The time course of binocular rivalry during the phases of the menstrual cycle

Jocelyn L. Sy Department of Psychology, Vanderbilt University, Nashville, TN, USA

Andrew J. Tomarken Department of Psychology, Vanderbilt University, Nashville, TN, USA

Vaama Patel Department of Psychology, Vanderbilt University, Nashville, TN, USA

Randolph Blake Department of Psychology, Vanderbilt University, Nashville, TN, USA

Binocular rivalry occurs when markedly different inputs to the two eyes initiate alternations in perceptual dominance between the two eyes' views. A link between individual differences in perceptual dynamics of rivalry and concentrations of GABA, a prominent inhibitory neurotransmitter in the brain, has highlighted binocular rivalry as a potential tool to investigate inhibitory processes in the brain. The present experiment investigated whether previously reported fluctuations of GABA concentrations in a healthy menstrual cycle (Epperson et al., 2002) also are associated with measurable changes in rivalry dynamics *within* individuals. We obtained longitudinal measures of alternation rate, dominance, and mixture durations in 300 rivalry tracking blocks measured over 5 weeks from healthy female participants who monitored the start of the follicular and luteal phases of their cycle. Although we demonstrate robust and stable individual differences in rivalry dynamics, across analytic approaches and dependent measures, we found no significant change or even trends across menstrual phases in the temporal dynamics of dominance percepts. We found only sparse between-phase differences in skew and kurtosis on mixture percepts when data were pooled across sessions and blocks. These results suggest a complex dynamic between hormonal steroids, binocular rivalry, and GABAeric signaling in the brain and thus implicate the need to consider a systemic perspective when linking GABA with perceptual alternations in binocular rivalry.

Introduction

Binocular rivalry is a remarkable phenomenon in which visual confusion created by the presence of conflicting inputs to each eye sparks alternations in perceptual dominance between the two eyes' views. It is widely believed that rivalry results from differences in activity levels among neural ensembles representing the competing rival stimuli—differences that are modulated by a dynamically changing interplay between excitation and inhibition (Kang & Blake, 2010; Klink, Brascamp, Blake, & van Wezel, 2010; Lehky, 1988; Said & Heeger, 2013; Wilson, 2003). There is active debate concerning the locus of this interplay within the visual hierarchy, with some favoring neural interactions within early stages of processing including the primary visual cortex (e.g., Lehky & Blake, 1991) whereas others pinpoint the lynchpin stages higher within the visual hierarchy (e.g., Leopold & Logothetis, 1999). But essentially, all accounts of rivalry include this notion of neural competition mediated by reciprocal inhibition, including hybrid models in which rivalry is characterized as the culmination of neural events distributed over different stages of the visual hierarchy (Blake & Logothetis, 2002; Hohwy, Roepstorff, & Friston, 2008; Tong, Meng, & Blake, 2006).

According to reciprocal inhibition, dominance of one stimulus is the consequence of excitation of the neural ensemble coding for its percept and the reciprocal inhibition of the competing percept arising from dissimilar stimulation of the other eye. Neural

Citation: Sy, J. L., Tomarken, A. J., Patel, V., Blake, R. (2016). The time course of binocular rivalry during the phases of the menstrual cycle. *Journal of Vision*, *16*(15):22, 1–19, doi:10.1167/16.15.22.

doi: 10.1167/16.15.22 Received June 9, 2016; published December 22, 2016 ISSN 1534-7362

excitation associated with the dominant stimulus decreases over time owing to adaptation, causing the inhibition exerted by that stimulus to weaken as well. Eventually, the relative strength of the two representations reverses, triggering a perceptual state change. According to extant neural models of rivalry, the rate of binocular rivalry alternations in dominance should depend on levels of inhibition exerted upon neurons whose activity represents the currently suppressed stimulus: Stronger inhibition promotes longer dominance durations and thus slower rates of alternation. To be sure, other factors, too, govern the rivalry rate, including intrinsic noise and adaptation time constants. But inhibition is a major ingredient in rivalry dynamics and one that is potentially relatable to endogenous neurochemical factors (e.g., Klink et al., 2010; van Loon et al., 2013).

Inhibitory processes are ubiquitous within the nervous system and play essential roles in the regulation of neural activity and neural plasticity. Within the brain, inhibition is mediated predominantly by gamma amino butyric acid (GABA), a neurotransmitter that activates several receptor types. Consistent with inhibitory models of rivalry alternations, recent research has demonstrated that short-term, monocular eye patching temporarily affects the subsequently measured temporal dynamics of dominance in favor of the previously patched eye and, at the same time, decreases GABA concentrations in the occipital cortex (Lunghi, Emir, Morrone, & Bridge, 2015). A possible relationship between GABA and perceptual rivalry alternations is also suggested by increased durations of rivalry dominance induced pharmacologically by lorazepam, a $GABA_A$ agonist (van Loon et al., 2013). From an individual differences perspective, individuals with greater resting concentrations of GABA in the occipital cortex exhibit longer durations of perceptual dominance in rivalry (van Loon et al., 2013). These correlations between GABA levels and rivalry dynamics suggest that binocular rivalry may provide a reliable proxy for GABA concentrations in the human brain. If so, given prior evidence indicating atypical patterns of rivalry dynamics linked to schizophrenia (Tononi & Edelman, 2000) and autism (Freyberg, Robertson, & Baron-Cohen, 2015; Robertson, Kravitz, Freyberg, Baron-Cohen, & Baker, 2013), the linkage between GABA and rivalry might promote investigations of the neurochemical and inhibitory imbalances associated with these disorders. However, the evidence for atypical rivalry dynamics linked to these conditions is not uniformly consistent (Miller et al., 2003; Said, Egan, Minshew, Behrmann, & Heeger, 2013), suggesting that the relation between GABA and perceptual alternations may be more nuanced and thus require further consideration. Toward that end, we have taken a different approach to the study of this putative relationship, one based on endogenous neuropharmacological changes in ovulating females.

