominance periods within a trial) and averaged across 40 participants for the first 5 trials, and 16 participants for the last 5 trials (resulting in larger error bars for trials 6 through 10; see, Suzuki et al., 2007, for more details). **(E)** Decreases in switching rate across dominance instances within individual trials (each data point is an average across 21 sessions measured for one participant). **(F)** Long-term speed-up of switching rate across sessions for one participant (individual data point for each session is average across all dominance periods in all 20 trials of each session, resulting in small error bars). In D-E, error bars are mean $\pm$ SEM, and data adopted from Suzuki et al. (2007).

perceptual selection (Dayan, 2012). This action is accompanied by reduced observation noise (see section Sensory, perceptual and decision processes), which, amongst other consequences, entails a stronger belief about the perceived state (dominant percept). The action is also endowed with sources of rewards or punishments, which are themselves dynamic. For instance, each image has an aesthetic value (how much we like it), which changes over time (e.g., as we get bored after multiple exposures; Brielmann, Berentelg, et al., 2023).

The agent needs to take into account all contributing factors (the momentary total utility of each possible percept, belief strength, etc) to decide which percept should dominate. These factors include conventional components from classical Helmholtzian treatments to perceptual multistability, as well as components originating in our value-centred conceptualisation. An example conventional component is the conventional loss that purely inference-based models of perceptual multistability emphasize (Dayan, 1998; Schrater et al., 2007; Sundareswara et al., 2008; Hohwy et al., 2008a; Reichert et al., 2011; Moreno-Bote et al., 2011; Gershman et al., 2014; Weilnhammer et al., 2017; Leptourgos et al., 2020; Cao et al., 2021), which counts the unlikelihood of the agent's observations given the current decision of the agent about the state of the world. An example of value-centered component is an approximate form of the value of information that is associated with a stimulus which is incompletely known (Dayan and Sejnowski, 1996, also see section Sensory, perceptual and decision processes, for more details).

The policy of the agent – i.e., the systematic (and potentially probabilistic) mapping of the images and the history of its past perceptual choices to its next sensory decision – is the solution of the POMDP. This means, finding the optimal actions, i.e., selecting actions that lead to the maximum utility for the agent (see section Partially observable Markov decision process, for more details). We show that this policy replicates, and thereby explains, several aspects of rivalry (as discussed in the next sections).

Simplified POMDP and multistability hallmarks

We start from a version of our POMDP (described in Methods) in which the reward function is simplified. This can already capture some key hallmarks of multistability. Perhaps the simplest contributor to reward is the aesthetic value of a percept (Brielmann and Dayan, 2022). This should be present across a wide range of experimental settings (since participants are always exposed to some form of sensory input). For our simplified model, we consider this value to evolve according to a simple, phenomenological, case of exponential decay. The rationale behind this choice is that aesthetic value wanes with boredom and/or fatigue – at least for the simple stimuli that are often used in binocular rivalry experiments (such as grating patterns).

Given this exponentially decaying reward and belief fluctuations, the agent spontaneously switches between percepts, leading to a variety of dominance periods (***Figure 3***A). The randomness in the switches comes from belief fluctuations rooted in observation noise. More precisely, when the aesthetic value decays sufficiently, fluctuations induced by the observation noise can lead to a switch (***Figure 3***B). Thus, there is a mean dominance period that depends on the characteristic decay of phenomenological aesthetic value, with variability stemming from observation noise (***Figure 3***A-B). Interestingly, spontaneous dominance periods are also (approximately) gamma-distributed (***Figure 3***C, green curve) as also observed in humans and monkeys (Blake et al., 2002). This is also preserved under further manipulations (see e.g., Supplementary Figure 1, for a setting with asymmetric reward).

The design of this simple model resonates with the key features of classical biophysical models of binocular rivalry, where noise and adaptation are the key factors shaping the dynamics. Our formulation offers a functional interpretation for traditional biophysical characteristics such as fatigue, e.g., putting such biophysical constructs next to other cognitive factors modulating perception (Scocchia et al., 2014; Safavi and Dayan, 2022). Notably, as discussed later (section Perceptual privileges through reward asymmetries and Levelt's propositions), this formulation allows us to explain the essential aspects of perceptual multistability such as Levelt's propositions (Lev-

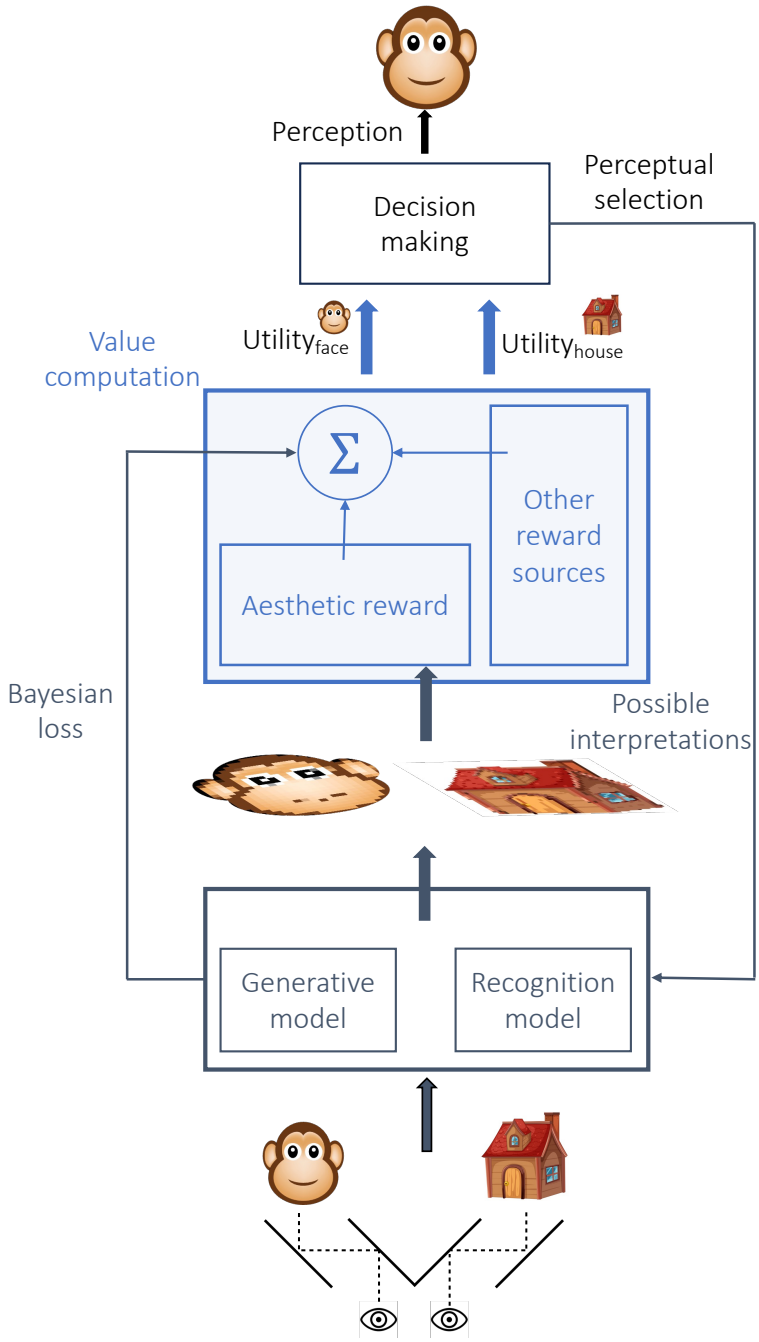


Figure 2. Schematic demonstration of the decision-theoretic model.
In our model, the agent receives two sensory inputs (one per eye). As in conventional Bayesian accounts of perception, we assume our agent has a generative and recognition model that leads to two possible interpretations of sensory inputs (also affected by noise). In the next stage, each possible interpretation of sensory input goes through a *valuation* process that determines the utility of each perception and how it evolves over time. The utility of each perception depends on various factors, such as associated reward (e.g., external and aesthetic reward), error of the generative model in explaining the input, and other factors. Perceptual action is the selection of a perception with a larger utility.

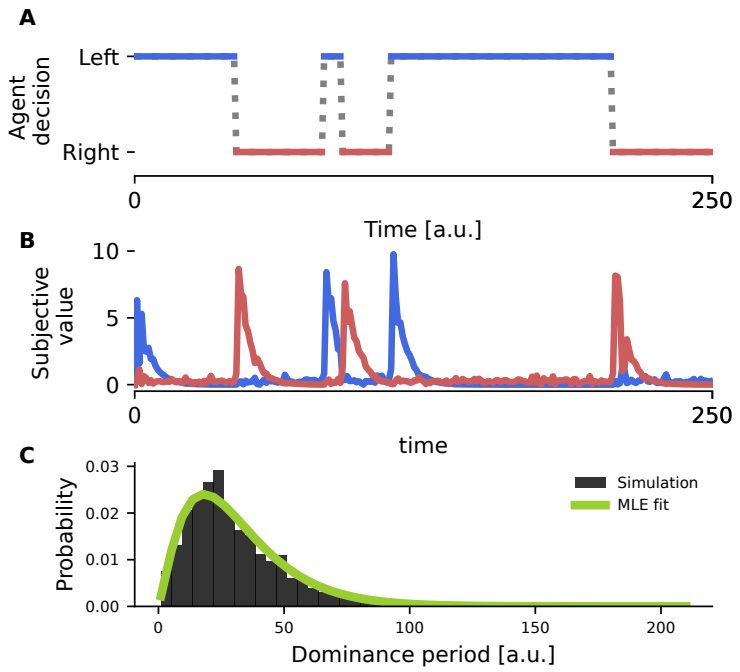


Figure 3. A simple POMDP captures the statistical hallmarks of perceptual multistability. **(A)** An example trajectory of the agent's decision sequence between left and right images in a binocular rivalry task. **(B)** Reward function of the simplified POMDP, which is a phenomenological aesthetic value modelled with exponential decay, which fluctuates due to observation noise. **(C)** Distribution of dominance periods for an extended length of simulation. Gamma distribution fitted with maximum likelihood estimation (MLE) and indicated with a green line.

elt, 1965), and modern conceptualization such as reward modulation of perception (Balcetis et al., 2012; Wilbertz, van Slooten, et al., 2014; Marx et al., 2015; Wilbertz, van Kemenade, et al., 2017), based on a common computational principle in a single framework (that is inspired by reinforcement learning).

Perceptual events are orchestrated by dynamics of the competing value

To explain a broader range of observations in multistablity we incorporate a more complex reward function in the model. Although the simple decision-making model of ***Figure 3*** can capture hallmarks of binocular rivalry (apparently spontaneous switches and approximately gamma-distributed dominance periods), the phenomenological account of aesthetic value does not have sufficient capacity to capture all the intricate aspects of multistability. Thus, in the following, we employ a more complex model of aesthetic value and demonstrate how it shapes the dynamic of perceptual switches.

A recent theory of the aesthetic value of sensory objects (Brielmann and Dayan, 2022) has suggested that, aesthetic value can be derived from the task of optimising a sensory system to process present and future stimuli well. In this framework, an agent is assumed to have a momentary or contextual probabilistic model of what it expects to sense now (called the 'system state') and a more general probabilistic model for the sensory input it expects in the long run (the 'expected true distribution'). Perceiving a stimulus (or in the case of binocular rivalry, one of the two stimuli) leads to a mandatory adaptation or learning of the system state. This then leads to two underlying components for the aesthetic value: a) the fluency with which it can be processed (i.e., how well it can be encompassed by the system state); and b) the effect of plasticity occasioned by perceiving this stimulus for the fluency with which likely future stimuli will be processed. As the system state moves, the values of the two components change, and so the aesthetic value changes too (described in more detail in Model of aesthetic value).

Within this model of aesthetic value, we make the substantial simplification of representing

stimuli (and the agent's estimate of the stimuli) in a two-dimensional feature space. In binocular rivalry, two stimuli are available to the agent, thus they are characterised by two distinct coordinates in this feature space (***Figure 4***A, blue and red circles). The initial location of the system state's mean (***Figure 4***A, black circle) is chosen such that it is equidistant from the location of both stimuli (which implies that the stimuli have equal aesthetic value).

The agent enjoys the dynamically changing aesthetic reward of the stimulus that is perceived at each moment. For instance, when the agent perceives the blue stimulus, it starts consuming its aesthetic reward. In the course of the perception of the blue stimulus, the system state gradually evolves toward the location of the blue stimulus in the feature space. The change in the location of the system state in relation to the blue stimulus and the expected true distribution determines how the aesthetic value changes over time. As the system state moves toward the blue stimulus (gets closer), the agent better accounts for and processes the blue stimulus (fluency component of the aesthetic value), however, the dynamism of the system state dwindles (learning component of the aesthetic value). This implies that the learning component of the aesthetic reward for the blue stimulus decreases over the period that the agent perceives this stimulus, and it implies that the aesthetic value (which is the sum of both the fluency and the learning components) of a dominant stimulus decreases over time (***Figure 4***B). This creates a *pushing* for a perceptual switch. At the same time, the system state gets further from the red stimulus, which also changes the aesthetic reward that the agent *could accrue* from the red stimulus *if* it were switched into. Thus, the agent continuously estimates the (counterfactual) aesthetic reward of the alternative perception (red stimulus in this case). Given that the system state is far and getting even further from the red stimulus, the (counterfactual) evolution of the system state toward the red stimulus would be large (see Model of aesthetic value, ***Equation 19***). In terms of the formulation of aesthetic value we use, it means that perception of the red stimulus would generate substantial learning, or in other words, this change has a high long-term utility for the agent (see Model of aesthetic value, and in particular ***Equation 22***). Thus, in the course of the perception of the currently dominant stimulus, i.e., closure of the system state to the dominant perception and distance from the suppressed one, the (counterfactual) aesthetic reward of the suppressed (red stimulus) gradually increases (due to learning component of aesthetic value). This is therefore *pulling* for a perceptual switch. At some point, more rewards would be available by switching to the alternative percept. These together create a periodic (yet random, due to belief fluctuations) push-pull mechanism, and thus perceptual alternations (***Figure 4***D).

Our proposed push-pull mechanism for perceptual switching also resonates with dynamic systems models of perceptual multistability (e. g., Brascamp, van Ee, et al., 2006). In these models, the adaptation of the dominant percept may correspond to a push factor, and recovery from that adaptation may act as a pull factor, particularly noting the gradual nature of perceptual switches described in these models. However, the pull factor in our model can be cognitively richer. For instance, in continuous flash suppression (CFS; see Tsuchiya and Koch, 2004, a similar phenomenon to binocular rivalry, where a contentful image presented to one eye is suppressed by dynamic patterns presented to the other eye), faces with fearful expressions have shorter suppression periods compared to neutral or happy expressions. In such cases, it has been suggested that fearful faces are privileged because of their emotional content (Yang et al., 2007; Sterzer et al., 2011), which, in the context of our framework, can be considered a stronger pulling force by the suppressed percept for switching (or, more precisely for CFS, breaking through awareness). The same applies to other forms of cognitive associations (e. g., for the association of external reward, see Lunghi et al., 2023).

Overall, our results suggest that the dynamics of perception are shaped (at least in part) by the competing and dynamic value of sensory stimuli in binocular rivalry. Notably, this relies not only on the value of the perceived stimuli, but also, on the value of the alternative perception and its dynamics (notably that can be cognitively rich, e.g., through affective charges). This provides an *active* perspective compared to classical biophysical and normative models where switches are driven

primarily by (manually included) adaptation/fatigue and noise, i.e., a push due to the cessation of the currently dominant percept, and recovery from adaptation as the pull factor.

Perceptual privileges through reward asymmetries and Levelt's propositions

As mentioned earlier, multiple sensory and cognitive factors can favour particular percepts in the context of perceptual multistablity. We consider asymmetries in reward function by manipulating different components of this function (e.g., differential aesthetic preferences) and demonstrate it can explain several established aspects of perceptual multistability.

We start with the instrumental imperative to execute actions associated with an aspect of the percept (e. g., to avoid a negative outcome or gain a positive one). In the simplest setting where no instrumental reward is involved (***Figure 5***Ai, resembling the most conventional binocular rivalry experiments), we observe equi-dominance (***Figure 5***Bi, as seen in previous studies, Moreno-Bote et al., 2010). However, consider the case where there is an explicitly learned association between percept and reward. This can institute an asymmetry in the rewards associated with perceiving a stimulus, so that either the left (***Figure 5***Aii) right (***Figure 5***Aiii) images have a higher value (as empirically tested in, Wilbertz, van Slooten, et al., 2014; Marx et al., 2015), leading to increased dominance of the corresponding perception (***Figure 5*** Bii and Biii, Wilcoxon rank-sum test, $p < 10^{-20}$).

We then investigate the perceptual privilege attributed to the aesthetic value of sensory input. For instance, it has been shown that aesthetic aspects of an image can bias the perceptual preference in binocular rivalry and other related paradigms (Huang et al., 2013; Mo et al., 2016; Shang et al., 2020; Tsikandilakis et al., 2019). To replicate the key findings of this category of studies (i.e., longer dominance of aesthetically preferred images), we chose the parameters of the aesthetic component of our model, such that, the aesthetic value of one image is higher than the other one. We did this by slightly changing one of the coordinates of one of the stimuli, altering the distance of stimulus from the mean of the system state distribution (turning the symmetric case ***Figure 5***Ci into the asymmetric Cii and Ciii). Here again, we have equi-dominance for the symmetric case (***Figure 5***Di), and biased dominance durations for asymmetric reward structures (***Figure 5***Dii-Diii), similar to external reward (***Figure 5***A-B).

Through asymmetries in the reward function and the dynamics of the reward function, we can also explain all Levelt (1965)'s propositions, that state how perception during binocular rivalry depends on the attributes of the presented stimuli. Levelt's propositions state (adapted from, Brascamp, Klink, et al., 2015):

I. Increasing stimulus strength for one eye will increase the perceptual predominance of that eye's stimulus.
II. Increasing the difference in stimulus strength between the two eyes will primarily act to increase the average duration of perceptual dominance of the stronger stimulus.
III. Any manipulation that increases the difference between the two eyes' stimulus strengths will reduce the perceptual alternation rate.
IV. Increasing the stimulus strength in both eyes while keeping it equal between eyes will generally increase the perceptual alternation rate, but this effect may reverse at near-threshold stimulus strengths.

The key to compatibility with Levelt's propositions is having the Helmholtzian notion of sensation (Dayan, 1998; Brascamp, Sterzer, et al., 2018) in our model. The Helmholtzian character of our reward function comes from two sources. One is as part of the aesthetic value (weighted likelihood of the input, or fluency component of the aesthetic reward, see section Model of aesthetic value). The other, is as the agent's estimation of the unlikelihood of its sensory input (for a more formal treatment, see Methods section). The former contributes as a reward for the portion of the sensory input that the agent can explain, and the latter, as the penalty, for the portion that the agent cannot explain (the mismatch between percept and actual sensory input).