It is well established that hormone steroids and their metabolites affect various neurotransmitter systems, including GABAergic systems (Akk et al., 2005; Smith, Waterhouse, Chapin, & Woodward, 1987; Smith, Waterhouse, & Woodward, 1988). Because progesterone and estradiol fluctuate dramatically within a menstrual cycle, it is not surprising that the menstrual cycle has been linked to variable behavioral performance in a variety of cognitive, motor, auditory, and visual spatial tasks (Hampson, 1990; Maki, Rich, & Rosenbaum, 2002). Of prime importance in the present context, Epperson and colleagues (2002) have reported that the follicular phase of the menstrual cycle, in which estradiol is high and progesterone is low, has greater in vivo GABA concentrations in the occipital cortex compared with the mid- and late-luteal phases in the menstrual cycles of normal, healthy women, in which estradiol is low and progesterone is high. Our experiment capitalized on the reported fluctuations in GABA that occur during the course of a woman's menstrual cycle to test the extent to which the dynamics of binocular rivalry also vary during the course of a menstrual cycle. Given the relationship of GABA concentration in the occipital cortex to perceptual alternations in rivalry and the measured fluctuations of occipital GABA concentrations in the menstrual cycle, we surmised that the higher concentration of GABA during the follicular phase compared with the mid- and late-luteal phases should be the basis of strengthened cortical inhibition during that phase. Based on the results reported by Epperson and colleagues (2002) and those results summarized in the previous paragraph, we surmised in turn that relatively stronger inhibition during the follicular phase could produce longer periods of perceptual dominance and thus slower alternation rates during that phase. To test this hypothesis, we obtained repeated measures of binocular rivalry state dynamics over the course of the female menstrual cycle.

Methods

Participants

Sixteen female participants (mean age = 22) with normal or corrected-to-normal visual acuity and no history of ocular pathology volunteered to participate in this study. These young adult participants reported that they were not actively taking birth control or fertility medications, were not pregnant or breastfeeding, and had not previously experienced irregular cycles in the past 6 months. Participants also had no

history of premenstrual dysphoric, neurological, or psychiatric disorders; no history of drug or alcohol abuse; and no history of consumption of tobacco (smoking or chewing), mood stabilizers, benzodiazepines, or anticonvulsant medications. One participant was excluded because she lacked data sampled from one menstrual phase used for analysis. A second participant was excluded because she began taking birth control halfway through the testing period. A third participant was excluded from analysis because more than 50% of her dominance durations were shorter than 500 ms, placing her data well outside the range of durations evidenced by all others in our sample and, for that matter, far outside the range routinely encountered in other rivalry studies from our laboratory.

Each participant gave written and informed consent in compliance with the experimental protocol approved by the Institutional Review Board at Vanderbilt University. Participants received $10 per 30-min session and an additional $40 bonus for completion of the entire, 5-week study.

Apparatus and stimuli

Stimuli were generated on a Macintosh computer running MATLAB in conjunction with routines from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Observers viewed the display in a darkened room on a gamma-corrected CRT (21-in. Sony Multi-Scan; refresh rate 100 Hz). The display was viewed through a mirror stereoscope that presented half of the monitor display exclusively to one eye and the other half of the display to the other eye.

To induce binocular rivalry, orthogonally oriented, circular sinusoidal gratings (±45°; 4.5 cycles/°; 30% contrast; 1.4° diameter) were dichoptically presented to corresponding retinal areas of the two eyes; a small circular fixation mark appeared in the center of each grating. To stabilize binocular alignment, each rival target was framed by a textured fusion figure (3.8° × 3.8°; see Figure 1) that was identical for each eye. Each test session started with administration of an automated alignment task wherein the participant adjusted the X/Y positions of the fusion frames to which were added a pair of nonius markers. Those adjustments were made so as to achieve stable binocular alignment as those fusion frames were alternately presented and removed. Fusion was signified by invariance in perceived visual position of the left- and right-eye frames and the alignment of the nonius markers. Participants performed this alignment procedure three times in succession, with the positions of each dichoptic frame offset by a variable amount before each alignment adjustment. The average of the X/Y positions of the three settings was used to situate the pair of rival targets and fuser frames on the video monitor during the experimental tracking task. In each experimental trial, the rival targets appeared centered within the circular fusion frame. The order of pairings of eye and orientation was randomized within each testing session, with the stipulation that each grating orientation was presented to each eye an equal number of times.

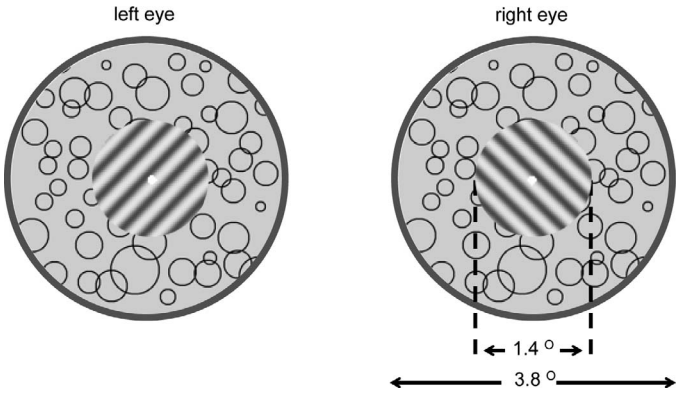

Figure 1. Schematic of left and right eye dichoptic stimuli. The outer "fusion" patterns were identical in both eyes' views, and the interior rival gratings were orthogonally oriented (with eye and orientation counterbalanced over trials for each test session).

Procedure

Each observer participated in 15 experimental testing sessions (three times a week for 5 weeks), and each daily session comprised 20 individual blocks of rivalry tracking. During a given block of tracking, the participant pressed and held down one of two arrow keys on the computer keyboard to designate which grating orientation was exclusively dominant. Observers withheld key presses when perceptual mixtures were experienced, and the durations of those "no-press" episodes were also recorded. It was carefully explained to participants that a button should be pressed only when one orientation or the other was exclusively visible and that mixtures referred to simultaneous visibility of portions of each grating, either as a patchwork mosaic or as a superimposition of both orientations. We also explained that one might experience a sequence of rivalry states in which dominance of one stimulus transitions into a mixture comprising both stimuli followed by a transition back to the previously dominant stimulus. Referred to as "return transitions" in the literature (Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006), these comprised 13.88% of the dominance states experienced by our observers. Each tracking block lasted a

minimum of 60 s. The first state change reported after 60 s ended the tracking block (e.g., mixture to dominance, or dominance of the left eye percept to dominance of the right eye percept). This trial termination rule allowed us to include the last dominance state in that tracking period without prematurely truncating it by turning off the display.

Prior to the first day of experimental testing, each observer underwent a routine visual screening battery using the Keystone Orthorator to confirm that she had normal acuity in both eyes and good stereopsis. During this pretest session, we also familiarized the observer with the rivalry displays, the alignment procedure for ensuring accurate binocular alignment of the dichoptic displays, and the tracking task. This pretest session also included five practice blocks of binocular rivalry tracking.

Because of variation between women in the duration of their cycle, longitudinal analyses of alternation rate were based on the day in which a participant measured a spike in her luteinizing hormone and the date of the first day of menses. During the training session, participants were also instructed how to monitor the start of their menstrual cycle and a surge in the luteinizing hormone, the hormone known to trigger ovulation (within 24–48 hr) in a normal healthy menstrual cycle. The surge in luteinizing hormone can be measured with over-the-counter ovulation tests of urine samples; for our study, we used the BFP ovulation test strips by Fairhaven Health, which reports 99.9% accuracy detecting 25 miu/mL. Participants were instructed to use the ovulation test, at home, on a daily basis for a week. Ovulation testing began just prior to halfway through their menstrual cycle. For example, if a participant reported an average menstrual cycle that lasts 28 days, and ovulation typically occurs halfway through the cycle, the participant was instructed to use the ovulation tests on day 10–16 of their cycle or until the luteinizing hormone spike is first measured. Participants were asked to report the first day of any menstrual flow or any spike in luteinizing hormone during the 5 weeks of testing via a confidential internet-based survey. This reporting procedure permitted experimenters to be naïve to the menstrual phase of the participant.

Analyses

By way of preview, our initial examination of the results following the end of the study revealed no obvious differences in rivalry dynamics among the three menstrual cycle phases of interest. Thus, to minimize the likelihood of failing to reject the null hypothesis if, in fact, it were false, and to provide an optimally comprehensive assessment of possible differences among phases, we assembled a variety of different statistical strategies (parametric and nonparametric) and measures for comparing the three phases. This section describes those strategies in some detail, starting with the procedure for dealing with aberrant durations.