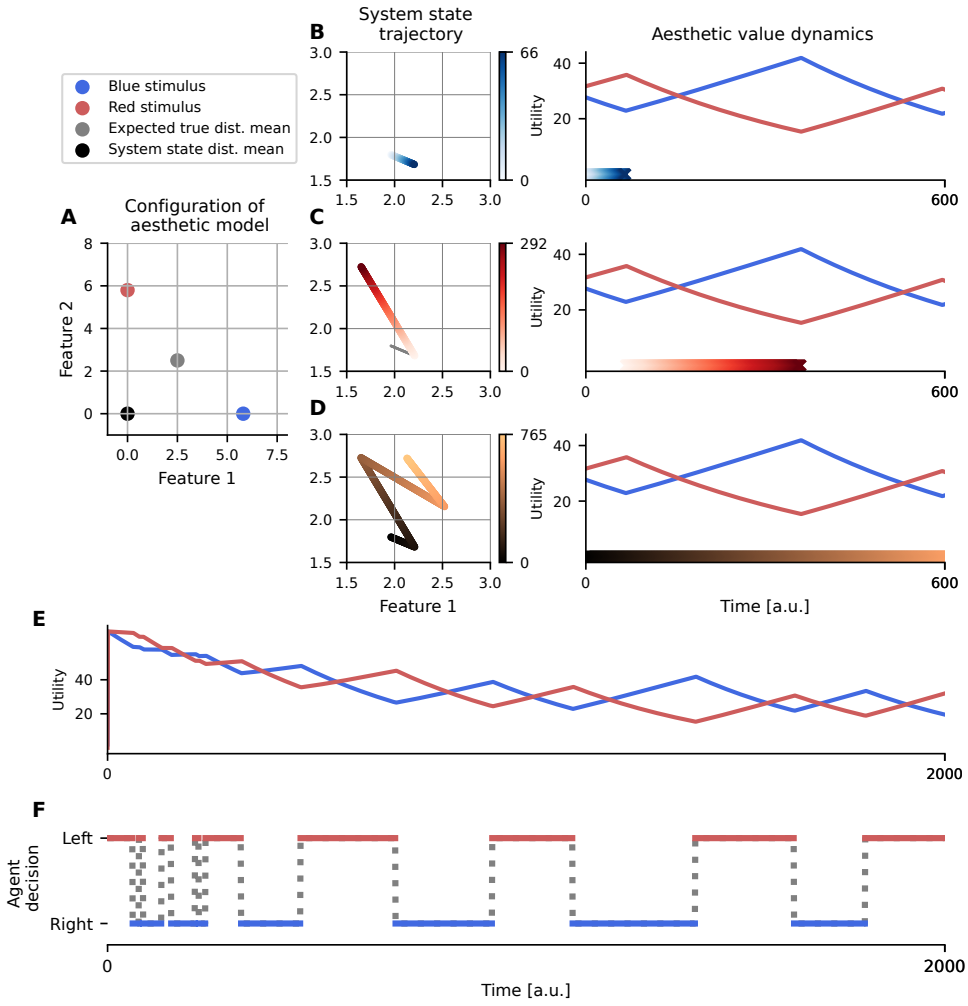


Figure 4. Push-pull mechanism of perceptual alternations governed by dynamics of value. **(A)** Representation of two rivalling stimuli (red and blue circles) and the initial location of the system state mean and expected true distribution mean in the model of aesthetic value (black and grey circle respectively). **(B)** When the blue stimulus is the dominant percept, the system state moves toward its representation in the feature space (left figure) and the aesthetic value of the blue stimuli decays (blue trace in the right figure). Simultaneously, the system state gets further from the red stimulus, which leads to an increase in aesthetic value for the red stimulus if it becomes dominant (red trace in the right figure). Temporal evolution is colour-coded in both left and right figures according to the gradient colourbar in the middle. **(C-D)** Similar to (B), but for the second perceptual switch (C) and both switches together (D). **(E)** Global dynamics of the aesthetic reward for red and blue stimuli during an extended period of binocular stimulation. **(F)** Similar to ***Figure 3***A, an exemplary trajectory of the agent's decision sequence temporally aligned with (E) during the binocular rivalry task.

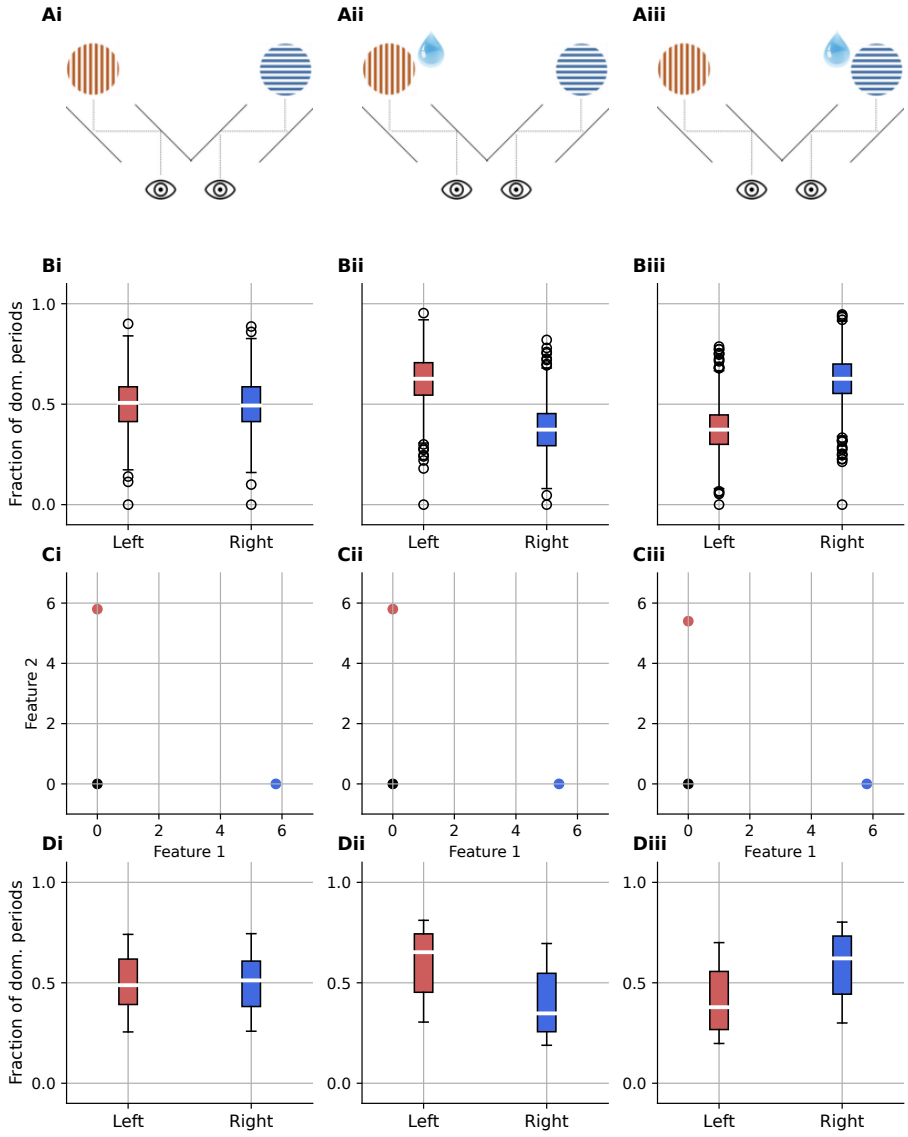


Figure 5. Modulation of dominance periods by different sources of reward.
(Ai-Aiii) Reward structure of the binocular simulations: (Ai) no reward asymmetry, (Aii) reward associated with the left stimulus, and (Aiii) right stimulus. **(Bi-Biii)** Distribution of dominance periods (as a fraction of the entire trial) for left and right stimuli in different reward structures corresponding to Ai-Aiii. **(Ci-Ciii)** Different geometry of parameters in the aesthetic model to create similar (symmetric or asymmetric) reward structures as in Ai-Aiii. Each panel represents the location of two stimuli of binocular rivalry (red and blue circles) and the mean of the system state (black circle) as follows: (Ci) The locations of the stimuli (red and blue circles) are chosen such that they are equidistant from the location of the mean of the system state (black circle), which implies that the stimuli have equal aesthetic values. (Cii) Similar to Ci, but with a slightly different location for blue stimuli to create an asymmetry in reward structure. (Ciii) Similar to Cii, but with a slightly different location of the red stimulus (instead of the blue one). **(Di-Diii)** Similar to Bi-Biii, but with asymmetry in reward structure induced with aesthetic reward.

In conventional, symmetric settings for binocular rivalry, the Helmholtzian contribution from unlikelihood estimation may not play a significant role. If we assume in a typical setting of binocular rivalry, sensory evidence provided by either of the presented stimuli is equally unlikely (e.g., they are equally noisy, dim, etc), then any of the possible interpretations of the sensory input leave almost half of the sensory input unexplained. Thus, in such settings, irrespective of the perceptual selection of the agent, the agent's estimation of the unlikelihood of its sensory inputs remains almost constant over time, and thus, it may not play a significant role in shaping the fine-grained temporal dynamics of switching. This is not the case for the Helmholtzian contribution attributed to aesthetic value.

In the majority of Levelt's propositions (except the last one), an asymmetry in the strength of the evidence for the stimuli is the central condition of the proposition (e.g., lower or higher contrast of one of the stimuli). Thus, by mapping the asymmetries in Levelt's propositions (i.e., asymmetries in the attribute of stimuli) to the unlikelihood term of our reward function (***Equation 11***) we can readily explain these three propositions.

Corresponding settings to the first and second propositions are quite similar to the simple reward asymmetry discussed earlier (***Figure 5***), however instead of the task reward, the source of privilege is stronger evidence on one percept (the unlikelihood term) that lead to an excess of overall reward magnitude of one percept (or more precisely, more penalty for worse explanation of the weaker input). This leads to the first dominance (predominance) of the privileged stimulus (for proposition I), and overall longer dominance of the privileged stimulus (for proposition II), similar to ***Figure 5*** A-B. Our model is also compatible with the more delicate aspects of proposition II, which emphasizes the *difference* between the strength of the left and right stimuli. In our model, the difference between the expected utilities is also central (given the expected utilities of percepts are compared to each other). Thus, any changes in the strength of stimuli that lead to an asymmetry, result in a difference in the utility of left and right perception and, consequently, an asymmetric dominance duration.

The third proposition can also be explained similarly based on induced reward asymmetries that are also compatible with previous accounts (e.g., Leptourgos et al., 2020). Previous accounts of the third proposition relied on neural adaptation (Brascamp, Klink, et al., 2015). As we discussed earlier (see ***Figure 3*** and ***Figure 4***), adaptation is also present in our framework. Through this, manipulations that increase the difference between the strengths of the two stimuli, arrange for the value of one percept to remain higher than that of the other for a longer period of time (as compared to the case where both have equal strength and decay together, see ***Figure 4***E).

We can explain the last proposition (IV) based on the potential effect of the stimulus strength on aesthetic value and its temporal dynamics. We implicitly assumed that the utility stemming from basic aspects of sensory input (e.g., contrast) is part of the classical Bayesian loss function (our Helmholtzian unlikelihood components, ***Equation 11***), however, it can also be relevant in terms of aesthetic judgments (see, e. g., Tinio et al., 2011), and as a consequence, can also affect the temporal dynamics of perception. Thus, we refine our previous assumption and propose stronger stimuli (e.g., images with higher contrast), at least for a range of values have higher aesthetic values (Tinio et al., 2011).

In the model of aesthetic value that we employed, we can change the magnitude of aesthetic value by changing the geometry of the model (e.g., changing the coordinates of the stimuli in the feature space). Here, we need only change the distance between rivalling stimuli and the system state's mean (i.e., making red and blue circles in ***Figure 4***A, closer or further from the black circle, as we also used in ***Figure 5***C). As a consequence, a change in the aesthetic value (due to distance) also leads to a change in the rate of decay of the aesthetic value, i.e., the higher the aesthetic value, the greater the decay rate. When both stimuli have the same strength (as stated in the fourth proposition), a faster decay of the aesthetic value of the perceived (dominant) stimulus also implies a faster growth of the (counterfactual) aesthetic reward of the suppressed percept (***Figure 4***E). Similarly, when the agent is exposed to weak stimuli in both eyes, the value of the dominant percept

decays slowly, and the counterfactual value of the suppressed percept grows proportionally sluggishly. These together lead to faster (or slower) alternation when the strength of stimuli on both sides is higher (or weaker). The contribution of the suppressed precept (through counterfactual reward) could be even more important, given that abbreviated periods of suppression have been more emphasised in previous studies (see, e. g., Blake et al., 2002). For near-threshold stimulus strengths where a reversal has been observed (i.e., Brascamp, Klink, et al., 2015, a decrease rather increase in alternation rate was observed), other factors may contribute more (e.g., changes in the level of exploration due to low belief value in a POMDP or other aspects of inference processes) and need to be investigated further. However, it should be noted that the reversal was not observed in all cases (e.g., changing dot luminance contrast did not elicit a reversal Brascamp, Klink, et al., 2015).

Rich temporal dynamics of binocular rivalry

As noted earlier, perceptual multistability exhibits rich temporal dynamics (see, Suzuki et al., 2007, and see ***Figure 6*** right panels) that is not completely understood. These rich dynamics in the rate of perceptual switches span a wide range of timescales: an initial increase over the course of dozens of trials (***Figure 6***A, right); a gradual decrease within a single trial (tens of seconds, ***Figure 6***B, right); a slow increases across several sessions (***Figure 6***C, right). We demonstrate, by employing a sufficiently complex reward function we are able to replicate and explain this rich temporal dynamics.

We suggest that the key factor shaping the initial speed-up of the switching in naive participants (***Figure 6***A) is the tendency of the agent to explore or better understand the scene (noting that Suzuki et al., 2007, only observed it in naive participants). As discussed in the Methods section, we approximate the value of information associated with a stimulus that is incompletely known (Dayan and Sejnowski, 1996) with an exponentially decreasing exploration bonus. At the beginning, stimuli enjoy a high exploration bonus; but this decays as the agent perceives them, and so gets to know them better. This implies that the rate of perceptual switches is initially low (the agent experiences longer dominance periods) and gradually increases when the exploration bonus of both stimuli is equally low. This is captured by our simulation (***Figure 6***A left) as reported by Suzuki et al. (2007, ***Figure 6***A right). There are also more principled approaches to potentially explain these dynamics, for instance, based on richer hierarchical inference for novel stimuli that are further discussed in section Limitations and future extensions (also see our more elaborate description of the state, section Thick state).

We suggest the gradual slow-down in the rate of perceptual switches is governed by the way that the dynamics of value evolves within the timescale of a single trial (tens of seconds). As discussed above (***Figure 4***E-F), the push-pull dynamics induced by the aesthetic value shapes the dominance periods and switching times. Along the same lines, but across several switches, aesthetic value can have a characteristic dynamic that shapes the length of dominance periods and timing of the switches in the course of a trial (***Figure 4***E-F). At the beginning of the trial, switches are fast due to the strength of the push-pull mechanism. But aesthetic value also globally decays (***Figure 4***E). This implies that the push-pull mechanism gradually loses its strength and the switches primarily are governed by noise. This is also well captured by our model (***Figure 6***B left) as was observed empirically (***Figure 6***B right).

For the longest timescale of perceptual dynamics, we presume that extensive perception of stimuli over several sessions (across days), increases the familiarity of the stimuli for the agent with the effect of reducing the switching cost. To be concrete, we made this be subject to exponential decay over multiple sessions. This heuristic leads to an increase in the tendency of the agent to switch and so captures the slow dynamics of the switching rate (***Figure 6***B left). Suzuki et al. (2007) observed a large variability in the characteristics of the acceleration of the switching rate between the participants. We speculate such individual differences can be attributed to the agent's parameter settings. For instance, a speed-up in switching rate with a larger dynamic range can be reproduced by a lower initial value of the switching cost (Supplementary Figure 2).

Overall, these results demonstrate that a sufficiently complex reward function can also capture the rich temporal dynamics of perception during binocular rivalry, and offer an explanation for the underlying computational mechanism. This implies a novel perspective on perceptual switches, where transitions from one percept to other ones are affected by a variety of cognitive factors in addition to basic bottom-up sensory processes (such as perceptual modulation due to contrast). Furthermore, these cognitive factors not only concern the attributes of the dominant percept (which has been the main focus in most previous studies), but also the suppressed percept.

Methods

Perceptual multistability as a POMDP

Partially observable Markov decision process

Humans and animals (generically, agents) frequently encounter situations in which they repeatedly need to choose one action among multiple alternatives. Each action in a sequence must be made not only based on its immediate effect in terms of gaining reward, but also based on any long-term ramifications from changing the circumstances or state of the agent. A Markov decision process (MDP, Sutton et al., 2018) is a general framework for modelling a wide range of such decision processes. The four components of an MDP are: states (specifying the circumstances of the human or animal in the external world); actions (the choice between available alternatives); transition probabilities (characterising how the state of the world might change given each possible action); and rewards, which specify the loss or gain associated with the state, action, and transition. Long-run values, a measure of the expected summed future reward, are a consequence of the agent's policy (i.e., the way it chooses to act at each state); and this policy can be optimised to maximise those long-term values.

Although MDPs have proved very useful for characterising many environments, one restrictive assumption is that the actor must know the state completely at all times (for instance, based on sensory input). This is not valid for many real-world problems, including our characterisation of perceptual multistability, where we assume that agents also make decisions about what to perceive and for how long (Martin et al., 2021; Safavi and Dayan, 2022). We therefore use a more complicated but more realistic class of models called *partially observable* MDPs or POMDPs (Littman, 2009). POMDPs are MDPs in which the agent may be uncertain about the state of the world, for instance, due to noisy sensory information. This noise is the product of an observation function that assigns probabilities to possible observations depending on the agent's states and actions. The agent then needs to form and update a belief about the state of the world in light of observations. It can also choose actions based on the information they are expected to provide about the state, i.e., on the value of the reduction in uncertainty about the state to which they are expected to lead. The agent's task remains to determine a policy that maximises long-term values. In the following, we describe the key components of the POMDP in relation to perceptual multistability, and leave more generic aspects and procedures for the Appendix (see section Partially observable Markov decision process).