As explained in the Methods section, participants were instructed to press one of two buttons to signal exclusive dominance of one or the other of the two rival patterns and to press neither key when mixtures were experienced. During rivalry, mixture periods sometimes intervene between states of exclusive dominance but other times do not. For some state transitions, in other words, the previously suppressed stimulus can abruptly and unpredictably achieve complete dominance. When that happens, the observer requires some minimum decision/motor response time to register an abrupt transition and to generate the sequence of key press changes required to signify that transition. Such transitions inevitably produce a small but nontrivial proportion of no-press periods lasting a fraction of a second. Because we did not want to include those kinds of transitions in the mixture category, we needed to designate some brief criterion duration to specify the minimum duration to be accepted as a mixture. We set that criterion duration at 500 ms, a value based on results from a rivalry replay condition created and tested during execution of another experiment in our laboratory (Dieter & Blake, 2015). During replay trials in that experiment, observers pressed one of two buttons to track unpredictable physical switches between two dissimilarly oriented gratings that were physically interchanged following a temporal pattern that matched previously measured records of transitions associated with genuine binocular rivalry. The replay sequence also included mixture transitions mimicked by local spatial morphing between the two dissimilar gratings. From the tracking records and the knowledge about when the physical changes in states were happening, we were able to measure the lag times between the tracking responses to these transitions. Among the 15 observers tested on that task, the median lag time was 512 ms, prompting us to discard all durations 500 ms or shorter from the statistical analyses of state durations in the present study. To be on the safe side, we repeated statistical tests using data sets created by lowering this minimum duration value to 300 ms; results from those analyses led to the same conclusions as those associated with the 500-ms criterion. We also excluded any durations in excess of 60 s across participants, an unprecedented event in normal rivalry tracking and one suggestive of finger mistakes on the part of the participant ($<0.01\%$ of events).

Next, we focused our analyses on portions of the menstrual cycle during which there was reason to

believe GABA levels were fluctuating. According to Epperson et al. (2002), healthy females demonstrated higher GABA concentrations during the follicular phase (+3 to +8 days from menstruation), compared with mid-luteal (+3 to +8 days from luteinizing hormone spike) and late-luteal (−5 to −1 days from menstruation) phases. We used these demarcations to bin data in each of the three phases, and data from sessions that did not fall into one of the three phases were excluded from analyses. To test for between-phase differences on the average durations of rivalry epochs, we specified linear mixed models (LMMs) using SAS PROC GLIMMIX, Version 9.4, of the SAS System for Windows™ (Copyright © 2002–2015 SAS Institute Inc., SAS and all other SAS Institute Inc. products or service names are registered trademarks of SAS Institute Inc., Cary, NC; e.g., Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006; Stroup, 2013). Analyses were performed on the average durations of dominance and mixture epochs per block, thus yielding 20 scores per measure per session for each participant. We analyzed block averages instead of the raw, epoch-level data because preliminary assessments indicated that aggregation of the observations within a block facilitated model convergence, substantially reduced computation time, and provided estimates of random effect and residual parameters with better fit and likely better reproducibility. The block means were log-transformed to better meet the LMM assumption of normality. The results of the LMM analyses were unchanged when we weighted each block mean by the number of epochs within a given block of the given state (dominance or mixtures). In the Results section, we report the outcomes from the unweighted analyses for clarity and because they simplify the computation of proportion of variance measures.

Each LMM model specified fixed effects denoting the effects of phase, session, and block. Because preliminary analyses indicated no significant higher-order interactions involving the phase and block factors, we dropped such terms from the model and specified only main effect terms for phase and block. Because only those sessions that fell within one of the three phases were included (on average 7.84 sessions per participant were included in one of the three phases and 2.61 sessions per participant were included per phase), the data were rather sparse in some cells formed by the crossing of session and phase (e.g., a few session numbers were not represented in a given phase). Because it can be problematic to model interactions with such sparse data, we treated session only as a main effect term and view it primarily as a control variable. Because of the high number of levels of the block (20) and session (15) factors and the apparent nonlinearity of changes in mean duration across blocks and sessions, we tested for the effects of each using spline functions. Splines can reveal nonlinear patterns of change while avoiding the distortions imposed by higher-order polynomials because splines emphasize local rather than global features of the data (for reviews, see, e.g., Keele, 2008; Ruppert, Wand, & Carroll, 2003). We specified truncated polynomial cubic splines yielding smooth functions that were continuous at specific points of the predictors (block and session) known as knots. Preliminary comparisons using the Akaike information criterion (AIC; Akaike, 1974) indicated that for both block and session and for both dominance and mixture durations, three interior knots specified at the 25th, 50th, and 75th percentiles of block (corresponding to Blocks 6, 11, and 16) and session (corresponding to Sessions 5, 8, and 12) provided an optimal tradeoff between bias and precision by demonstrating sufficient sensitivity to variations in rivalry across blocks or sessions without overfitting the data. Spline terms were specified in the design matrix using the spline facility of PROC GLIMMIX. The spline terms denoting effects of session number and block number were included in models because these factors were potential predictors of variation in rivalry durations. If so, their inclusion would likely heighten statistical power to detect the effects of phase. In addition, although a repeated-measures analysis of variance (ANOVA) testing for between-phase differences in the average session number indicated only moderate differences that were not statistically significant ($\text{mean}_{\text{follicular}} = 8.01$, $\text{mean}_{\text{mid-luteal}} = 10.08$, $\text{mean}_{\text{late-luteal}} = 6.56$), $F(2, 12) = 2.53$, $p = 0.12$, it was still important to include session as a factor to ensure that the effects of phase were not confounded by even minor between-phase differences in the distribution of sessions on either a between- or within-subjects basis. To further minimize confounding, we included in the analysis two aggregate scores as between-subjects predictors that reflected the relative number of blocks across the three phases and the average session number of those included in the analysis.