Sensory, perceptual and decision processes

As in conventional Bayesian accounts of perception, we assume that our decision-making agent has a generative model of the world (Hinton et al., 1997; Dayan, 1997; Rao et al., 1999; Doya et al., 2007; Yuille et al., 2006). which specifies what input activity would be generated given a particular state of the world. Perception is then the recognition task that inverts the generative model, i.e., mapping the activities in the sensory system to the percept (or strictly, a distribution over possible percepts). We consider there to be a single actual state $s_t \in \mathcal{S}$ of the world at any one time t, that must be inferred from sensory observations. However, rather than sticking to the regular probabilistic recognition model (inverse of the generative model, Dayan, 1997), we consider the fuller body of Bayesian decision theory for recognition. This is a key difference between this model,

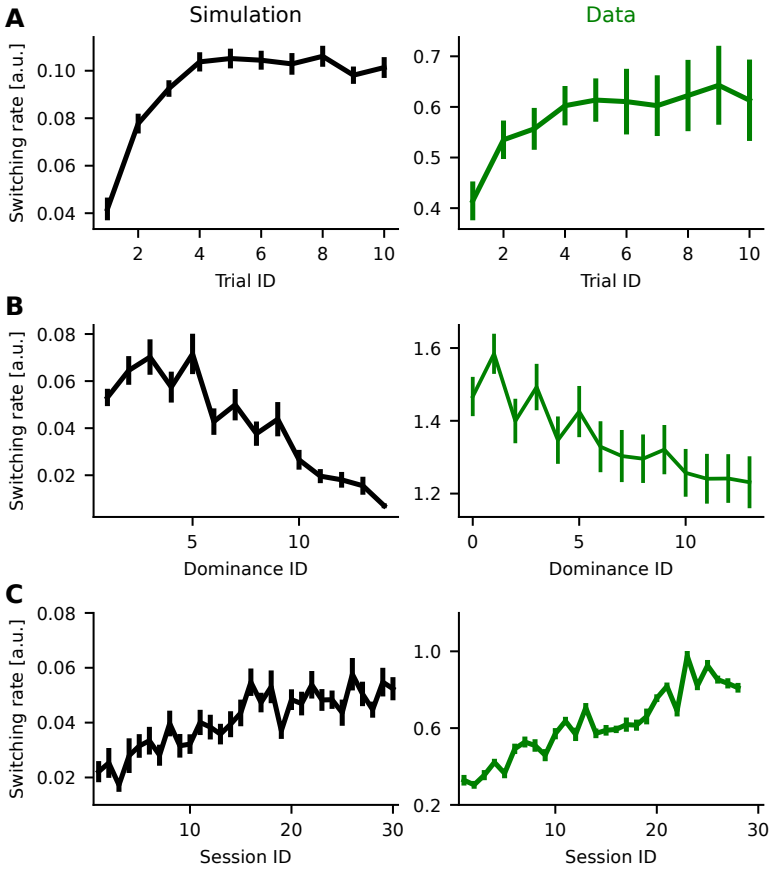


Figure 6. Capturing temporal dynamics of perceptual multistability at multiple timescales. **(A-C)** Modelled (left) and behavioural data of switching rates (right, from Suzuki et al., 2007) quantified as the reciprocal of the duration of perceptual dominance. **(A)** The initial increase of switching rate over 10 trials,(including all dominance periods within a trial) and averaged across multiple participants. The number of participants, varies across trials in empirical data, and it is 40 for simulation. For the data panel (right, green), switching rate were averaged across 40 participants for the first 5 trials, and 16 participants for last 5 trials (resulting in larger error bars for trials 6 through 10; see, Suzuki et al., 2007, for more details). **(B)** Decreases in switching rate across dominance instances within individual trials. **(C)** Long-term speed-up of switching rate across sessions for one participant, and a single simulated participant (individual data point for each session is average across 20 trials). The left panels are from ***Figure 1***D-F, and are presented for easier comparison with simulations (right panels). In all panels, error bars are mean ± SEM.

and previous models of perceptual multistability with the same spirit (Dayan, 1997; Gershman et al., 2014).

We build our account in two stages: first, assuming a very simple generative model containing only two possible states (see Thin state); and second, considering that the states have extra dimensions of internal structure (see Thick state) that need to be inferred over time from sensory input in a way that can benefit from the allocation of attention (which we treat itself as another component of the action space for the POMDP).

Thin state

For the simplest formulation of binocular rivalry, we assume there are only two plausible, but incompatible candidates for this state, $\mathcal{S} = \{+1, -1\}$ (Dayan, 1998). Thus, the agent should make an inference about the actual state at time t – this is the perceptual decision $a_t^P \in \mathcal{A} = \{+1, -1\}$. $\mathcal{S}$ and $\mathcal{A}$ are the state and perceptual action spaces of our perceptual POMDP; what remains is to specify the observations, transitions, and rewards.

In general, the observation function of a POMDP, $O = O(s_t, a_t^P, \mathbf{o}_t)$, expresses the probability distribution over possible observations $\mathbf{o}_t$ made at time t, given that the agent took action a_t^P and ended up in state s_t. In our case, the observations only depend on the current state s_t of the world, and not the perceptual action. We make the simplifying assumption that there is one scalar input for each eye, with Gaussian noise, so

$$O(s_t, a_{t-1}^P, \mathbf{o}_t) = P(\mathbf{o}_t | s_t) \sim \mathcal{N}\left[s_t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma \right] \tag{1}$$

The mean of this distribution, expresses the expectation that the eyes should see the same stimulus. The covariance matrix $\Sigma = \sigma^2 \mathcal{I}$ expresses the noise surrounding this stimulus; for the moment, we consider this to be proportional to the identity matrix ($\mathcal{I}$). Note that, in the case of binocular rivalry, a characteristic observation might express the conflict between the content of the two eyes (and the noise).

The transition function $T = T(s_{t-1}, a_t^P, s_t)$ of the POMDP expresses the probability distribution over the ending state, s_t, given that the agent started from state s_{t-1} and took action a_t^P. This accommodates the possibility of change in the environment, and so will be intimately tied to the dynamic nature of our perceptual experience (Pastukhov et al., 2022). Again, for our perceptual POMDP, the transition is independent of the perceptual action, leaving

$$T(s_{t-1}, s_t) = \begin{bmatrix} p_{\text{stay}} & 1 - p_{\text{stay}} \\ 1 - p_{\text{stay}} & p_{\text{stay}} \end{bmatrix}, \tag{2}$$

where p_{stay} denotes the probability of staying in the original state of the world, and we assume is known to the agent.

The reward function has a number of components: a loss function $\mathcal{L}(a_t^P, s_t)$ associated with a mismatch between the percept a_t^P and the state s_t as informed by the history of observations up to time t, $\mathcal{O}_t = \{\mathbf{o}_1, \ldots, \mathbf{o}_t\}$, a cost $c(a_t^P, a_{t-1}^P, t)$ of switching from percept a_{t-1}^P to percept a_t^P which can vary over a long time-scale, an aesthetic reward $R_A(a_t^P, t)$, an (heuristic) exploration bonus $R_{\text{exp}}(a_t^P, t)$, and an external reward $R_{\text{ext}}(a_t^P, t)$ associated with what the agent can achieve the world based on the percept (as in various of the experiments described above). We cover these in turn.

The standard Bayesian decision theory articulates the loss associated with choosing action a_t^P when the true state of the world is s_t. In a filtering context, one Helmholtzian possibility for this (slightly abusing the notation) is

$$\mathcal{L}(a_t^P, s_t) = -\log P(a_t^P \neq s_t | \mathcal{O}_t) \tag{3}$$

effectively counting the unlikelihood of the observations $\mathcal{O}_t$ given the current assignment of state a_t^P. A prominent component of this is the negative log probability of the current stimulus $\mathbf{o}_t$ under

a_t^P

$$-\log P(\mathbf{o}_t|s_t = a_t^P) = \mathcal{K} + \frac{1}{2}(\mathbf{o}_t - a_t^P\mathbf{n})^T\Sigma^{-1}(\mathbf{o}_t - a_t^P\mathbf{n}) \tag{4}$$

where $\mathbf{n} = (1,1)^T$ and $\mathcal{K}$ is a constant. Variants of this unlikelihood can explain perceptual preferences for higher (over lower) contrast inputs in rivalry, when the penalty for not explaining the high-contrast input is higher.

Switching between percepts (i.e., if $a_t^P \neq a_{t-1}^P$) is a form of costly *internal action* (Dayan, 2012) – the equivalent of the external action of moving the eye gaze from one object to another (e.g., Callaway et al., 2021). We allow the switching cost, $c(a_t^P, a_{t-1}^P, t)$, to change over a very long timescale (i.e., experimental sessions conducted across days). That is, we assume that, within each experimental session (as structured in Suzuki et al., 2007) with multiple trials (each with length τ_{tr}, with an inter-trial interval of a few seconds), the switching cost remains constant. However, this constant switching cost decreases exponentially over multiple sessions (at least spread over days, as in Suzuki et al., 2007), for instance from an increase in the familiarity of the stimuli.

Balancing out these costs are three rewards. The aesthetic reward, $R_A(a_t^P, t)$, follows the model of Brielmann and Dayan (2022, as briefly described in section Model of aesthetic value in the Appendix). This theory suggests that aesthetic value is derived from optimising a sensory system to process present and future stimuli well, assuming, consistent with our Helmholtzian picture, that the agent has a *momentary or contextual* probabilistic model over the whole collection of stimuli it thinks it might sense (denoted as the system state) and a permanent probabilistic model for the sensory input it expects in the future (called the expected true distribution). Thus, the aesthetic reward is computed on the basis of current and past percepts and agent's generative models. At every moment, the agent computes and accrues the aesthetic reward from the perceived (dominant) stimulus, but also computes a counterfactual aesthetic reward for the suppressed percept. Both consumed and counterfactual aesthetic rewards are computed as described in section Model of aesthetic value in Appendix, with the key difference that the consumed one is computed based on the location of dominant stimulus in the feature space and the counterfactual one, based on the location of the suppressed stimulus.

We also consider a heuristic exploration bonus, $R_{\text{exp}}(a_t^P, t)$. This is an approximation to the value of information associated with a stimulus that is incompletely known (Dayan and Sejnowski, 1996). This utility is attributed to all possible states, and decreases during the time that state is being perceived. With the current, binary, state space, this term is notional, and we include it for completeness; when there is information content to a state, this stands in for the benefit of learning about this information. For the moment, we consider that, as the dominance time of a given state increases, the stimulus becomes better known, at least up to forgetting, and thus less novel. We define it as:

$$R_{\text{exp}}(a_t^P, t) = \omega_{\text{exp}} \exp(-T_a(t)/\tau_{\text{exp}}), \tag{5}$$

where ω_{exp} determines the gross contribution of this component, $T_a(t)$ is the integrated total dominance time of percept a, and τ_{exp} is the time constant according to which the bonus decreases. We assume the agent computes the integrated dominance $T_a(t)$ with a forgetting rate of η, given the limitation on the memory on past perceptual experience within a session and the possibility of change. Let $T_a^j(t)$ be the total dominance period of percept a in trial j up to time t; then we assume the cumulative dominance $T_a(t)$ (within the session) is the sum of the dominance period of the current trial and the discounted value from the previous trial (due to forgetting),

$$T_a(t) = T_a^j(t) + \eta T_a^{j-1}(\tau_{\text{tr}}). \tag{6}$$

where τ_{tr} is the length of a trial.

Finally, the external reward $R_{\text{ext}}(a_t^P, t)$ is, for our current purposes, a constant term (to be assigned to one particular percept) to account for rewards induced by the experimentalist such as associating reward to a stimulus or a percept (similar to the experimental design of previous studies, Balcetis et al., 2012; Wilbertz, van Slooten, et al., 2014; Marx et al., 2015; Wilbertz, van Kemenade,

et al., 2017; Wilbertz and Sterzer, 2018) and/or rewards associated with an external, visually-guided task (as introduced in previous studies Paffen et al., 2008; Dieter et al., 2016; Moreno-Sánchez et al., 2019; Wang et al., 2021). Elements of this external reward (for instance, other remunerated decisions that can be made on the basis of the nature of the state) are rather nominal in the simplistic binary formulation. The full reward at time t is then

$$R(s_t, a_t^P, t) = -\mathcal{L}(a_t^P, s_t) - c(a_t^P, a_{t-1}^P, \tau^s) + R_A(a_t^P, t) + R_{\text{exp}}(a_t^P, t) + R_{\text{ext}}(a_t^P, t) \tag{7}$$

The solution of the POMDP is the optimal perceptual policy that maximizes the long-run summed expected reward (see Partially observable Markov decision process in the Appendix for more details). In the general formulation of a POMDP, both the immediate and the expected future reward (which here is dynamic, as not all the terms in the reward function are stationary) should be considered, but, to keep the model as simple as possible, we only consider the immediate reward, i.e., solving it myopically.

Thick state

Although the version of the model with a thin, binary, state is convenient to illustrate the workings of the model, it lacks the depth to encompass some key effects associated with the three forms of reward. In a more complete setting, we would have richer states $\mathbf{s}$, which have internal structure that can be inferred from the observations. Here, given the state $\mathbf{s}_t$, the mean of the observations would be $\mathbf{s}_t \otimes \mathbf{n}$. However, in rivalry, the actual input would be a noisy version of $\mathbf{s}^l$ to the left eye and $\mathbf{s}^r$ to the right eye (just as it was $+1$ to the left eye and -1 to the right eye in the thin state model). The agent has to make inferences about $\mathbf{s}_t$ given this conflicting input over time, and its expectations about change.

Here we ignore the possibility of mixed or averaged percepts (but see, Dayan, 1998), and only consider ocular rather than stimulus rivalry (Logothetis et al., 1996). Thus, we consider that the agent makes a composite action: one part is choosing from which of the eyes the percept will come, and, since it will then take some time to assimilate and process the information from that eye, the other part is a probabilistic description or belief about what the stimulus is, given the observations and the ocularity decisions. We do not introduce this probabilistic description formally (as it is not needed for the purposes of this simple model). However, it conceptually underpins certain aspects of the aesthetic reward. For instance, this probabilistic description of the content of the stimulus enriches and justifies the aesthetic reward, engenders an exploration bonus (since it may take some time to accumulate the relevant information), and also may provide the substrate for a suitably remunerated external decision.

One important facet of this perceptual decision-making process is that choosing that the percept is associated with the observation in one eye is to treat the components of the input from the other eye that do not match as noise. One formal possibility for this is to enrich the observation function of the generative model to allow for the possibility that the input of one eye can sometimes have excessive noise, such that

$$\Sigma = \Sigma^l = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix} \quad \text{or} \quad \Sigma = \Sigma^r = \sigma^2 \begin{pmatrix} \omega & 0 \\ 0 & 1 \end{pmatrix}, \tag{8}$$

where $\omega \gg 1$ represents excess noise. Then, we can treat the ocular component of the perceptual action as also determining the version of the covariance matrix used to determine the Bayesian loss function, in the equivalent of ***Equation 12***. This would have a rather similar effect to the way that Dayan and Solomon (2010) treated attention as suppressing noise, albeit in their case from confounded receptive fields, and with a different generative model.

Discussion

Our account of perceptual multistability which is grounded in value-based decision-making, provides a more comprehensive treatment of the computational and algorithmic facets of this phe-

nomenon. This decision-theoretic approach addresses several shortcomings of the more traditional, purely probabilistic, Helmholtzian accounts (Brascamp, Sterzer, et al., 2018). These shortcomings encompass both behavioural and neural aspects of perceptual multistablity. On the behavioural side (which was the main focus of this study), growing evidence suggests perceptual dominance and perceptual dynamics can be modulated by a wide range of cognitive, or at least not purely probabilistic, factors (Scocchia et al., 2014; Safavi and Dayan, 2022), for instance, reward (Balcetis et al., 2012; Wilbertz, van Slooten, et al., 2014; Marx et al., 2015; Wilbertz, van Kemenade, et al., 2017, also see the Results section). Neurally, several neural systems beyond sensory areas that are implicated in general aspects of decision-making (e.g., ACC, PFC) are also involved in perceptual multistability. Crucially, the computational side of our framework offers an explanation for the elusive behavioural aspects, and we suggest that its natural implementation in the brain will account for the rich collection of neural findings (see section Implication for the underlying neural substrate).

Implication for the underlying neural substrate

Although areas of the brain are far from singular in their computational roles, it remains true that if percepts arise as choices, then regions participating in the processes of decision-making should also be engaged in multistability. This is true for the ACC (Lyu et al., 2022; Drew et al., 2021; Gelbard-Sagiv et al., 2018), striatum (Baker et al., 2015; Sekutowicz et al., 2016), LC (at least using pupil size as proxy of its activation, Turi et al., 2018; Einhauser et al., 2008; Einhäuser et al., 2008; Naber et al., 2011; Schütz et al., 2018), PFC (see, e. g., Zaretskaya and Narinyan, 2014; Safavi, Kapoor, et al., 2014; Tsuchiya, Wilke, et al., 2015; Odegaard et al., 2017; Weilnhammer, Fritsch, et al., 2021b; Block, 2019; Michel, 2022; Kapoor et al., 2022; Dwarakanath et al., 2023) amygdala (Kreiman et al., 2002; Pasley et al., 2004), parietal cortex (Kanai et al., 2010; Zaretskaya, Thielscher, et al., 2010; Bahmani et al., 2019), pulvinar (Wilke et al., 2009; Qian et al., 2023) and various neuromodulators that were examined through pharmacology: serotonin (Carter et al., 2005), acetylcholine (Sheynin et al., 2020) catecholamine (Pfeffer et al., 2018).

We can speculate about how these various components are marshalled. For instance, the cascade leading to a perceptual switch might ultimately be triggered by the LC, noting early signs in pupil dilation (Einhauser et al., 2008) and the evidence that pharmacological manipulation of catecholamine systems increases the switching rate (Pfeffer et al., 2018). One computational correlate of this is the substantial uncertainty involved, particularly in binocular rivalry, as almost half the sensory input typically remains unexplained given either perceptual decision. Norepinephrine (NE), in concert with acetylcholine (ACh), has been suggested they reflect aspects of uncertainty, and occasioning the resetting of brain states over a large scale (Yu et al., 2005; Bouret et al., 2005). There are potential mechanistic substrates for aspects of their action (e. g., a modulation of excitation-inhibition balance by cholinergic and catecholamine systems, as suggestd by Sheynin et al., 2020; Pfeffer et al., 2018), although the exact circuits involved are as yet unclear.

Next, we suggest that the ACC (which exerts an important influence over the LC) may be involved in reevaluating perceptual actions, just as for other actions (Fouragnan et al., 2019; Brockett et al., 2020; Holroyd et al., 2021b; Akam et al., 2021). For instance, Fouragnan et al. (2019) combined fMRI, and transcranial ultrasound stimulation, in the context of a 3-arm restless bandit task to show that the values of options that are not currently available, but may potentially be chosen in the future, are represented in ACC. In their experiment, only 2 of the 3 bandits were available for choice, but the value of the third bandit still needed to be tracked. The authors showed that the ACC is critically involved in translating the value of such unavailable choices into a *counterfactual plan* that can be executed in the future. This has an obvious resonance with rivalry, with the key difference that, here, choices are a set of internal actions. Furthermore, the involvement of ACC in perceptual multistability is supported by a variety of experimental modalities (Lumer et al., 1998; Gelbard-Sagiv et al., 2018; Drew et al., 2021; Lyu et al., 2022). These data not only support the engagement of ACC in perceptual switches but also the processing of the suppressed percept (Levinson et al.,

2021), which would be necessary for computing the counterfactual plan.