For mixed-effects models to yield valid estimates and thus valid hypothesis tests, it is important to specify a structure for the random and residual variances and covariances that accommodates serial correlation and other potential sources of nonindependence among observations (e.g., Gurka, Edwards, & Muller, 2011). Preliminary analyses comparing alternative structures using the AIC indicated that the optimal structure specified (a) an autoregressive (AR) lag-1 random effects structure to model the across-session correlations among scores of a given participant and (b) an autoregressive moving average (ARMA) (1,1) structure on the residual covariance matrix to model the within-session correlations among the scores of a given participant. The combination of these two structures specified that all the observations of a given participant were intercorrelated, with the magnitude of correlations declining as blocks within a session were

Measure	Follicular	Mid-luteal	Late-luteal
Duration of dominance epochs (s)	2.31 (1.11)	2.26 (1.34)	2.40 (1.45)
Duration of mixture epochs (s)	1.77 (0.86)	1.62 (0.75)	1.75 (0.85)
Proportion of dominance epochs that were return transitions	.15 (.15)	.13 (.12)	.13 (.13)
Switch rate per min	18.99 (6.54)	19.82 (9.05)	19.39 (6.53)
Proportion of epochs that were mixtures	.31 (.16)	.32 (.21)	.32 (.16)

Table 1. Means for duration and proportion measures. *Notes*: $N = 13$. Means are computed by first computing averages, proportions, or counts per block per session per participant, then averaging across all the blocks of a given phase per participant, and finally averaging across participants. Standard deviations across participants are noted in parentheses.

increasingly distant from one another and as sessions within a subject were increasingly distant from one another.

Restricted maximum likelihood was used to estimate all parameters. The first-order Kenward-Rogers procedure (Harville & Jeske, 1992; Kenward & Roger, 1997; for a review, see Stroup, 2013) was used to provide bias-corrected standard errors and degrees of freedom. General F statistics were used to test omnibus fixed effects and contrasts among means. Based on the results of Epperson et al. (2002) indicating higher GABA concentrations in the follicular relative to the mid- and late-luteal phases, we conducted a planned complex contrast comparing mean durations of the follicular phase and the other two phases (contrast coefficients equaled 1–0.5 and −0.5). We also conducted pairwise contrasts between each of the three phases. When such contrasts follow a significant omnibus effect for phase, they are multiplicity-corrected because this is the Fisher least significant difference approach that provides an optimal combination of control of family-wise Type I errors and power when the number of levels of a factor equals three (Levin, Serlin, & Seaman, 1994; Seaman, Levin, & Serlin, 1991). When omnibus tests were not significant, we also report below the results of pairwise comparisons, but any significant effects should be regarded as more exploratory in nature.

We assessed between-phase differences separately on dominance and mixture durations for five reasons: (a) Nearly all rivalry studies have focused on dominance durations (indeed, few studies have assessed mixtures), and dominance states have provided the primary corpus of evidence for evaluating theories of rivalry dynamics (Brascamp, Klink, & Levelt, 2015). Thus, dominance states were our main interest and the focus of our a priori hypotheses. (b) Recent work has reported a differential influence of alcohol, a proposed GABAergic agonist (see Mihic, 1999, for review), on the duration and incidence of dominance and mixtures, thus warranting their distinction (Cao, Zhuang, Kang, Hong, & King, 2016). (c) The distributions of dominance and mixture durations differed according to several parametric and nonparametric indices. For example, an LMM model that included state (dominance vs. mixture) as a factor indicated significant mean differences between dominance and mixtures, $F(1, 24) = 5.77$, $p = 0.024$, and additional type-specific random effect components. Overall, mean durations tended to be longer for dominance epochs than mixture epochs (see Table 1). (d) Average dominance and mixture durations were only weakly correlated across subjects ($r = 0.12$). (e) Dominance states were denoted by specific responses, whereas mixture states had to be estimated from the absence of responses taking into account response times for releasing one button and then pressing another button.

In addition to the LMM analyses on mean duration, generalized linear mixed models (GLMMs) were specified that assessed between-phase differences on three additional quantities: the proportion of dominance states that were return transitions (described in the Procedure section), the switch rate for dominance states (defined as the number of switches in dominance states per minute), and the proportion of mixture states in each block. GLMM models are a liberalization of the standard LMM model that can accommodate nonlinear functional forms and nonnormal distributions. The GLMM analysis on the two proportion measures (return transitions and proportion of mixture states) specified a binomial distribution for the dependent variable, a fixed-effects structure identical to that of the LMM analyses, and AR1 and ARMA(1,1) structures to model the cross- and within-session dependencies among the observations of a given participant. The analysis of the switch rate used an identical fixed effect, random effect, and residual structure but specified a Poisson distribution for the dependent variable. We directly analyzed the number of switches rather than number per second or minute because the final switch into a dominance percept in a block had to occur before the 60-s mark even if the total duration of the block was more than 60 s. Model fit diagnostics verified the absence of overdispersion on the GLMM analyses. We used a restricted pseudo-likelihood approach (Wolfinger & O'Connell, 1993) for estimation because it allowed us to model fully the random effects and residual

structure. We verified, however, that similar results were yielded when Laplace integral approximations were used to maximize the marginal likelihood and obtain estimates (for a review, see Stroup, 2013). Complex and pairwise comparisons between phases were identical to those used in the LMM analyses.