Equally, we suggest that sub-regions of the PFC such as ventrolateral PFC (vlPFC) execute the perceptual actions, and thus influence the timing of perceptual switches. There is growing evidence of the involvement of various parts of the PFC in perceptual multistability (Zaretskaya and Narinyan, 2014; Safavi, Kapoor, et al., 2014; Tsuchiya, Wilke, et al., 2015; Odegaard et al., 2017; Block, 2019; Michel, 2022). A particular role has been suggested for vlPFC, a region involved in cognitive control, reflexive reorienting, motor inhibition, and action updating (Levy et al., 2011). Neural responses in this region reflect the value of visual objects (Ghazizadeh et al., 2021), are selective for sensory features in conventional visual paradigms (Pigarev et al., 1979; Ó Scalaidhe et al., 1997; Hussar et al., 2009; Safavi, Dwarakanath, et al., 2018), represent perceptual content during binocular rivalry (Kapoor et al., 2022), and exhibit characteristic network dynamics in the vicinity of perceptual transitions (Safavi, Logothetis, et al., 2021; Safavi, Panagiotaropoulos, et al., 2023; Dwarakanath et al., 2023).

Although there is substantial supporting evidence, other studies suggest perceptual switches per se do not heavily involve the fronto-parietal network (see, e. g., Brascamp, Blake, et al., 2015), which is believed to be intimately involved in the sort of cognitive control that we have suggested should be engaged in multistability (e. g., Zanto et al., 2013). Most of these studies shape their questions and experiments in the context of the recent debate on the role of the frontal lobe as a neural correlate of consciousness (NCC) and visual awareness. However, our focus here is mostly orthogonal, i. e., the role of value in determining dominance and switching. Even in that context, the involvement of the frontal-parietal network has been suggested to be intimately related to the *planning* of motor actions (without execution, Brascamp, Blake, et al., 2015) which can be considered as being compatible with our *goal-centered* approach to perceptual multistability (also inline with other proposals, see, e.g., Leopold et al., 1999; Jensen et al., 2023).

Finally, neuromodulators other than norepinephrine and acetylcholine, which play a critical role in shaping distributed information processing (Dayan, 2012), and are deeply involved in many psychiatric disorders (Iglesias et al., 2017), are also involved in perceptual multistablity (Carter et al., 2005; Sheynin et al., 2020; Pfeffer et al., 2018; Novicky et al., 2022), albeit in ways that are not completely clear. It is tempting to speculate that dopamine might be involved in at least some aspects of the complex temporal dynamics of the rate of switching (van Ee, 2005; Mamassian et al., 2005; Suzuki et al., 2007; Brascamp, Knapen, et al., 2008; Klink, Brascamp, et al., 2010). The basis for this is the suggestion that tonic levels of dopamine, by representing the local average reward rate, act as an opportunity cost for the passage of time, and thereby influence the vigour or alacrity of action (Niv et al., 2007; Mazzoni et al., 2007; Shadmehr et al., 2019). Intuitively, if the average rate of reward is high, every unit of time in which a reward is not delivered is costly for the agent, thus, it is worth the agent making decisions faster even if the costs of doing so (e. g., perceptual switches) are greater. The converse is true if the average rate of reward is low (Niv et al., 2007).

Limitations and future extensions

This is only the first formal realization of our theory, and so suffers from various limitations. First, we assume that there are only two plausible candidates for the state of the world. The space of candidates should, in principle, be considerably larger (in keeping with the potential richness of our perceptual experiences). Furthermore, in binocular rivalry, in addition to exclusive [dominant] percepts, humans (and perhaps also non-human primates) also experience, piecemeal, superposition (see, e. g., Klink, Brascamp, et al., 2010; Skerswetat et al., 2023) and fused percepts (for special stimuli like rivalling faces, Klink, Boucherie, et al., 2017, participants perceive morphed faces that slowly switch). In fact, the factors affecting piecemeal perception may shed light on the cognitive aspects of multistability. For instance, participants who are more open to exploration (based on measures of personality, Antinori et al., 2017), exhibit longer durations of piecemeal percepts. Thus, a pressing direction is to allow and investigate a richer state space.

Second, we assumed that the agent's observation function (***Equation 9***) has a known mean,

however, the mean of this distribution should also be estimated by the agent (for instance, using a similar procedure as Khalvati et al., 2021). The resulting decreasing exploration, which we characterized using a simplistic heuristic, might then contribute to the temporal dynamics of perception in early phases, and would interact with expectations about subsequent change.

Third, in our formulation, we consider the consequences of the agent's action to be both making a decision that the state is as implied by the associated percept, and also lowering observation noise about it (compared to the suppressed percept) which can be considered as a form of attention (Dayan and Solomon, 2010; Chebolu et al., 2022). The latter operation is computationally richer in hierarchical accounts of perception (Dayan, Hinton, et al., 1995; Dayan, 1998; Lee et al., 2003; Hohwy et al., 2008b). In such accounts, uncertainty about the unperceived image will be progressively greater at higher levels in the hierarchy, in a way that might influence the temporal dynamics. Given the explicit computational account of attention in our model, this could also provide an insight, if and how the underlying computational operation is related to the still controversial (Maier et al., 2021) question in multistability about the potential role of attention, a construct that itself is enduringly complex (Wu, 2023).

Fourth, our approximate POMDP solution only takes into account the immediate horizon; however, each action should really be taken based not only on its immediate return, but also on its long-term ramifications. The effects of this on the aesthetic value component of reward would be particularly complicated since this component is dynamic itself. However, some facets of perceptual dynamics that we explained based on heuristics (e.g., exploration bonus) might derive naturally through a more presbyopic solution of the POMDP.

Fifth, there are delicate aspects of the temporal dynamics of perception (such as the unit of time and variability of dominance periods across participants) that merit further investigation. For instance, the unit of time in the current setting of the model is a free parameter, so the model's temporal dynamics follow the experimental data (***Figure 6***), but only up to an arbitrary scaling factor. It would be interesting to relate the actual timescales in the model, such as the exponential decay and the learning rate of the system state, to empirical observations of switching behaviour. Furthermore, it is also interesting to investigate whether and how this timescale is related to the characteristic timescale of other cognitive functions, and, indeed, whether it is influenced by dopaminergic neuromodulation, given the latter's effect on other aspects of timing behaviour (Meck, 1983).

Finally, although the incorporation of aesthetic value (Brielmann and Dayan, 2022) was a key step towards a novel approach to sensory processing in multistability, the full potential of the aesthetic model was not explored in this study. Upon exposure of the agent to a new sensory object, the construction of a percept likely require some initial computation to stabilize the corresponding percept (an initial inference process). As a consequence, the value of the sensory object is not known to the agent. Given in our model, the entire inference process is boiled down to a simple Bayesian belief update (see Partially observable Markov decision process), the aesthetic model can be based on a *hierarchical* percept formation that requires some time to be realized. In the current formulation, for simplicity, we use a heuristic (***Equation 13***) to incorporate this initial computation. However, this can be incorporated more naturally into the model, for instance, as a richer hierarchical inference process in the POMDP (also related to the third point discussed above) or in the aesthetic model through the initial dynamics of the system state as a consequence of an initial inference process (also see our more elaborate description of the state, section Thick state). Moreover, the choice of parameters for the aesthetic component of the model was limited to a form of decay. This was motivated by the simple choices of visual images (e.g., gratings) that were used in a large body of previous experiments. However, the few studies that also looked at the temporal dynamics induced by more complex stimuli (like faces) reported non-monotonic temporal dynamics (see, e. g., Klink, Brascamp, et al., 2010; Klink, Boucherie, et al., 2017). The model of aesthetic reward can also have a nonmonotonic temporal structure, under more complex parameter settings (see, e. g., Brielmann and Dayan, 2022; Brielmann, Berentelg, et al., 2023). There is also room for further expanding the aesthetic model to better explain the temporal dynamic of multistablity.

For instance, to explain the long-term speed-up of switching (***Figure 6***), we introduced a heuristic (gradual change of switching cost across sessions). Such slow dynamics could be potentially explained through incorporating a slow learning process that affects the coordinate of the expected true distribution in the feature space of the aesthetic model (the expected true distribution was assumed to be constant in the original model, Brielmann and Dayan, 2022). Presumably, such slow learning should happen through offline processes, which also matches the slow dynamic in sessions distributed across days. Thus, it is pressing to investigate if and how more complex temporal dynamics of perception in binocular rivalry are shaped by the temporal dynamics of the aesthetic value.

Prospects

Perceptual multistability, which has been an object of study for many decades, arises in a wide range of species, across almost all sensory modalities, and is also relevant for a wide range of neural computations (Safavi and Dayan, 2022). It therefore has the potential to provide a unique window for understanding a wide range of cognitive functions and dysfunctions. In the following, we discuss some of the opportunities our novel perspective opens up.

First, having a handle on the temporal dynamics of multistability, may provide new understanding of distinctions between healthy and psychiatric populations (see, e. g., Ye et al., 2019; Jia et al., 2020), as most reported differences concern these facets. For instance, we speculate that, in particular forms of depression, value estimates will be biased lower, explaining the slower switching rate; and in anxiety, the informational value may have a higher weight (particularly for individually-scary stimuli), leading to elevation in the rate of perceptual switches (Jia et al., 2020). Equally, conditions in which dopamine release or reception differ (including forms of schizophrenia and depression; Maia et al., 2017; Pizzagalli et al., 2019) may exhibit different switching rates, given the putative tie to vigour and timing mentioned above (Niv et al., 2007; Meck, 1983). Altered rates of perceptual switches have also been reported with animal models of psychiatric conditions, including mouse models of autism (Palagina et al., 2017; Bogatova et al., 2023) (also in keeping with human studies; Robertson et al., 2013; Spiegel et al., 2019). The recent development of no-report paradigms (Tsuchiya, Wilke, et al., 2015), makes multistability an appealing paradigm for studies with animals and even children (Hudak et al., 2011; Beers et al., 2014; Karaminis et al., 2017).

Second, our formulation of perceptual multistability offers a functional and computational interpretation for some traditional biophysical characteristics such as adaptation. Besides the Helmholtzian approaches (Brascamp, Sterzer, et al., 2018), a large body of previous mathematical approaches (e.g., work on circuit and synaptic principles such as attractor dynamics, adaptation, and mutual inhibition; Braun et al., 2010; Deco et al., 2013; Theodoni, 2014) could account for many characteristics of perceptual multistability (e. g., distribution of dominance periods). However, they were largely agnostic to the sort of computational purposes on which we have focused. For instance, as we discussed earlier (see ***Figure 3*** and ***Figure 4***), we introduce adaptation in our framework in the cognitive context of boredom. Such formulations tie together biophysical and cognitive constructs (Scocchia et al., 2014; Safavi and Dayan, 2022). This is in keeping with other recent data – for instance, traditional visual adaptation (Keller et al., 2017) and pupil modulation by brightness (Turi et al., 2018) has been shown to be controlled by other cognitive processes. Future studies should address how such biophysical principles can be fully integrated with our normative computational framework.

Third, our approach suggests a link between modern conceptions of internal reinforcement learning (Hazy et al., 2007; Dayan, 2012; Lieder et al., 2018), sensory processing, and even neurofeedback (Lubianiker et al., 2022). Thus, it would be particularly important to examine cognitive control in this context (also given other suggestions, see, e. g., Botvinick et al., 2001; Drew et al., 2021). Indeed, the three components of cognitive control (monitoring, specification, and regulation) proposed by Shenhav et al. (2013), are also applicable to perceptual multistability. Monitoring concerns the assessment of the current circumstances and the evaluation of the performance of

the ongoing strategy. For perceptual multistability, this means (self-)assessing the quality of inference about the state of the world (which is the primary role of perception). Specification deals with the choice of actions or sub-tasks, that in the case of multistability are, for instance, the exclusive perception of one of the stimuli, or a perceptual switch. Finally, regulation concerns the adjustment of lower-level information processing machinery. In terms of perceptual multistability, this would be critical when elements of the task, or the statistics of the world, change. At a mechanistic level, adjustments could involve several processes, for instance, adaptation and tuning of the excitation-inhibition balance, possibly with computational implications (e.g., see the earlier discussion on adaptation).

Finally, since our proposed computational framework (and indeed the aesthetics model that we suggest endows percepts with value; Brielmann and Dayan, 2022) exploits Helmholtzian notions of sensation as inverse generation, our results are consistent with many others in that tradition (Brascamp, Sterzer, et al., 2018; Safavi and Dayan, 2022), and can readily be extended to incorporate the developments in these accounts. This would include, for instance, a more thoroughly hierarchical treatment (Dayan, Hinton, et al., 1995; Dayan, 1998; Lee et al., 2003).

In sum, in this study, we have introduced a decision-theoretic perspective on perceptual multistability, using binocular rivalry as an illustration of our computational framework. This framework not only explains several elusive aspects of the phenomenon (such as reward modulation), but also opens up modern conceptions of internal reinforcement learning in the service of understanding perceptual phenomena and sensory processing in health and disease.

Methods

Perceptual multistability as a POMDP

Partially observable Markov decision process

Humans and animals (generically, agents) frequently encounter situations in which they repeatedly need to choose one action among multiple alternatives. Each action in a sequence must be made not only based on its immediate effect in terms of gaining reward, but also based on any long-term ramifications from changing the circumstances or state of the agent. A Markov decision process (MDP, Sutton et al., 2018) is a general framework for modelling a wide range of such decision processes. The four components of an MDP are: states (specifying the circumstances of the human or animal in the external world); actions (the choice between available alternatives); transition probabilities (characterising how the state of the world might change given each possible action); and rewards, which specify the loss or gain associated with the state, action, and transition. Long-run values, a measure of the expected summed future reward, are a consequence of the agent's policy (i.e., the way it chooses to act at each state); and this policy can be optimised to maximise those long-term values.

Although MDPs have proved very useful for characterising many environments, one restrictive assumption is that the actor must know the state completely at all times (for instance, based on sensory input). This is not valid for many real-world problems, including our characterisation of perceptual multistability, where we assume that agents also make decisions about what to perceive and for how long (Martin et al., 2021; Safavi and Dayan, 2022). We therefore use a more complicated but more realistic class of models called *partially observable* MDPs or POMDPs (Littman, 2009). POMDPs are MDPs in which the agent may be uncertain about the state of the world, for instance, due to noisy sensory information. This noise is the product of an observation function that assigns probabilities to possible observations depending on the agent's states and actions. The agent then needs to form and update a belief about the state of the world in light of observations. It can also choose actions based on the information they are expected to provide about the state, i.e., on the value of the reduction in uncertainty about the state to which they are expected to lead. The agent's task remains to determine a policy that maximises long-term values. In the following, we describe the key components of the POMDP in relation to perceptual multistability, and leave

more generic aspects and procedures for the Appendix (see section Partially observable Markov decision process).

Sensory, perceptual and decision processes

As in conventional Bayesian accounts of perception, we assume that our decision-making agent has a generative model of the world (Hinton et al., 1997; Dayan, 1997; Rao et al., 1999; Doya et al., 2007; Yuille et al., 2006). which specifies what input activity would be generated given a particular state of the world. Perception is then the recognition task that inverts the generative model, i.e., mapping the activities in the sensory system to the percept (or strictly, a distribution over possible percepts). We consider there to be a single actual state $s_t \in \mathcal{S}$ of the world at any one time t, that must be inferred from sensory observations. However, rather than sticking to the regular probabilistic recognition model (inverse of the generative model, Dayan, 1997), we consider the fuller body of Bayesian decision theory for recognition. This is a key difference between this model, and previous models of perceptual multistability with the same spirit (Dayan, 1997; Gershman et al., 2014).

We build our account in two stages: first, assuming a very simple generative model containing only two possible states (see Thin state); and second, considering that the states have extra dimensions of internal structure (see Thick state) that need to be inferred over time from sensory input in a way that can benefit from the allocation of attention (which we treat itself as another component of the action space for the POMDP).

Thin state

For the simplest formulation of binocular rivalry, we assume there are only two plausible, but incompatible candidates for this state, $\mathcal{S} = \{+1, -1\}$ (Dayan, 1998). Thus, the agent should make an inference about the actual state at time t – this is the perceptual decision $a_t^P \in \mathcal{A} = \{+1, -1\}$. $\mathcal{S}$ and $\mathcal{A}$ are the state and perceptual action spaces of our perceptual POMDP; what remains is to specify the observations, transitions, and rewards.

In general, the observation function of a POMDP, $O = O(s_t, a_t^P, \mathbf{o}_t)$, expresses the probability distribution over possible observations $\mathbf{o}_t$ made at time t, given that the agent took action a_t^P and ended up in state s_t. In our case, the observations only depend on the current state s_t of the world, and not the perceptual action. We make the simplifying assumption that there is one scalar input for each eye, with Gaussian noise, so

$$O(s_t, a_{t-1}^P, \mathbf{o}_t) = P(\mathbf{o}_t | s_t) \sim \mathcal{N}\left[s_t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma \right] \tag{9}$$

The mean of this distribution, expresses the expectation that the eyes should see the same stimulus. The covariance matrix $\Sigma = \sigma^2 \mathcal{I}$ expresses the noise surrounding this stimulus; for the moment, we consider this to be proportional to the identity matrix ($\mathcal{I}$). Note that, in the case of binocular rivalry, a characteristic observation might express the conflict between the content of the two eyes (and the noise).

The transition function $T = T(s_{t-1}, a_t^P, s_t)$ of the POMDP expresses the probability distribution over the ending state, s_t, given that the agent started from state s_{t-1} and took action a_t^P. This accommodates the possibility of change in the environment, and so will be intimately tied to the dynamic nature of our perceptual experience (Pastukhov et al., 2022). Again, for our perceptual POMDP, the transition is independent of the perceptual action, leaving

$$T(s_{t-1}, s_t) = \begin{bmatrix} p_{\text{stay}} & 1 - p_{\text{stay}} \\ 1 - p_{\text{stay}} & p_{\text{stay}} \end{bmatrix}, \tag{10}$$

where p_{stay} denotes the probability of staying in the original state of the world, and we assume is known to the agent.

The reward function has a number of components: a loss function $\mathcal{L}(a_t^P, s_t)$ associated with a mismatch between the percept a_t^P and the state s_t as informed by the history of observations up to time t, $\mathcal{O}_t = \{\mathbf{o}_1, \ldots, \mathbf{o}_t\}$, a cost $c(a_t^P, a_{t-1}^P, t)$ of switching from percept a_{t-1}^P to percept a_t^P which can vary over a long time-scale, an aesthetic reward $R_A(a_t^P, t)$, an (heuristic) exploration bonus $R_{\text{exp}}(a_t^P, t)$, and an external reward $R_{\text{ext}}(a_t^P, t)$ associated with what the agent can achieve the world based on the percept (as in various of the experiments described above). We cover these in turn.

The standard Bayesian decision theory articulates the loss associated with choosing action a_t^P when the true state of the world is s_t. In a filtering context, one Helmholtzian possibility for this (slightly abusing the notation) is

$$\mathcal{L}(a_t^P, s_t) = -\log P(a_t^P \neq s_t | \mathcal{O}_t) \tag{11}$$

effectively counting the unlikelihood of the observations $\mathcal{O}_t$ given the current assignment of state a_t^P. A prominent component of this is the negative log probability of the current stimulus $\mathbf{o}_t$ under a_t^P

$$-\log P(\mathbf{o}_t | s_t = a_t^P) = \mathcal{K} + \frac{1}{2}(\mathbf{o}_t - a_t^P \mathbf{n})^T \Sigma^{-1} (\mathbf{o}_t - a_t^P \mathbf{n}) \tag{12}$$

where $\mathbf{n} = (1, 1)^T$ and $\mathcal{K}$ is a constant. Variants of this unlikelihood can explain perceptual preferences for higher (over lower) contrast inputs in rivalry, when the penalty for not explaining the high-contrast input is higher.

Switching between percepts (i.e., if $a_t^P \neq a_{t-1}^P$) is a form of costly *internal action* (Dayan, 2012) – the equivalent of the external action of moving the eye gaze from one object to another (e.g., Callaway et al., 2021). We allow the switching cost, $c(a_t^P, a_{t-1}^P, t)$, to change over a very long time-scale (i.e., experimental sessions conducted across days). That is, we assume that, within each experimental session (as structured in Suzuki et al., 2007) with multiple trials (each with length τ_{tr}, with an inter-trial interval of a few seconds), the switching cost remains constant. However, this constant switching cost decreases exponentially over multiple sessions (at least spread over days, as in Suzuki et al., 2007), for instance from an increase in the familiarity of the stimuli.

Balancing out these costs are three rewards. The aesthetic reward, $R_A(a_t^P, t)$, follows the model of Brielmann and Dayan (2022, as briefly described in section Model of aesthetic value in the Appendix). This theory suggests that aesthetic value is derived from optimising a sensory system to process present and future stimuli well, assuming, consistent with our Helmholtzian picture, that the agent has a *momentary or contextual* probabilistic model over the whole collection of stimuli it thinks it might sense (denoted as the system state) and a permanent probabilistic model for the sensory input it expects in the future (called the expected true distribution). Thus, the aesthetic reward is computed on the basis of current and past percepts and agent's generative models. At every moment, the agent computes and accrues the aesthetic reward from the perceived (dominant) stimulus, but also computes a counterfactual aesthetic reward for the suppressed percept. Both consumed and counterfactual aesthetic rewards are computed as described in section Model of aesthetic value in Appendix, with the key difference that the consumed one is computed based on the location of dominant stimulus in the feature space and the counterfactual one, based on the location of the suppressed stimulus.

We also consider a heuristic exploration bonus, $R_{\text{exp}}(a_t^P, t)$. This is an approximation to the value of information associated with a stimulus that is incompletely known (Dayan and Sejnowski, 1996). This utility is attributed to all possible states, and decreases during the time that state is being perceived. With the current, binary, state space, this term is notional, and we include it for completeness; when there is information content to a state, this stands in for the benefit of learning about this information. For the moment, we consider that, as the dominance time of a given state increases, the stimulus becomes better known, at least up to forgetting, and thus less novel. We define it as:

$$R_{\text{exp}}(a_t^P, t) = \omega_{\text{exp}} \exp(-T_a(t)/\tau_{\text{exp}}), \tag{13}$$

where ω_{exp} determines the gross contribution of this component, $T_a(t)$ is the integrated total dominance time of percept a, and τ_{exp} is the time constant according to which the bonus decreases. We assume the agent computes the integrated dominance $T_a(t)$ with a forgetting rate of η, given the limitation on the memory on past perceptual experience within a session and the possibility of change. Let $T_a^j(t)$ be the total dominance period of percept a in trial j up to time t; then we assume the cumulative dominance $T_a(t)$ (within the session) is the sum of the dominance period of the current trial and the discounted value from the previous trial (due to forgetting),

$$T_a(t) = T_a^j(t) + \eta T_a^{j-1}(\tau_{\text{tr}})\,. \tag{14}$$

where τ_{tr} is the length of a trial.

Finally, the external reward $R_{\text{ext}}(a_t^P, t)$ is, for our current purposes, a constant term (to be assigned to one particular percept) to account for rewards induced by the experimentalist such as associating reward to a stimulus or a percept (similar to the experimental design of previous studies, Balcetis et al., 2012; Wilbertz, van Slooten, et al., 2014; Marx et al., 2015; Wilbertz, van Kemenade, et al., 2017; Wilbertz and Sterzer, 2018) and/or rewards associated with an external, visually-guided task (as introduced in previous studies Paffen et al., 2008; Dieter et al., 2016; Moreno-Sánchez et al., 2019; Wang et al., 2021). Elements of this external reward (for instance, other remunerated decisions that can be made on the basis of the nature of the state) are rather nominal in the simplistic binary formulation. The full reward at time t is then

$$R(s_t, a_t^P, t) = -\mathcal{L}(a_t^P, s_t) - c(a_t^P, a_{t-1}^P, \tau^s) + R_A(a_t^P, t) + R_{\text{exp}}(a_t^P, t) + R_{\text{ext}}(a_t^P, t) \tag{15}$$

The solution of the POMDP is the optimal perceptual policy that maximizes the long-run summed expected reward (see Partially observable Markov decision process in the Appendix for more details). In the general formulation of a POMDP, both the immediate and the expected future reward (which here is dynamic, as not all the terms in the reward function are stationary) should be considered, but, to keep the model as simple as possible, we only consider the immediate reward, i.e., solving it myopically.

Thick state

Although the version of the model with a thin, binary, state is convenient to illustrate the workings of the model, it lacks the depth to encompass some key effects associated with the three forms of reward. In a more complete setting, we would have richer states $\mathbf{s}$, which have internal structure that can be inferred from the observations. Here, given the state $\mathbf{s}_t$, the mean of the observations would be $\mathbf{s}_t \otimes \mathbf{n}$. However, in rivalry, the actual input would be a noisy version of $\mathbf{s}^l$ to the left eye and $\mathbf{s}^r$ to the right eye (just as it was $+1$ to the left eye and -1 to the right eye in the thin state model). The agent has to make inferences about $\mathbf{s}_t$ given this conflicting input over time, and its expectations about change.

Here we ignore the possibility of mixed or averaged percepts (but see, Dayan, 1998), and only consider ocular rather than stimulus rivalry (Logothetis et al., 1996). Thus, we consider that the agent makes a composite action: one part is choosing from which of the eyes the percept will come, and, since it will then take some time to assimilate and process the information from that eye, the other part is a probabilistic description or belief about what the stimulus is, given the observations and the ocularity decisions. We do not introduce this probabilistic description formally (as it is not needed for the purposes of this simple model). However, it conceptually underpins certain aspects of the aesthetic reward. For instance, this probabilistic description of the content of the stimulus enriches and justifies the aesthetic reward, engenders an exploration bonus (since it may take some time to accumulate the relevant information), and also may provide the substrate for a suitably remunerated external decision.

One important facet of this perceptual decision-making process is that choosing that the percept is associated with the observation in one eye is to treat the components of the input from

the other eye that do not match as noise. One formal possibility for this is to enrich the observation function of the generative model to allow for the possibility that the input of one eye can sometimes have excessive noise, such that

$$\Sigma = \Sigma^l = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix} \quad \text{or} \quad \Sigma = \Sigma^r = \sigma^2 \begin{pmatrix} \omega & 0 \\ 0 & 1 \end{pmatrix}, \tag{16}$$

where $\omega \gg 1$ represents excess noise. Then, we can treat the ocular component of the perceptual action as also determining the version of the covariance matrix used to determine the Bayesian loss function, in the equivalent of ***Equation 12***. This would have a rather similar effect to the way that Dayan and Solomon (2010) treated attention as suppressing noise, albeit in their case from confounded receptive fields, and with a different generative model.

Acknowledgments

This work was supported by the Max Planck Society and the Alexander von Humboldt Foundation. SS acknowledges the add-on fellowship from the Joachim Herz Foundation. We would like to thank Satoru Suzuki for generously providing the data used presented in ***Figure 1***D-E (as well as, ***Figure 6***); Jan Brascamp for their valuable comments on an earlier version of this manuscript; scidraw.io for providing a free repository of high-quality scientific drawings (in particular Akshay Markanday for providing his "Monkey face cartoon" in this repository); and “macrovector_official” / Freepik for providing free vector graphics (icons and avatars) that were used in ***Figure 2***.

References

Akam, Thomas et al. (2021). “The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection”. In: *Neuron* 109.1, 149–163.e7. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2020.10.013. URL: https://www.cell.com/neuron/abstract/S0896-6273(20)30809-6 (visited on 2023) (cit. on p. 19).

Alais, David, Keys, Robert, et al. (2021). “Vestibular and Active Self-Motion Signals Drive Visual Perception in Binocular Rivalry”. In: *iScience* 24.12, p. 103417. ISSN: 2589-0042. DOI: 10.1016/j.isci.2021.103417. URL: https://www.sciencedirect.com/science/article/pii/S2589004221013882 (visited on 2023) (cit. on p. 1).

Alais, David and Melcher, David (2007). “Strength and Coherence of Binocular Rivalry Depends on Shared Stimulus Complexity”. In: *Vision Research* 47.2, pp. 269–279. ISSN: 0042-6989. DOI: 10.1016/j.visres.2006.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0042698906004159 (visited on 2021) (cit. on p. 2).

Antinori, Anna, Carter, Olivia L., and Smillie, Luke D. (2017). “Seeing It Both Ways: Openness to Experience and Binocular Rivalry Suppression”. In: *Journal of Research in Personality* 68, pp. 15–22. ISSN: 00926566. DOI: 10.1016/j.jrp.2017.03.005 (cit. on p. 20).

Bahmani, Hamed et al. (2019). “Responses of Neurons in Lateral Intraparietal Area Depend on Stimulus-Associated Reward During Binocular Flash Suppression”. In: *Frontiers in Systems Neuroscience* 13. ISSN: 1662-5137. DOI: 10.3389/fnsys.2019.00009. URL: https://www.frontiersin.org/articles/10.3389/fnsys.2019.00009/full (visited on 2020) (cit. on p. 19).

Baker, Daniel H. et al. (2015). “Brain Networks Underlying Bistable Perception”. In: *NeuroImage* 119, pp. 229–234. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2015.06.053. URL: https://www.sciencedirect.com/science/article/pii/S1053811915005583 (visited on 2021) (cit. on p. 19).

Balcetis, Emily, Dunning, David, and Granot, Yael (2012). “Subjective Value Determines Initial Dominance in Binocular Rivalry”. In: *Journal of Experimental Social Psychology* 48.1, pp. 122–129. ISSN: 0022-1031. DOI: 10.1016/j.jesp.2011.08.009. URL: http://www.sciencedirect.com/science/article/pii/S0022103111002228 (visited on 2021) (cit. on pp. 2, 7, 17, 19, 26).

Beers, Amanda et al. (2014). “Characterizing Perceptual Alternations during Binocular Rivalry in Children”. In: *Journal of Vision* 14.15, p. 47. ISSN: 1534-7362. DOI: 10.1167/14.15.47. URL: https://doi.org/10.1167/14.15.47 (visited on 2023) (cit. on p. 22).

Blake, R. and Logothetis, N. (2002). "Visual Competition". In: *Nat. Rev. Neurosci.* 3.1, pp. 13–21. ISSN: 1471-003X (Print) 1471-003X (Linking). DOI: 10.1038/nrn701 (cit. on pp. 1–3, 5, 13).

Block, Ned (2019). "What Is Wrong with the No-Report Paradigm and How to Fix It". In: *Trends in Cognitive Sciences* 23.12, pp. 1003–1013. ISSN: 1364-6613. DOI: 10.1016/j.tics.2019.10.001. URL: https://www.sciencedirect.com/science/article/pii/S1364661319302360 (visited on 2021) (cit. on pp. 3, 19, 20).

Bogatova, Daria, Smirnakis, Stelios M., and Palagina, Ganna (2023). *Tug-of-Peace: Visual Rivalry and Atypical Visual Motion Processing in MECP2 Duplication Syndrome of Autism*. DOI: 10.1101/2022.09.23.509144. URL: https://www.biorxiv.org/content/10.1101/2022.09.23.509144v2 (visited on 2023) (cit. on p. 22).

Botvinick, Matthew M. et al. (2001). "Conflict Monitoring and Cognitive Control". In: *Psychological Review* 108.3, pp. 624–652. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/0033-295X.108.3.624 (cit. on p. 22).

Bouret, Sebastien and Sara, Susan J. (2005). "Network Reset: A Simplified Overarching Theory of Locus Coeruleus Noradrenaline Function". In: *Trends in Neurosciences* 28.11, pp. 574–582. ISSN: 0166-2236. DOI: 10.1016/j.tins.2005.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0166223605002432 (visited on 2023) (cit. on p. 19).

Brascamp, J. W., Klink, P. C., and Levelt, W. J. (2015). "The 'laws' of Binocular Rivalry: 50 Years of Levelt's Propositions". In: *Vision research* 109, pp. 20–37. ISSN: 1878-5646 (Electronic) 0042-6989 (Linking). DOI: 10.1016/j.visres.2015.02.019 (cit. on pp. 1, 9, 12, 13).

Brascamp, J. W., van Ee, R., et al. (2006). "The Time Course of Binocular Rivalry Reveals a Fundamental Role of Noise". In: *J Vis* 6.11, pp. 1244–56. ISSN: 1534-7362 (Electronic) 1534-7362 (Linking). DOI: 10.1167/6.11.8 (cit. on p. 8).

Brascamp, Jan, Blake, Randolph, and Knapen, Tomas (2015). "Negligible Fronto-Parietal BOLD Activity Accompanying Unreportable Switches in Bistable Perception". In: *Nat Neurosci*. ISSN: 1097-6256. DOI: doi:10.1038/nn.4130 (cit. on p. 20).

Brascamp, Jan, Sterzer, Philipp, et al. (2018). "Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference". In: *Annual review of psychology* 69, pp. 77–103 (cit. on pp. 1–3, 9, 19, 22, 23, 40).

Brascamp, Jan W., Ee, Raymond van, et al. (2005). "Distributions of Alternation Rates in Various Forms of Bistable Perception". In: *Journal of Vision* 5.4, pp. 1–1. ISSN: 1534-7362. DOI: 10.1167/5.4.1. URL: https://jov.arvojournals.org/article.aspx?articleid=2192715 (visited on 2021) (cit. on p. 3).

Brascamp, Jan W., Knapen, Tomas H. J., et al. (2008). "Multi-Timescale Perceptual History Resolves Visual Ambiguity". In: *PLOS ONE* 3.1, e1497. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0001497. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0001497 (visited on 2023) (cit. on p. 20).

Braun, J. and Mattia, M. (2010). "Attractors and Noise: Twin Drivers of Decisions and Multistability". In: *Neuroimage* 52, pp. 740–751. ISSN: 1053-8119. DOI: DOI10.1016/j.neuroimage.2009.12.126 (cit. on p. 22).

Brielmann, Aenne A., Berentelg, Max, and Dayan, Peter (2023). "Modelling Individual Aesthetic Judgements over Time". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379.1895, p. 20220414. DOI: 10.1098/rstb.2022.0414. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2022.0414 (visited on 2024) (cit. on pp. 5, 21).

Brielmann, Aenne A. and Dayan, Peter (2022). "A Computational Model of Aesthetic Value". In: *Psychological Review*, No Pagination Specified–No Pagination Specified. ISSN: 1939-1471. DOI: 10.1037/rev0000337 (cit. on pp. 2, 5, 7, 17, 21–23, 25, 40).

Brockett, Adam T. et al. (2020). "Anterior Cingulate Cortex Is Necessary for Adaptation of Action Plans". In: *Proceedings of the National Academy of Sciences* 117.11, pp. 6196–6204. DOI: 10.1073/pnas.1919303117. URL: https://www.pnas.org/doi/full/10.1073/pnas.1919303117 (visited on 2023) (cit. on p. 19).

Brown, Kenneth T. (1955). "Rate of Apparent Change in a Dynamic Ambiguous Figure as a Function of Observation-Time". In: *The American Journal of Psychology* 68.3, pp. 358–371. ISSN: 0002-9556. DOI: 10.2307/1418520. JSTOR: 1418520. URL: https://www.jstor.org/stable/1418520 (visited on 2022) (cit. on p. 3).

Callaway, Frederick, Rangel, Antonio, and Griffiths, Thomas L. (2021). "Fixation Patterns in Simple Choice Reflect Optimal Information Sampling". In: *PLOS Computational Biology* 17.3, e1008863. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008863. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008863 (visited on 2021) (cit. on pp. 17, 25).

Cao, Robin et al. (2021). "Binocular Rivalry Reveals an Out-of-Equilibrium Neural Dynamics Suited for Decision-Making". In: *eLife* 10. Ed. by Rinzel et al., e61581. ISSN: 2050-084X. DOI: 10.7554/eLife.61581. URL: https://doi.org/10.7554/eLife.61581 (visited on 2021) (cit. on pp. 2, 3, 5).

Carmel, David et al. (2010). "How to Create and Use Binocular Rivalry". In: *JoVE (Journal of Visualized Experiments)* 45, e2030. ISSN: 1940-087X. DOI: 10.3791/2030. URL: https://www.jove.com/de/v/2030/how-to-create-and-use-binocular-rivalry (visited on 2021) (cit. on p. 40).

Carter, Olivia et al. (2020). "Perceptual Rivalry across Animal Species". In: *Journal of Comparative Neurology* 528.17, pp. 3123–3133 (cit. on p. 1).

Carter, Olivia L et al. (2005). "Modulating the Rate and Rhythmicity of Perceptual Rivalry Alternations with the Mixed 5-HT2A and 5-HT1A Agonist Psilocybin". In: *Neuropsychopharmacology* 30.6, pp. 1154–1162. ISSN: 0893-133X, 1740-634X. DOI: 10.1038/sj.npp.1300621. URL: http://www.nature.com/articles/1300621 (visited on 2020) (cit. on pp. 19, 20).

Chebolu, Sahiti, Dayan, Peter, and Lloyd, Kevin (2022). "Vigilance, Arousal, and Acetylcholine: Optimal Control of Attention in a Simple Detection Task". In: *PLOS Computational Biology* 18.10, e1010642. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010642. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010642 (visited on 2023) (cit. on p. 21).

Chittka, Lars and Walker, Julian (2006). "Do Bees like Van Gogh's Sunflowers?" In: *Optics & Laser Technology*. Colour and Design in the Natural and Man-Made Worlds 38.4, pp. 323–328. ISSN: 0030-3992. DOI: 10.1016/j.optlastec.2005.06.020. URL: https://www.sciencedirect.com/science/article/pii/S0030399205001258 (visited on 2023) (cit. on p. 2).

Cloutier, Jasmin et al. (2008). "Are Attractive People Rewarding? Sex Differences in the Neural Substrates of Facial Attractiveness". In: *Journal of Cognitive Neuroscience* 20.6, pp. 941–951. ISSN: 0898-929X. DOI: 10.1162/jocn.2008.20062. URL: https://doi.org/10.1162/jocn.2008.20062 (visited on 2021) (cit. on p. 3).

Daini, Roberta et al. (2010). "Neuropsychological Evidence of High-Level Processing in Binocular Rivalry". In: *Behavioural Neurology* 23.4, pp. 233–235. ISSN: 0953-4180, 1875-8584. DOI: 10.1155/2010/148485. URL: http://www.hindawi.com/journals/bn/2010/148485/ (visited on 2021) (cit. on p. 2).