Using nonparametric approaches, we also assessed whether the three menstrual phases differed in other distributional features. Because these analyses necessitated simpler models, they were performed on summary indices that were averaged across all the epoch-level observations of a given phase for a given individual. Thus, each participant had three scores, one per phase. Such measures were generated for measures of the median, variance, skew, and kurtosis of duration values. We then used the Friedman test (Marascuilo & McSweeney, 1977) to conduct nonparametric repeated-measures analyses assessing between-phase differences on each measure. We also conducted Wilcoxon signed-rank tests comparing each pair of phases and a planned analysis that compared the follicular measures to the average of the mid- and late-luteal measures.

We also assessed whether the distributions as a whole differed across phases using the following procedure. First, for each pairing of phases (e.g., follicular vs. mid-luteal) and for each subject, we computed the Kolmogorov–Smirnov (K-S) D_{max} statistic comparing the two distributions of duration values. Across subjects, we were less interested in whether the average D_{max} values across participants were greater than 0 (indicating *some* overall difference between distributions) because that could be found even if there were no consistent pattern across subjects. To generate a more meaningful test of a systematic effect, we attached a sign (+ or −) to the D_{max} values depending on the relative location of the two distributions. For example, in a comparison between the follicular and mid-luteal distributions, if the follicular distribution was shifted to the right relative to mid-luteal, D_{max} was given a positive sign; if the follicular distribution was shifted to the left relative to mid-luteal, D_{max} was given a negative sign. We then conducted one-sample Wilcoxon tests of the null hypothesis that the average signed D_{max} value equaled 0. Rejection of this hypothesis would imply that the distributions for a given phase differed in a consistent manner across participants. In addition to pairwise comparisons between phases, we compared the D_{max} values for the follicular versus combined mid- and late-luteal phases. Conceptually similar tests comparing kernel density estimates of individual distributions on a measure of distributional overlap (e.g., Schmid & Schmidt, 2006) yielded identical results. For the sake of brevity, we report only the K-S analyses.

Results

Analyses of mean duration

The following exposition of results focuses on dominance and mixture durations for all sessions falling within the three phases of interest. In the aggregate, those totaled 61,679 individual durations compiled over all 13 participants and all 20 blocks for all sessions falling within the three phases.

Figure 2 presents dominance duration density plots for each observer, grouped and color coded by menstrual cycle phase. Several aspects of these results are noteworthy. First, these distributions conform to the signature shape characteristic of binocular rivalry durations (Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006; Fox & Herrmann, 1967; Levelt, 1965). Second, we see pronounced individual differences in alternation rate among our subject sample, as indicated by marked differences in the peaks of the histograms across participants (peaks at longer durations imply slower alternations). Such heterogeneity is also routinely seen in larger samples of participants (Carter & Pettigrew, 2003; Hancock, Gareze, Findlay, & Andrews, 2012). Of relevance for our purposes, we also see no consistent tendency for durations measured during the follicular phase to be longer than those measured during mid- or late-luteal phases (i.e., the prediction motivating this study). This tendency is borne out in the statistical results reported below. We also created duration density plots for mixture states for each participant, and those are shown in Figure 3. These, too, have the rightward skew seen for dominance durations, and some of the mixture durations are as long as dominance duration states. To our knowledge, this is the first time in the literature that mixture distributions have been plotted, so we have no basis for comparison in terms of shape. Table 1 displays means of the raw (i.e., untransformed) rivalry duration values for dominance and mixture durations for each of the three phases. To generate the values in this table, means for each block within a given session were first computed for each participant. Then, all the block means within a given session were averaged for each participant. In the final step, means for each phase were computed by averaging the session means that fell within each phase bin across participants. The means and standard deviations shown in this table indicate small differences between phases, particularly for dominance durations.

These descriptive observations were corroborated by the results of the LMM ANOVA performed on the log of the average duration of dominance epochs per block. There was no significant effect for phase, $F(2, 55.7) < 1$, $p = 0.71$, on the omnibus test, and no significant effects