Dayan, P. (2012). "Twenty-Five Lessons from Computational Neuromodulation". In: *Neuron* 76, pp. 240–56. ISSN: 1097-4199 (Electronic) 0896-6273 (Linking). DOI: 10.1016/j.neuron.2012.09.027 (cit. on p. 20).

Dayan, Peter (1997). "Recognition in Hierarchical Models". In: *Foundations of Computational Mathematics*. Ed. by Cucker and Shub. Berlin, Heidelberg: Springer, pp. 43–62. ISBN: 978-3-642-60539-0. DOI: 10.1007/978-3-642-60539-0_4 (cit. on pp. 14, 16, 24).

— (1998). "A Hierarchical Model of Binocular Rivalry". In: *Neural Comput* 10.5, pp. 1119–1135. ISSN: 0899-7667 1530-888X. DOI: 10.1162/089976698300017377. URL: https://direct.mit.edu/neco/article-abstract/10/5/1119/6171/A-Hierarchical-Model-of-Binocular-Rivalry (cit. on pp. 2, 3, 5, 9, 16, 18, 21, 23, 24, 26).

— (2012). "How to Set the Switches on This Thing". In: *Current Opinion in Neurobiology*. Decision Making 22.6, pp. 1068–1074. ISSN: 0959-4388. DOI: 10.1016/j.conb.2012.05.011. URL: https://www.sciencedirect.com/science/article/pii/S0959438812000992 (visited on 2021) (cit. on pp. 2, 3, 5, 17, 22, 25).

Dayan, Peter, Hinton, Geoffrey E, et al. (1995). "The helmholtz machine". In: *Neural computation* 7.5, pp. 889–904 (cit. on pp. 21, 23, 40).

Dayan, Peter and Sejnowski, Terrence J (1996). "Exploration bonuses and dual control". In: *Machine Learning* 25, pp. 5–22 (cit. on pp. 5, 13, 17, 25).

Dayan, Peter and Solomon, Joshua A. (2010). "Selective Bayes: Attentional Load and Crowding". In: *Vision Research*. Mathematical Models of Visual Coding 50.22, pp. 2248–2260. ISSN: 0042-6989. DOI: 10.1016/j.visres.2010.04.014. URL: https://www.sciencedirect.com/science/article/pii/S0042698910001902 (visited on 2023) (cit. on pp. 18, 21, 27).

Deco, G. et al. (2013). "Brain Mechanisms for Perceptual and Reward-Related Decision-Making". In: *Prog Neurobiol* 103, pp. 194–213. ISSN: 1873-5118 (Electronic) 0301-0082 (Linking). DOI: 10.1016/j.pneurobio.2012.01.010 (cit. on p. 22).

Dieter, Kevin C., Melnick, Michael D., and Tadin, Duje (2016). "Perceptual Training Profoundly Alters Binocular Rivalry through Both Sensory and Attentional Enhancements". In: *Proceedings of the National Academy of Sciences* 113.45, pp. 12874–12879. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1602722113. URL: https://www.pnas.org/content/113/45/12874 (visited on 2021) (cit. on pp. 18, 26).

Doya, Kenji et al. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press. ISBN: 978-0-262-04238-3 (cit. on pp. 14, 24).

Drew, Alice et al. (2021). "Conflict Monitoring and Attentional Adjustment during Binocular Rivalry". In: *European Journal of Neuroscience* n/a.n/a. ISSN: 1460-9568. DOI: 10.1111/ejn.15554. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.15554 (visited on 2021) (cit. on pp. 2, 3, 19, 22).

Dwarakanath, Abhilash et al. (2023). "Bistability of prefrontal states gates access to consciousness". In: *Neuron* 111.10, pp. 1666–1683. DOI: 10.1016/j.neuron.2023.02.027 (cit. on pp. 19, 20).

Einhauser, W. et al. (2008). "Pupil Dilation Reflects Perceptual Selection and Predicts Subsequent Stability in Perceptual Rivalry". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.5, pp. 1704–9. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.0707727105 (cit. on pp. 3, 19).

Einhäuser, Wolfgang et al. (2008). "Reply to Hupé et al.: The Predictive Correlation of Pupil Dilation and Relative Dominance Durations in Rivalry Is Not a Statistical Artifact". In: *Proceedings of the National Academy of Sciences* 105.28, E44–E44. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0803927105. URL: https://www.pnas.org/content/105/28/E44 (visited on 2021) (cit. on p. 19).

Fouragnan, Elsa F. et al. (2019). "The Macaque Anterior Cingulate Cortex Translates Counterfactual Choice Value into Actual Behavioral Change". In: *Nature Neuroscience* 22.5, pp. 797–808. ISSN: 1546-1726. DOI: 10.1038/s41593-019-0375-6. URL: https://www.nature.com/articles/s41593-019-0375-6 (visited on 2023) (cit. on p. 19).

Gelbard-Sagiv, Hagar et al. (2018). "Human Single Neuron Activity Precedes Emergence of Conscious Perception". In: *Nat. Commun.* 9.1, p. 2057. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03749-0 (cit. on pp. 2, 3, 19).

Gershman, Samuel, Vul, Ed, and Tenenbaum, Joshua B. (2014). "Perceptual Multistability as Markov Chain Monte Carlo Inference". In: *Advances in Neural Information Processing Systems*, pp. 611–619 (cit. on pp. 2, 3, 5, 16, 24).

Ghazizadeh, Ali and Hikosaka, Okihide (2021). "Common Coding of Expected Value and Value Uncertainty Memories in the Prefrontal Cortex and Basal Ganglia Output". In: *Science Advances* 7.20, eabe0693. ISSN: 2375-2548. DOI: 10.1126/sciadv.abe0693. URL: https://advances.sciencemag.org/content/7/20/eabe0693 (visited on 2021) (cit. on p. 20).

Hazy, Thomas E, Frank, Michael J, and O'reilly, Randall C (2007). "Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1485, pp. 1601–1613 (cit. on pp. 3, 22).

Hinton, Geoffrey E. and Ghahramani, Zoubin (1997). "Generative Models for Discovering Sparse Distributed Representations". In: *Philosophical Transactions of the Royal Society of London. Series*

B: Biological Sciences 352.1358, pp. 1177–1190. DOI: 10.1098/rstb.1997.0101. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1997.0101 (visited on 2023) (cit. on pp. 14, 24).

Hohwy, J., Roepstorff, A., and Friston, K. (2008a). "Predictive Coding Explains Binocular Rivalry: An Epistemological Review". In: *Cognition* 108.3, pp. 687–701. ISSN: 0010-0277 (Print) 0010-0277 (Linking). DOI: 10.1016/j.cognition.2008.05.010 (cit. on pp. 2, 3, 5).

— (2008b). "Predictive Coding Explains Binocular Rivalry: An Epistemological Review". In: *Cognition* 108.3, pp. 687–701. ISSN: 0010-0277 (Print) 0010-0277 (Linking). DOI: 10.1016/j.cognition.2008.05.010 (cit. on p. 21).

Holroyd, Clay B. and Verguts, Tom (2021a). "The Best Laid Plans: Computational Principles of Anterior Cingulate Cortex". In: *Trends in Cognitive Sciences* 25.4, pp. 316–329. ISSN: 1364-6613. DOI: 10.1016/j.tics.2021.01.008. URL: https://www.sciencedirect.com/science/article/pii/S1364661321000103 (visited on 2021) (cit. on p. 3).

— (2021b). "The Best Laid Plans: Computational Principles of Anterior Cingulate Cortex". In: *Trends in Cognitive Sciences* 25.4, pp. 316–329. ISSN: 1364-6613. DOI: 10.1016/j.tics.2021.01.008. URL: https://www.sciencedirect.com/science/article/pii/S1364661321000103 (visited on 2023) (cit. on p. 19).

Huang, Yu-Feng and Hsieh, Po-Jang (2013). "The Mere Exposure Effect Is Modulated by Selective Attention but Not Visual Awareness". In: *Vision Research* 91, pp. 56–61. ISSN: 0042-6989. DOI: 10.1016/j.visres.2013.07.017 (cit. on pp. 2, 9).

Hudak, Mariann et al. (2011). "Increased Readiness for Adaptation and Faster Alternation Rates Under Binocular Rivalry in Children". In: *Frontiers in Human Neuroscience* 5. ISSN: 1662-5161. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2011.00128 (visited on 2023) (cit. on p. 22).

Hussar, C. R. and Pasternak, T. (2009). "Flexibility of Sensory Representations in Prefrontal Cortex Depends on Cell Type". In: *Neuron* 64, pp. 730–43. ISSN: 1097-4199 (Electronic) 0896-6273 (Linking). DOI: 10.1016/j.neuron.2009.11.018 (cit. on p. 20).

Iglesias, Sandra et al. (2017). "Models of Neuromodulation for Computational Psychiatry". In: *WIREs Cognitive Science* 8.3, e1420. ISSN: 1939-5086. DOI: 10.1002/wcs.1420. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1420 (visited on 2024) (cit. on p. 20).

Iigaya, Kiyohito, Yi, Sanghyun, Wahle, Iman A., Tanwisuth, Koranis, et al. (2021). "Aesthetic Preference for Art Can Be Predicted from a Mixture of Low- and High-Level Visual Features". In: *Nature Human Behaviour* 5.6, pp. 743–755. ISSN: 2397-3374. DOI: 10.1038/s41562-021-01124-6. URL: https://www.nature.com/articles/s41562-021-01124-6 (visited on 2022) (cit. on p. 2).

Iigaya, Kiyohito, Yi, Sanghyun, Wahle, Iman A., Tanwisuth, Sandy, et al. (2023). "Neural Mechanisms Underlying the Hierarchical Construction of Perceived Aesthetic Value". In: *Nature Communications* 14.1, p. 127. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35654-y. URL: https://www.nature.com/articles/s41467-022-35654-y (visited on 2023) (cit. on p. 3).

Jensen, Ole and Ferrante, Oscar (2023). "Prefrontal Cortex Oscillations Update and Stabilize Consciousness during Binocular Rivalry". In: *Neuron* 111.10, pp. 1519–1520. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2023.04.027. URL: https://www.sciencedirect.com/science/article/pii/S0896627323003379 (visited on 2023) (cit. on p. 20).

Jia, Ting et al. (2020). "Difference in Binocular Rivalry Rate between Major Depressive Disorder and Generalized Anxiety Disorder". In: *Behavioural Brain Research* 391, p. 112704. ISSN: 0166-4328. DOI: 10.1016/j.bbr.2020.112704. URL: http://www.sciencedirect.com/science/article/pii/S0166432820304034 (visited on 2020) (cit. on pp. 2, 22).

Kaelbling, Leslie Pack, Littman, Michael L., and Cassandra, Anthony R. (1998). "Planning and Acting in Partially Observable Stochastic Domains". In: *Artif. Intell* 101.1, pp. 99–134. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(98)00023-X (cit. on p. 41).

Kanai, R., Bahrami, B., and Rees, G. (2010). "Human Parietal Cortex Structure Predicts Individual Differences in Perceptual Rivalry". In: *Curr Biol* 20.18, pp. 1626–30. ISSN: 1879-0445 (Electronic) 0960-9822 (Linking). DOI: 10.1016/j.cub.2010.07.027 (cit. on p. 19).

Kapoor, Vishal et al. (2022). "Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports". In: *Nature Communications* 13.1, p. 1535. DOI: 10.1038/s41467-022-28897-2 (cit. on pp. 4, 19, 20).

Karaminis, Themelis et al. (2017). "Binocular Rivalry in Children on the Autism Spectrum". In: *Autism Res.* 10.6, pp. 1096–1106. ISSN: 1939-3806. (Visited on 2023) (cit. on p. 22).

Keller, Andreas J et al. (2017). "Stimulus Relevance Modulates Contrast Adaptation in Visual Cortex". In: *eLife* 6. Ed. by Kleinfeld, e21589. ISSN: 2050-084X. DOI: 10.7554/eLife.21589. URL: https://doi.org/10.7554/eLife.21589 (visited on 2022) (cit. on p. 22).

Khalvati, Koosha, Kiani, Roozbeh, and Rao, Rajesh P. N. (2021). "Bayesian Inference with Incomplete Knowledge Explains Perceptual Confidence and Its Deviations from Accuracy". In: *Nature Communications* 12.1, p. 5704. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25419-4. URL: https://www.nature.com/articles/s41467-021-25419-4 (visited on 2022) (cit. on p. 21).

Klink, P. Christiaan, Boucherie, Daphne, et al. (2017). "Interocularly Merged Face Percepts Eliminate Binocular Rivalry". In: *Sci. Rep.* 7.1, p. 7585. ISSN: 2045-2322. DOI: 10.1038/s41598-017-08023-9 (cit. on pp. 20, 21, 40).

Klink, P. Christiaan, Brascamp, Jan W., et al. (2010). "Experience-Driven Plasticity in Binocular Vision". In: *Current Biology* 20.16, pp. 1464–1469. ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.06.057 (cit. on pp. 20, 21).

Kreiman, Gabriel, Fried, Itzhak, and Koch, Christof (2002). "Single-Neuron Correlates of Subjective Vision in the Human Medial Temporal Lobe". In: *Proceedings of the National Academy of Sciences* 99.12, pp. 8378–8383. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.072194099. URL: https://www.pnas.org/content/99/12/8378 (visited on 2021) (cit. on p. 19).

Krug, Kristine (2020). "Coding Perceptual Decisions: From Single Units to Emergent Signaling Properties in Cortical Circuits". In: *Annual Review of Vision Science* 6.1, pp. 387–409. DOI: 10.1146/annurev-vision-030320-041223. URL: https://doi.org/10.1146/annurev-vision-030320-041223 (visited on 2022) (cit. on p. 1).

Lee, Daeyeol, Seo, Hyojung, and Jung, Min Whan (2012). "Neural Basis of Reinforcement Learning and Decision Making". In: *Annual review of neuroscience* 35, pp. 287–308. ISSN: 0147-006X. DOI: 10.1146/annurev-neuro-062111-150512. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490621/ (visited on 2021) (cit. on p. 3).

Lee, T. S. and Mumford, D. (2003). "Hierarchical Bayesian Inference in the Visual Cortex". In: *J Opt Soc Am A Opt Image Sci Vis* 20.7, pp. 1434–48. ISSN: 1084-7529 (Print) 1084-7529 (Linking). DOI: Doi10.1364/Josaa.20.001434 (cit. on pp. 21, 23).

Leopold, D. A. and Logothetis, N. K. (1999). "Multistable Phenomena: Changing Views in Perception". In: *Trends Cogn Sci* 3.7, pp. 254–264. ISSN: 1879-307X (Electronic) 1364-6613 (Linking). DOI: 10.1016/S1364-6613(99)01332-7 (cit. on pp. 1, 20).

Leptourgos, Pantelis et al. (2020). "A Functional Theory of Bistable Perception Based on Dynamical Circular Inference". In: *PLOS Computational Biology* 16.12, e1008480. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008480. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008480 (visited on 2020) (cit. on pp. 2, 3, 5, 12).

Levelt, Willem JM (1965). "On Binocular Rivalry". PhD thesis. Van Gorcum Assen (cit. on pp. 5, 9).

Levinson, Max et al. (2021). "Cortical and Subcortical Signatures of Conscious Object Recognition". In: *Nature Communications* 12.1, p. 2930. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23266-x. URL: https://www.nature.com/articles/s41467-021-23266-x (visited on 2021) (cit. on p. 19).

Levy, Benjamin J. and Wagner, Anthony D. (2011). "Cognitive Control and Right Ventrolateral Prefrontal Cortex: Reflexive Reorienting, Motor Inhibition, and Action Updating". In: *Annals of the New York Academy of Sciences* 1224.1, pp. 40–62. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.2011.05958.x. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079823/ (visited on 2022) (cit. on p. 20).

Lieder, Falk et al. (2018). "Rational metareasoning and the plasticity of cognitive control". In: *PLoS computational biology* 14.4, e1006043 (cit. on pp. 3, 22).

Littman, Michael L. (2009). "A Tutorial on Partially Observable Markov Decision Processes". In: *Journal of Mathematical Psychology*. Special Issue: Dynamic Decision Making 53.3, pp. 119–125. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2009.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0022249609000042 (visited on 2021) (cit. on pp. 3, 14, 23).

Logothetis, N. K., Leopold, D. A., and Sheinberg, D. L. (1996). "What Is Rivalling during Binocular Rivalry?" In: *Nature* 380.6575, pp. 621–4. ISSN: 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/380621a0 (cit. on pp. 18, 26).

Lubianiker, Nitzan et al. (2022). "Neurofeedback through the Lens of Reinforcement Learning". In: *Trends in Neurosciences* 45.8, pp. 579–593. ISSN: 0166-2236. DOI: 10.1016/j.tins.2022.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0166223622000595 (visited on 2024) (cit. on p. 22).

Lumer, E. D., Friston, K. J., and Rees, G. (1998). "Neural Correlates of Perceptual Rivalry in the Human Brain". In: *Science* 280.5371, pp. 1930–4. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.280.5371.1930 (cit. on pp. 3, 19).

Lunghi, Claudia and Pooresmaeili, Arezoo (2023). "Learned Value Modulates the Access to Visual Awareness during Continuous Flash Suppression". In: *Scientific Reports* 13.1, p. 756. ISSN: 2045-2322. DOI: 10.1038/s41598-023-28004-5. URL: https://www.nature.com/articles/s41598-023-28004-5 (visited on 2023) (cit. on p. 8).

Lyu, Dian et al. (2022). "Intrinsic Brain Dynamics in the Default Mode Network Predict Involuntary Fluctuations of Visual Awareness". In: *Nature Communications* 13.1, p. 6923. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34410-6. URL: https://www.nature.com/articles/s41467-022-34410-6 (visited on 2022) (cit. on pp. 2, 3, 19).

Maia, Tiago V. and Frank, Michael J. (2017). "An Integrative Perspective on the Role of Dopamine in Schizophrenia". In: *Biological Psychiatry*. Dopamine Hypothesis of Schizophrenia 81.1, pp. 52–66. ISSN: 0006-3223. DOI: 10.1016/j.biopsych.2016.05.021. URL: http://www.sciencedirect.com/science/article/pii/S0006322316324258 (visited on 2020) (cit. on p. 22).

Maier, Alexander and Tsuchiya, Naotsugu (2021). "Growing Evidence for Separate Neural Mechanisms for Attention and Consciousness". In: *Attention, Perception, & Psychophysics* 83.2, pp. 558–576. ISSN: 1943-393X. DOI: 10.3758/s13414-020-02146-4. URL: https://doi.org/10.3758/s13414-020-02146-4 (visited on 2021) (cit. on pp. 1, 21).

Mamassian, Pascal and Goutcher, Ross (2005). "Temporal Dynamics in Bistable Perception". In: *Journal of Vision* 5.4, p. 7. ISSN: 1534-7362. DOI: 10.1167/5.4.7. URL: https://doi.org/10.1167/5.4.7 (visited on 2022) (cit. on p. 20).

Martin, Joshua M., Solms, Mark, and Sterzer, Philipp (2021). "Useful Misrepresentation: Perception as Embodied Proactive Inference". In: *Trends in Neurosciences* 44.8, pp. 619–628. ISSN: 0166-2236. DOI: 10.1016/j.tins.2021.04.007. URL: https://www.sciencedirect.com/science/article/pii/S0166223621000795 (visited on 2021) (cit. on pp. 2, 14, 23).

Marx, Svenja and Einhäuser, Wolfgang (2015). "Reward Modulates Perception in Binocular Rivalry". In: *Journal of Vision* 15.1, p. 15.1.11. ISSN: 1534-7362. DOI: 10.1167/15.1.11 (cit. on pp. 2, 7, 9, 17, 19, 26).

Mazzoni, Pietro, Hristova, Anna, and Krakauer, John W. (2007). "Why Don't We Move Faster? Parkinson's Disease, Movement Vigor, and Implicit Motivation". In: *Journal of Neuroscience* 27.27, pp. 7105–7116. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0264-07.2007. URL: https://www.jneurosci.org/content/27/27/7105 (visited on 2024) (cit. on p. 20).

Meck, Warren H (1983). "Selective adjustment of the speed of internal clock and memory processes." In: *Journal of Experimental Psychology: Animal Behavior Processes* 9.2, p. 171 (cit. on pp. 21, 22).

Michel, Matthias (2022). "Conscious Perception and the Prefrontal Cortex A Review". In: *Journal of Consciousness Studies* 29.7-8, pp. 115–157. DOI: 10.53765/20512201.29.7.115 (cit. on pp. 2, 3, 19, 20).

Mo, Ce et al. (2016). "Natural Tendency towards Beauty in Humans: Evidence from Binocular Rivalry". In: *PLOS ONE* 11.3, e0150147. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0150147. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150147 (visited on 2021) (cit. on pp. 2, 9).

Monosov, Ilya E. et al. (2020). "Anterior Cingulate Cortex and the Control of Dynamic Behavior in Primates". In: *Current Biology* 30.23, R1442–R1454. ISSN: 0960-9822. DOI: 10.1016/j.cub.2020.10.009. URL: https://www.sciencedirect.com/science/article/pii/S096098222031513X (visited on 2021) (cit. on p. 3).

Montague, P. Read and Berns, Gregory S. (2002). "Neural Economics and the Biological Substrates of Valuation". In: *Neuron* 36.2, pp. 265–284. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(02)00974-1. URL: https://www.cell.com/neuron/abstract/S0896-6273(02)00974-1 (visited on 2021) (cit. on p. 3).

Moreno-Bote, R., Knill, D. C., and Pouget, A. (2011). "Bayesian Sampling in Visual Perception". In: *Proc Natl Acad Sci U S A* 108, pp. 12491–6. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1101430108 (cit. on pp. 2, 3, 5).

Moreno-Bote, Rubén et al. (2010). "Alternation Rate in Perceptual Bistability Is Maximal at and Symmetric around Equi-Dominance". In: *Journal of Vision* 10.11, pp. 1–1. ISSN: 1534-7362. DOI: 10.1167/10.11.1. URL: http://jov.arvojournals.org/article.aspx?articleid=2191530 (visited on 2020) (cit. on p. 9).

Moreno-Sánchez, Manuel, Aznar-Casanova, J. Antonio, and Valle-Inclán, Fernando (2019). "Attention to Monocular Images Bias Binocular Rivalry". In: *Frontiers in Systems Neuroscience* 13, p. 12. ISSN: 1662-5137. DOI: 10.3389/fnsys.2019.00012. URL: https://www.frontiersin.org/article/10.3389/fnsys.2019.00012 (visited on 2021) (cit. on pp. 18, 26).

Naber, M., Frassle, S., and Einhauser, W. (2011). "Perceptual Rivalry: Reflexes Reveal the Gradual Nature of Visual Awareness". In: *PLoS One* 6.6, e20910. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0020910 (cit. on p. 19).

Nath, Surabhi S. et al. (2024). "Relating Objective Complexity, Subjective Complexity and Beauty in Binary Pixel Patterns". In: DOI: 10.31234/osf.io/nuep7. URL: https://osf.io/nuep7 (visited on 2024) (cit. on p. 2).

Niv, Yael (2009). "Reinforcement Learning in the Brain". In: *Journal of Mathematical Psychology*. Special Issue: Dynamic Decision Making 53.3, pp. 139–154. ISSN: 0022-2496. DOI: 10.1016/j.jmp.2008.12.005. URL: http://www.sciencedirect.com/science/article/pii/S0022249608001181 (visited on 2020) (cit. on p. 3).

Niv, Yael et al. (2007). "Tonic Dopamine: Opportunity Costs and the Control of Response Vigor". In: *Psychopharmacology* 191.3, pp. 507–520. ISSN: 1432-2072. DOI: 10.1007/s00213-006-0502-4. URL: https://doi.org/10.1007/s00213-006-0502-4 (visited on 2021) (cit. on pp. 20, 22).

Novicky, Filip et al. (2022). *Bistable Perception, Precision and Neuromodulation*. DOI: 10.48550/arXiv.2212.09729. arXiv: 2212.09729 [q-bio]. URL: http://arxiv.org/abs/2212.09729 (visited on 2023) (cit. on p. 20).

Ó Scalaidhe, Séamas P., Wilson, Fraser A. W., and Goldman-Rakic, Patricia S. (1997). "Areal Segregation of Face-Processing Neurons in Prefrontal Cortex". In: *Science* 278.5340, pp. 1135–1138. DOI: 10.1126/science.278.5340.1135. URL: https://www.science.org/doi/full/10.1126/science.278.5340.1135 (visited on 2021) (cit. on p. 20).

Odegaard, Brian, Knight, Robert T., and Lau, Hakwan (2017). "Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?" In: *Journal of Neuroscience* 37.40, pp. 9593–9602. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3217-16.2017. URL: https://www.jneurosci.org/content/37/40/9593 (visited on 2020) (cit. on pp. 3, 19, 20).

Paffen, Chris L. E., Verstraten, Frans A. J., and Vidnyánszky, Zoltán (2008). "Attention-Based Perceptual Learning Increases Binocular Rivalry Suppression of Irrelevant Visual Features". In: *Journal of Vision* 8.4, pp. 25–25. ISSN: 1534-7362. DOI: 10.1167/8.4.25. URL: https://jov.arvojournals.org/article.aspx?articleid=2122424 (visited on 2021) (cit. on pp. 18, 26).

Palagina, Ganna, Meyer, Jochen F., and Smirnakis, Stelios M. (2017). "Complex Visual Motion Representation in Mouse Area V1". In: *Journal of Neuroscience* 37.1, pp. 164–183. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0997-16.2017. URL: https://www.jneurosci.org/content/37/1/164 (visited on 2023) (cit. on pp. 1, 22).

Pasley, Brian N, Mayes, Linda C, and Schultz, Robert T (2004). "Subcortical Discrimination of Unperceived Objects during Binocular Rivalry". In: *Neuron* 42.1, pp. 163–172. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(04)00155-2. URL: https://www.sciencedirect.com/science/article/pii/S0896627304001552 (visited on 2021) (cit. on p. 19).

Pastukhov, A. et al. (2013). "Multi-Stable Perception Balances Stability and Sensitivity". In: *Front Comput Neurosci* 7, p. 17. ISSN: 1662-5188 (Electronic) 1662-5188 (Linking). DOI: 10.3389/fncom.2013.00017 (cit. on p. 3).

Pastukhov, Alexander and Carbon, Claus-Christian (2022). "Change Not State: Perceptual Coupling in Multistable Displays Reflects Transient Bias Induced by Perceptual Change". In: *Psychonomic Bulletin & Review* 29.1, pp. 97–107. ISSN: 1531-5320. DOI: 10.3758/s13423-021-01960-7. URL: https://doi.org/10.3758/s13423-021-01960-7 (visited on 2023) (cit. on pp. 16, 24).

Persike, Malte, Meinhardt-Injac, Bozana, and Meinhardt, Günter (2014). "The Face Inversion Effect in Opponent-Stimulus Rivalry". In: *Frontiers in Human Neuroscience* 8, p. 295. ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00295. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4030207/ (visited on 2021) (cit. on p. 2).

Pfeffer, Thomas et al. (2018). "Catecholamines Alter the Intrinsic Variability of Cortical Population Activity and Perception". In: *PLOS Biology* 16.2, e2003453. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.2003453. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2003453 (visited on 2023) (cit. on pp. 19, 20).

Pigarev, Ivan N., Rizzolatti, Giacomo, and Scandolara, Cristiana (1979). "Neurons Responding to Visual Stimuli in the Frontal Lobe of Macaque Monkeys". In: *Neuroscience Letters* 12.2, pp. 207–212. ISSN: 0304-3940. DOI: 10.1016/0304-3940(79)96063-4. URL: https://www.sciencedirect.com/science/article/pii/0304394079960634 (visited on 2021) (cit. on p. 20).

Pizzagalli, Diego A et al. (2019). "Assessment of striatal dopamine transporter binding in individuals with major depressive disorder: in vivo positron emission tomography and postmortem evidence". In: *JAMA psychiatry* 76.8, pp. 854–861 (cit. on p. 22).

Qian, Chencan et al. (2023). *Hierarchical and Fine-Scale Mechanisms of Binocular Rivalry for Conscious Perception*. DOI: 10.1101/2023.02.11.528110. URL: https://www.biorxiv.org/content/10.1101/2023.02.11.528110v2 (visited on 2023) (cit. on p. 19).

Rao, Rajesh P. N. and Ballard, Dana H. (1999). "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects". In: *Nat Neurosci* 2.1, pp. 79–87. ISSN: 1546-1726. DOI: 10.1038/4580 (cit. on pp. 14, 24).

Reichert, David, Series, Peggy, and Storkey, Amos J. (2011). "Neuronal Adaptation for Sampling-Based Probabilistic Inference in Perceptual Bistability". In: *Advances in Neural Information Processing Systems* 24. URL: https://proceedings.neurips.cc/paper/2011/hash/b51a15f382ac914391a58850ab343b00 Abstract.html (visited on 2021) (cit. on pp. 2, 3, 5).

Rhodes, Gillian (2006). "The Evolutionary Psychology of Facial Beauty". In: *Annual Review of Psychology* 57.1, pp. 199–226. ISSN: 0066-4308. DOI: 10.1146/annurev.psych.57.102904.190208. URL: https://www.annualreviews.org/doi/10.1146/annurev.psych.57.102904.190208 (visited on 2021) (cit. on p. 2).

Robertson, C. E. et al. (2013). "Slower Rate of Binocular Rivalry in Autism". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33.43, pp. 16983–91. ISSN: 1529-2401 (Electronic) 0270-6474 (Linking). DOI: 10.1523/JNEUROSCI.0448-13.2013 (cit. on p. 22).

Rogers, Richard L., Rogers, Stephen W., and Tootle, John S. (1977). "Stimulus Complexity and Rate of Alternation in Binocular Rivalry". In: *Perceptual and Motor Skills* 44.2, pp. 669–670. ISSN: 0031-5125. DOI: 10.2466/pms.1977.44.2.669. URL: https://doi.org/10.2466/pms.1977.44.2.669 (visited on 2021) (cit. on p. 2).

Rudebeck, Peter H. and Izquierdo, Alicia (2022). "Foraging with the Frontal Cortex: A Cross-Species Evaluation of Reward-Guided Behavior". In: *Neuropsychopharmacology* 47.1, pp. 134–146. ISSN: 1740-634X. DOI: 10.1038/s41386-021-01140-0. URL: https://www.nature.com/articles/s41386-021-01140-0 (visited on 2022) (cit. on p. 3).

Safavi, Shervin and Dayan, Peter (2022). "Multistability, Perceptual Value, and Internal Foraging". In: *Neuron*. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2022.07.024. URL: https://www.sciencedirect.com/science/article/pii/S0896627322006699 (visited on 2022) (cit. on pp. 1, 2, 5, 14, 19, 22, 23).

Safavi, Shervin, Dwarakanath, Abhilash, et al. (2018). "Nonmonotonic Spatial Structure of Interneuronal Correlations in Prefrontal Microcircuits". In: *Proceedings of the National Academy of Sciences*, p. 201802356. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1802356115. URL: http://www.pnas.org/content/early/2018/03/26/1802356115 (visited on 2018) (cit. on p. 20).

Safavi, Shervin, Kapoor, Vishal, et al. (2014). "Is the Frontal Lobe Involved in Conscious Perception?" In: *Frontiers in Psychology* 5. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.01063. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01063/full (visited on 2019) (cit. on pp. 2, 3, 19, 20).

Safavi, Shervin, Logothetis, Nikos K., and Besserve, Michel (2021). "From Univariate to Multivariate Coupling between Continuous Signals and Point Processes: A Mathematical Framework". In: *Neural Computation*, pp. 1–67. ISSN: 0899-7667. DOI: 10.1162/neco_a_01389. URL: https://doi.org/10.1162/neco_a_01389 (visited on 2021) (cit. on p. 20).

Safavi, Shervin, Panagiotaropoulos, Theofanis I., et al. (2023). "Uncovering the Organization of Neural Circuits with Generalized Phase Locking Analysis". In: *PLOS Computational Biology* 19.4, e1010983. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010983. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010983 (visited on 2023) (cit. on p. 20).

Schrater, Paul R. and Sundareswara, Rashmi (2007). "Theory and Dynamics of Perceptual Bistability". In: *Advances in Neural Information Processing Systems*. Ed. by Schölkopf. MIT Press, pp. 1217–1224. ISBN: 978-0-262-25691-9 (cit. on pp. 2, 3, 5).

Schütz, Immo et al. (2018). "Visual Awareness in Binocular Rivalry Modulates Induced Pupil Fluctuations". In: *Journal of Cognition* 1.1, p. 12. ISSN: 2514-4820. DOI: 10.5334/joc.16. URL: http://www.journalofcognition.org/articles/10.5334/joc.16/ (visited on 2021) (cit. on p. 19).

Schwartz, Jean-Luc et al. (2012). "Multistability in Perception: Binding Sensory Modalities, an Overview". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1591, pp. 896–905. DOI: 10.1098/rstb.2011.0254. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2011.0254 (visited on 2021) (cit. on p. 1).

Scocchia, Lisa, Valsecchi, Matteo, and Triesch, Jochen (2014). "Top-down Influences on Ambiguous Perception: The Role of Stable and Transient States of the Observer". In: *Front. Hum. Neurosci.* 8. ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00979 (cit. on pp. 2, 5, 19, 22).

Sekutowicz, Maria et al. (2016). "Striatal Activation as a Neural Link between Cognitive and Perceptual Flexibility". In: *NeuroImage* 141, pp. 393–398. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.07.046. URL: http://www.sciencedirect.com/science/article/pii/S1053811916303603 (visited on 2021) (cit. on p. 19).

Seth, Anil K. and Bayne, Tim (2022). "Theories of Consciousness". In: *Nature Reviews Neuroscience*, pp. 1–14. ISSN: 1471-0048. DOI: 10.1038/s41583-022-00587-4. URL: https://www.nature.com/articles/s41583-022-00587-4 (visited on 2022) (cit. on p. 1).

Shadmehr, Reza et al. (2019). "Movement Vigor as a Reflection of Subjective Economic Utility". In: *Trends in Neurosciences* 42.5, pp. 323–336. ISSN: 0166-2236, 1878-108X. DOI: 10.1016/j.tins.2019.02.003. URL: https://www.cell.com/trends/neurosciences/abstract/S0166-2236(19)30019-0 (visited on 2024) (cit. on p. 20).

Shang, Junchen et al. (2020). "Perceptual Advantage of Animal Facial Attractiveness: Evidence From b-CFS and Binocular Rivalry". In: *Frontiers in Psychology* 11. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2020.01670. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01670/full (visited on 2021) (cit. on pp. 2, 9).

Shenhav, Amitai, Botvinick, Matthew M., and Cohen, Jonathan D. (2013). "The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function". In: *Neuron* 79.2, pp. 217–240. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2013.07.007. URL: https://www.sciencedirect.com/science/article/pii/S0896627313006077 (visited on 2023) (cit. on p. 22).

Sheynin, Yasha et al. (2020). "Cholinergic Modulation of Binocular Vision". In: *Journal of Neuroscience* 40.27, pp. 5208–5213. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2484-19.2020. URL: https://www.jneurosci.org/content/40/27/5208 (visited on 2020) (cit. on pp. 19, 20).

Skerswetat, Jan and Bex, Peter J. (2023). "InFoRM (Indicate-Follow-Replay-Me): A Novel Method to Measure Perceptual Multistability Dynamics Using Continuous Data Tracking and Validated Estimates of Visual Introspection". In: *Consciousness and Cognition* 107, p. 103437. ISSN: 1053-8100. DOI: 10.1016/j.concog.2022.103437. URL: https://www.sciencedirect.com/science/article/pii/S1053810022001696 (visited on 2022) (cit. on p. 20).

Spiegel, Alina et al. (2019). "Slower Binocular Rivalry in the Autistic Brain". In: *Current Biology* 29.17, 2948–2953.e3. ISSN: 09609822. DOI: 10.1016/j.cub.2019.07.026 (cit. on p. 22).

Sterzer, P. et al. (2011). "Access of Emotional Information to Visual Awareness in Patients with Major Depressive Disorder". In: *Psychological Medicine* 41.8, pp. 1615–1624. DOI: 10.1017/S0033291710002540 (cit. on p. 8).

Sundareswara, R. and Schrater, P. R. (2008). "Perceptual Multistability Predicted by Search Model for Bayesian Decisions". In: *J Vis* 8, pp. 12 1–19. ISSN: 1534-7362 (Electronic) 1534-7362 (Linking). DOI: 10.1167/8.5.12 (cit. on pp. 2, 3, 5).

Sutton, Richard S. and Barto, Andrew G. (2018). *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press. ISBN: 978-0-262-03924-6 (cit. on pp. 14, 23, 41).

Suzuki, Satoru and Grabowecky, Marcia (2007). "Long-Term Speeding in Perceptual Switches Mediated by Attention-Dependent Plasticity in Cortical Visual Processing". In: *Neuron* 56.4, pp. 741–753. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2007.09.028. URL: https://www.cell.com/neuron/abstract/S0896-6273(07)00754-4 (visited on 2021) (cit. on pp. 2–4, 13, 15, 17, 20, 25).

Theodoni, Panagiota (2014). "Fluctuations in Perceptual Decisions : Cortical Microcircuit Dynamics Mediating Alternations in Conscious Visual Perception". PhD thesis. Universitat Pompeu Fabra. URL: http://www.tdx.cat/handle/10803/145642 (visited on 2018) (cit. on p. 22).

Tinio, Pablo PL, Leder, Helmut, and Strasser, Marlies (2011). "Image quality and the aesthetic judgment of photographs: Contrast, sharpness, and grain teased apart and put together." In: *Psychology of Aesthetics, Creativity, and the Arts* 5.2, p. 165 (cit. on p. 12).