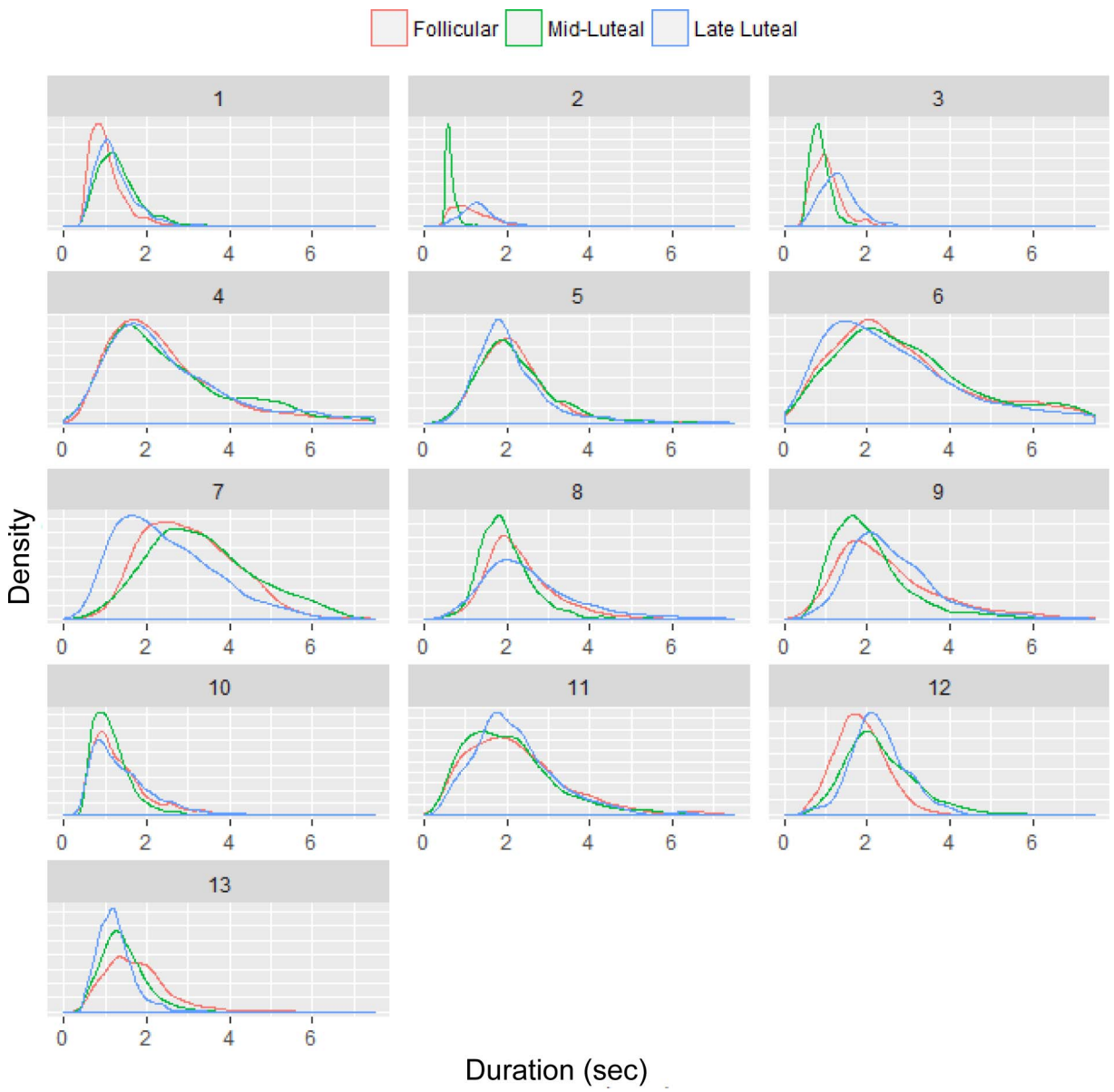

Figure 2. Distributions of dominance durations generated by kernel density plots (Wand & Jones, 1995) per phase for individual participants. Subject identification numbers are shown in the top border.

or trends were evident on the planned comparison between follicular and the average of mid-luteal and late-luteal phases, $F(1, 48.7) < 1$, $p = 0.41$, or on the pairwise comparisons among the three phases (all Fs < 1, all ps > 0.45). In contrast, there were significant effects for the spline functions of block, $F(6, 689.5) = 7.10$, $p < 0.0001$, and session, $F(6, 59.89) = 2.35$, $p = 0.04$. As shown in the top panel of Figure 4, mean durations increased across the first four blocks but then exhibited a progressive decline. The pattern of change across sessions was more complex with four shifts in direction (see the middle panel of Figure 4). Mean dominance values reached their peak nearly midway through the experiment during Session 7.

The LMM ANOVA on log of the average duration of mixture epochs per block also failed to reveal significant effects for phase, $F(2, 35.1) = 1.85$, $p = 0.19$. Although no pairwise contrasts among the three phases were significant (all ps > 0.10), there was a mild trend on the complex contrast between follicular and the average of mid-luteal and late-luteal phases, $F(1, 27.9) = 3.67$, $p = 0.07$. Contrary to phase-specific modulation of GABA and reciprocal inhibition models, the trend seemed to be due to higher estimated logged values of mixture durations for follicular (least square $\bar{X} = 0.43$) than mid-luteal ($\bar{X} = 0.35$) and late-luteal ($\bar{X} = 0.35$). This analysis also yielded significant spline effects for block, $F(6, 577) = 2.86$, $p = 0.01$, and session, $F(6, 37.1) = 3.59$, $p = 0.007$. As indicated by the top panel of