Toepfer, Franziska, Wolf, Reinhard, and Heisenberg, Martin (2018). "Multi-Stability with Ambiguous Visual Stimuli in Drosophila Orientation Behavior". In: *PLOS Biology* 16.2, e2003113. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.2003113. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2003113 (visited on 2022) (cit. on p. 1).

Tsikandilakis, Myron, Bali, Persefoni, and Chapman, Peter (2019). "Beauty Is in the Eye of the Beholder: The Appraisal of Facial Attractiveness and Its Relation to Conscious Awareness". In: *Perception* 48.1, pp. 72–92. ISSN: 0301-0066. DOI: 10.1177/0301006618813035. URL: https://doi.org/10.1177/0301006618813035 (visited on 2021) (cit. on pp. 2, 9).

Tsuchiya, Naotsugu and Koch, Christof (2004). "Continuous Flash Suppression". In: *Journal of Vision* 4.8, pp. 61–61. ISSN: 1534-7362. DOI: 10.1167/4.8.61. URL: https://jov.arvojournals.org/article.aspx?articleid=2131303 (visited on 2021) (cit. on p. 8).

Tsuchiya, Naotsugu, Wilke, Melanie, et al. (2015). "No-Report Paradigms: Extracting the True Neural Correlates of Consciousness". In: *Trends in Cognitive Sciences* 19.12, pp. 757–770. ISSN: 1364-6613. DOI: 10.1016/j.tics.2015.10.002. URL: http://www.sciencedirect.com/science/article/pii/S1364661315002521 (visited on 2020) (cit. on pp. 3, 19, 20, 22).

Turi, Marco, Burr, David Charles, and Binda, Paola (2018). "Pupillometry Reveals Perceptual Differences That Are Tightly Linked to Autistic Traits in Typical Adults". In: *eLife* 7. Ed. by Rust, e32399.

ISSN: 2050-084X. DOI: 10.7554/eLife.32399. URL: https://doi.org/10.7554/eLife.32399 (visited on 2022) (cit. on pp. 19, 22).

van Ee, Raymond (2005). "Dynamics of Perceptual Bi-Stability for Stereoscopic Slant Rivalry and a Comparison with Grating, House-Face, and Necker Cube Rivalry". In: *Vision Research* 45.1, pp. 29–40. ISSN: 0042-6989. DOI: 10.1016/j.visres.2004.07.039. URL: https://www.sciencedirect.com/science/article/pii/S0042698904003839 (visited on 2021) (cit. on pp. 3, 20).

Wang, Mengxin, McGraw, Paul, and Ledgeway, Timothy (2021). "Attentional Eye Selection Modulates Sensory Eye Dominance". In: *Vision Research* 188, pp. 10–25. ISSN: 0042-6989. DOI: 10.1016/j.visres.2021.06.006. URL: https://www.sciencedirect.com/science/article/pii/S0042698921001395 (visited on 2021) (cit. on pp. 18, 26).

Weilnhammer, V. et al. (2017). "A Predictive Coding Account of Bistable Perception - a Model-Based fMRI Study". In: *PLoS Comput Biol* 13, e1005536. ISSN: 1553-7358 (Electronic) 1553-734X (Linking). DOI: 10.1371/journal.pcbi.1005536 (cit. on pp. 2, 3, 5).

Weilnhammer, Veith, Chikermane, Meera, and Sterzer, Philipp (2021). "Bistable perception alternates between internal and external modes of sensory processing". In: *Iscience* 24.3 (cit. on p. 2).

Weilnhammer, Veith, Fritsch, Merve, et al. (2021a). "An Active Role of Inferior Frontal Cortex in Conscious Experience". In: *Current Biology*. ISSN: 0960-9822. DOI: 10.1016/j.cub.2021.04.043. URL: https://www.sciencedirect.com/science/article/pii/S0960982221005789 (visited on 2021) (cit. on p. 3).

— (2021b). "An Active Role of Inferior Frontal Cortex in Conscious Experience". In: *Current Biology*. ISSN: 0960-9822. DOI: 10.1016/j.cub.2021.04.043 (cit. on p. 19).

Wilbertz, Gregor and Sterzer, Philipp (2018). "Differentiating Aversive Conditioning in Bistable Perception: Avoidance of a Percept vs. Salience of a Stimulus". In: *Consciousness and Cognition* 61, pp. 38–48. ISSN: 1053-8100. DOI: 10.1016/j.concog.2018.03.010. URL: http://www.sciencedirect.com/science/article/pii/S105381001730394X (visited on 2021) (cit. on pp. 18, 26).

Wilbertz, Gregor, van Kemenade, Bianca M., et al. (2017). "fMRI-based Decoding of Reward Effects in Binocular Rivalry". In: *Neuroscience of Consciousness* 2017.1. DOI: 10.1093/nc/nix013. URL: https://academic.oup.com/nc/article/2017/1/nix013/3916729 (visited on 2020) (cit. on pp. 2, 7, 17, 19, 26).

Wilbertz, Gregor, van Slooten, Joanne, and Sterzer, Philipp (2014). "Reinforcement of Perceptual Inference: Reward and Punishment Alter Conscious Visual Perception during Binocular Rivalry". In: *Frontiers in Psychology* 5. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.01377. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01377/full (visited on 2020) (cit. on pp. 2, 7, 9, 17, 19, 26).

Wilke, M., Mueller, K. M., and Leopold, D. A. (2009). "Neural Activity in the Visual Thalamus Reflects Perceptual Suppression". In: *Proc Natl Acad Sci U S A* 106.23, pp. 9465–70. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.0900714106 (cit. on pp. 3, 19).

Wu, Wayne (2023). "We Know What Attention Is!" In: *Trends in Cognitive Sciences*. ISSN: 1364-6613. DOI: 10.1016/j.tics.2023.11.007. URL: https://www.sciencedirect.com/science/article/pii/S1364661323002863 (visited on 2023) (cit. on p. 21).

Yang, Eunice, Zald, David H., and Blake, Randolph (2007). "Fearful Expressions Gain Preferential Access to Awareness during Continuous Flash Suppression". In: *Emotion (Washington, D.C.)* 7.4, pp. 882–886. ISSN: 1528-3542. DOI: 10.1037/1528-3542.7.4.882 (cit. on p. 8).

Ye, Xing et al. (2019). "Slower and Less Variable Binocular Rivalry Rates in Patients With Bipolar Disorder, OCD, Major Depression, and Schizophrenia". In: *Frontiers in Neuroscience* 13. ISSN: 1662-453X. DOI: 10.3389/fnins.2019.00514. URL: https://www.frontiersin.org/articles/10.3389/fnins.2019.00514/full (visited on 2020) (cit. on pp. 2, 22).

Yu, Angela J. and Dayan, Peter (2005). "Uncertainty, Neuromodulation, and Attention". In: *Neuron* 46.4, pp. 681–692. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2005.04.026. URL: https://www.sciencedirect.com/science/article/pii/S0896627305003624 (visited on 2021) (cit. on p. 19).

Yuille, Alan and Kersten, Daniel (2006). "Vision as Bayesian inference: analysis by synthesis?" In: *Trends in cognitive sciences* 10.7, pp. 301–308 (cit. on pp. 3, 14, 24).

Zanto, Theodore P and Gazzaley, Adam (2013). "Fronto-parietal network: flexible hub of cognitive control". In: *Trends in cognitive sciences* 17.12, pp. 602–603 (cit. on p. 20).

Zaretskaya, N. and Narinyan, M. (2014). "Introspection, Attention or Awareness? The Role of the Frontal Lobe in Binocular Rivalry". In: *Front Hum Neurosci* 8, p. 527. ISSN: 1662-5161 (Electronic) 1662-5161 (Linking). DOI: 10.3389/fnhum.2014.00527 (cit. on pp. 3, 19, 20).

Zaretskaya, N., Thielscher, A., et al. (2010). "Disrupting Parietal Function Prolongs Dominance Durations in Binocular Rivalry". In: *Curr Biol* 20.23, pp. 2106–2111. ISSN: 0960-9822. DOI: DOI10.1016/j.cub.2010.10.046 (cit. on p. 19).

Zaretskaya, Natalia (2021). "Zooming-in on Higher-Level Vision: High-resolution fMRI for Understanding Visual Perception and Awareness". In: *Progress in Neurobiology*, p. 101998. ISSN: 0301-0082. DOI: 10.1016/j.pneurobio.2021.101998. URL: https://www.sciencedirect.com/science/article/pii/S0301008221000125 (visited on 2021) (cit. on p. 1).

Appendix

Method details

Model of aesthetic value

We used a modern theory of aesthetic value (Brielmann and Dayan, 2022), which suggests that the utility of a sensory object derives from its influence in adapting the brain's sensory processing system to the demands of the stimuli that it expects to encounter in the present and over the long run. As this model was introduced and described extensively in the study of Brielmann and Dayan (2022), we restrict ourselves to a brief exposition of its key aspects (and our own notation).

Here we represent stimuli, or in the context of this study, $\hat{\mathbf{s}}_t$, which is the estimate of the agent for the richer state $\mathbf{s}_t$ (introduced earlier in section Thick state). In principle, this formulation allows for different qualities of the perceptual experience through estimated stimuli. However, for simplicity, here we assumed that this estimate is limited to exact features of either left or right stimulus, i.e., we assume, for instance, there is no morph perception (e.g., as reported by Klink, Boucherie, et al., 2017) or piecemeal perception (Carmel et al., 2010).

Based on the theory of aesthetic value (Brielmann and Dayan, 2022), we assume that the agent has a *momentary or contextual* probabilistic model of what it thinks it might sense (called the system state, denoted by $\boldsymbol{X}_t$), and a second model of the sensory input for the sensory input that the agent expects over the long run (called the expected true distribution, denoted by p^{T}). Following Brielmann and Dayan (2022), we assume that both momentary and long-term distributions are normal

$$P(\mathbf{s};\mathbf{X}_t) = \mathcal{N}\left(\boldsymbol{\mu}_t^s, \Sigma^s\right) , \tag{17}$$

$$P^{\mathrm{T}}(\mathbf{s}) = \mathcal{N}\left(\boldsymbol{\mu}^{\mathrm{T}}, \Sigma^{\mathrm{T}}\right) . \tag{18}$$

The system state is dynamic, i.e., with continuous exposure to a stimulus (perception of a stimulus), the system state gradually evolves toward the presented stimulus, thus $\boldsymbol{\mu}_t^s$ continuously adapts with a constant learning rate of α,

$$\boldsymbol{\mu}_{t+1}^s = \boldsymbol{\mu}_t^s + \alpha(\mathbf{s} - \boldsymbol{\mu}_t^s) . \tag{19}$$

This change, relative to stimulus and the expected true distribution, determines the magnitude of aesthetic value.

The immediate perceptual reward, $r(t)$, is defined as the log-likelihood of a stimulus given the system state at time t,

$$r(\mathbf{s}_t) = k - {}^1\!/_2(\mathbf{s}_t - \boldsymbol{\mu}_t)^{\mathrm{T}}(\Sigma^s)^{-1}(\mathbf{s}_t - \boldsymbol{\mu}_t) , \tag{20}$$

where k is a constant. Notably, this has a similar formulation to the one we introduced earlier in ***Equation 12***, in which we treated the the *negative* log-likelihood of the observed ocular input as a component of the loss function – and, indeed, in an hierarchical formulation of a generative model of sensory input (Dayan, Hinton, et al., 1995; Brascamp, Sterzer, et al., 2018), these terms would also be directly combined in the loss function. In the simple characterisation here in which we do not explicitly model a belief state over the perceptual details of the input, we imagine that the internal percept is exactly the input to one of the two eyes, and is directly aesthetically valued.

Stimuli that generate substantial learning or change also have a high long-term utility, thus, Brielmann and Dayan (2022) defined the second component of the aesthetic reward as follows. The system state has an inherent value that is the average expected future sensory reward, which we write here as:

$$V(\mathbf{X}_t; t) = K - KL\left(P^{\mathrm{T}}(\mathbf{s};t) || P(\mathbf{s};\mathbf{X}_t)\right) , \tag{21}$$

where K is a constant, and KL is the Kullback-Leibler divergence between the distribution over those stimuli implied by the momentary system state, $P(\mathbf{s};\mathbf{X}_t)$ at time t and the expected true distribution $P^{\mathrm{T}}(\mathbf{s};t)$, that the agent believes at time t will arise in the long-term future (inspired by

temporal difference error from reinforcement learning, Sutton et al., 2018). Thus, the utility stemming from the change in system state, ΔV, follows as,

$$\Delta V_{\hat{\mathbf{s}}_t}(\mathbf{X}_t; t) = V(\mathbf{X}_{t+1}; t+1) - V(\mathbf{X}_t; t)\,. \tag{22}$$

acknowledging explicitly that the change to $\mathbf{X}_t$ is occasioned by $\hat{\mathbf{s}}_t$. Ultimately, the aesthetic value, $R_A(\mathbf{s}, t)$, is defined as the weighted sum of these two components (and potentially with the constant weight, w_0),

$$R_A(\mathbf{s}_t, t) = w_0 + w^r r(\mathbf{s}_t) + w^V \Delta_{\mathbf{s}_t} V(\mathbf{X}_t; t)\,. \tag{23}$$

In the description of our decision-making model (section Perceptual multistability as a POMDP) we used a^P (the perceptual action) instead of characteristics of the state ($\hat{\mathbf{s}}$) as a shorthand notation, thus we referred to aesthetic value by $R_A(a_t^P, t)$.

Partially observable Markov decision process

Belief formation and maintenance. The agent forms and maintains a belief about the state of the world (as described earlier, in section Thin state), i.e., the agent's posterior distribution over states, given its observations. Let $b_{t-1}(s)$ denote the probability assigned to state s by belief state b_{t-1} and let b_t be the updated belief state given the old belief state b_{t-1}, action a_t^P, and observation o_t. In a general form, the agent's belief can be updated as follows,

$$b_t(s) = \frac{\left[\sum_{s' \in \mathcal{S}} b_{t-1}(s') T\left(s', a_t^P, s\right)\right] O\left(s, a_t^P, \mathbf{o}_t\right)}{Z}\,, \tag{24}$$

where Z is a normalizing factor (ensuring the sum of beliefs remains 1). Given that the perceptual actions of the agent do not change the state of the world, the state transition function does not depend on the perceptual actions of the agent, therefore, belief update can be slightly simplified,

$$b_t(s) = \frac{\left[\sum_{s' \in \mathcal{S}} b_{t-1}(s') T\left(s', s\right)\right] O\left(s, a_t^P, \mathbf{o}_t\right)}{Z}\,. \tag{25}$$

Solving the POMDP. The optimal decision/policy allows us to create a mapping from the belief states to the optimal actions, i.e., selecting actions that lead to the maximum utility for the agent. Based on Kaelbling et al. (1998)'s formulation, we use the notion of state estimator function, $\mathrm{SE}(b_{t-1}, a_t^P, \mathbf{o}_t)$, which provides an updated belief state, based on the previous belief state at time $t-1$, the selected perceptual action and observation at time t:

$$p\left(b_t \mid \mathbf{o}_t, a_t^P, b_{t-1}\right) = \begin{cases} 1 & \text{if } \mathrm{SE}\left(b_{t-1}, a_t^P, \mathbf{o}_t\right) = b_t, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

The reward function in the belief MDP is derived from the original reward function via

$$\rho(b_t, a_t^P, t) = \sum_{s' \in \mathcal{S}} b_t(s') R(s', a_t^P, t)\,. \tag{27}$$

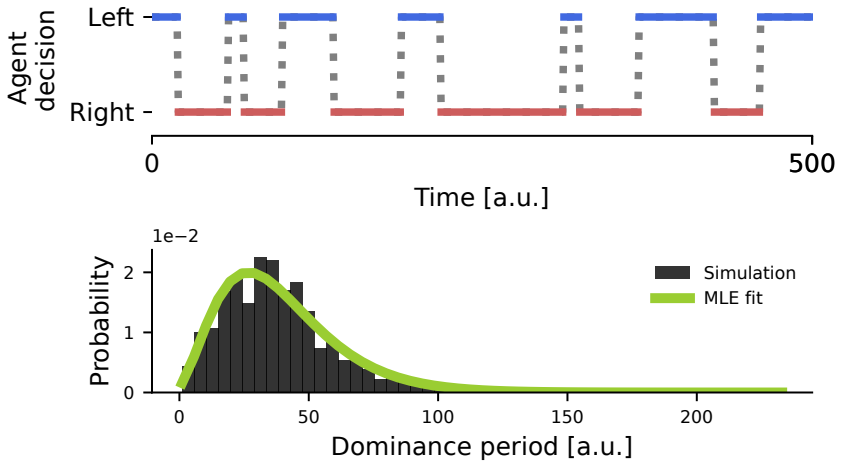
The objective is finding the optimal policy, π^*, which in this case is only the immediate action (as we solve the horizon 1 case). Thus we need to have an estimate for the expected return for performing some action in a given belief state b. These are called state-action or Q-values:

$$Q^\pi(b, a^p) = \mathbb{E}_\pi\left[\rho(b_t, a_t^P, t)\right]\,. \tag{28}$$

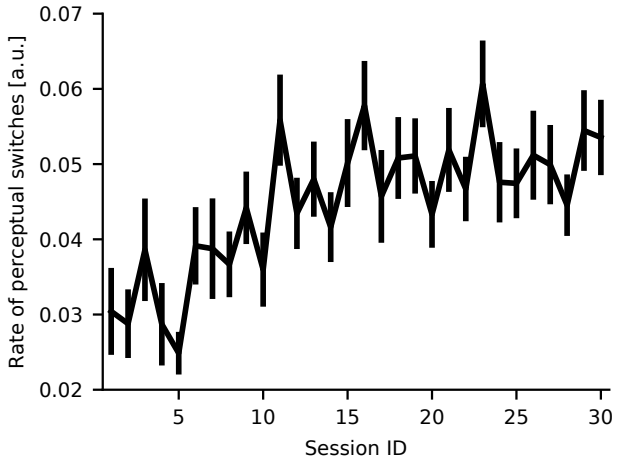
As the goal is to execute an optimal policy, π^*, the action leading to maximum return should be selected self-consistently:

$$(a^p)^* = \mathrm{argmax}_{(a^p)' \in \mathcal{A}}\, Q^{\pi^*}\left(b, (a^p)'\right)\,. \tag{29}$$

This provides a mapping from belief at time t to a perceptual action a_t^P.



Supplementary Figure 1. A simple POMDP with reward asymmetry.
(Top) An example trajectory of the agent's decision sequence between left and right images in a binocular rivalry task when one percept is rewarded (right percept is rewarded). **(Bottom)** Distribution of dominance periods for an extended length of simulation. Gamma distribution fitted with maximum likelihood estimation (MLE) and indicated with a green line.



Supplementary Figure 2. Temporal dynamics of perceptual multistability across sessions – an example.
Similar to ***Figure 6***, long-term speed-up of switching rate across sessions for one participant, and a single simulated participant. Switching rates are quantified as the reciprocal of perceptual dominance durations. Individual data points for each session are averages across 5 trials.