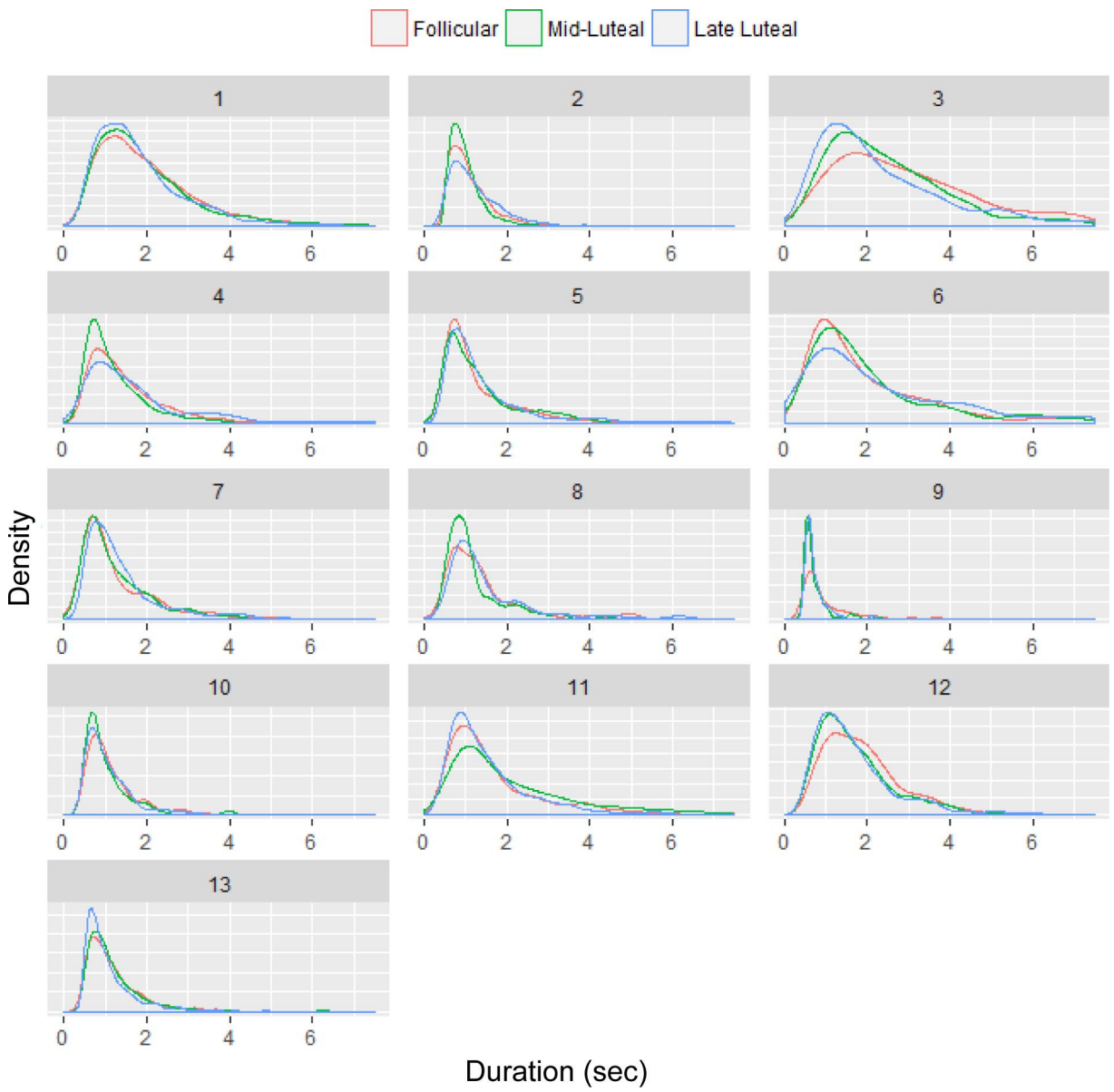

Figure 3. Distributions of mixture durations generated by kernel density plots (Wand & Jones, 1995) per phase for individual participants. Subject identification numbers are shown in the top border.

Figure 4, the pattern of mixture durations across blocks was characterized by an initial early decline followed by a late acceleration. This pattern was quite different from that shown for changes in dominance durations across blocks, and, indeed, the predicted values for the two states were negatively correlated across blocks ($r = -0.65$). The pattern of shifts in mixture durations across sessions was more complex with several shifts in slope (see middle panel of Figure 4). This pattern was also clearly different from that shown for dominance durations across sessions ($r = 0.04$).

One notable feature of the analyses of both dominance and mixture durations was the contrast between the low proportions of variance due to the effects of phase and the high proportions of variance due to the random factor of subjects. To compute the unique proportion of variance due to phase, we ran two LMM models, one in which phase was included as a factor and one in which phase was omitted. For each model, we computed the squared correlation between the predicted means (including both fixed and random effects) and actual block means. In other words, we computed an LMM analogue of R^2 (cf., Vonesh, Chinchilli, & Pu, 1996). Finally, we assessed the unique component of variance due to phase by computing the difference between the two squared *r*s. For dominance durations, the increments in squared *r*s due to phase were approximately 1.7%, and for mixture durations, the increment was approximately 1.8%. We computed the unique proportion of variance due to subjects by

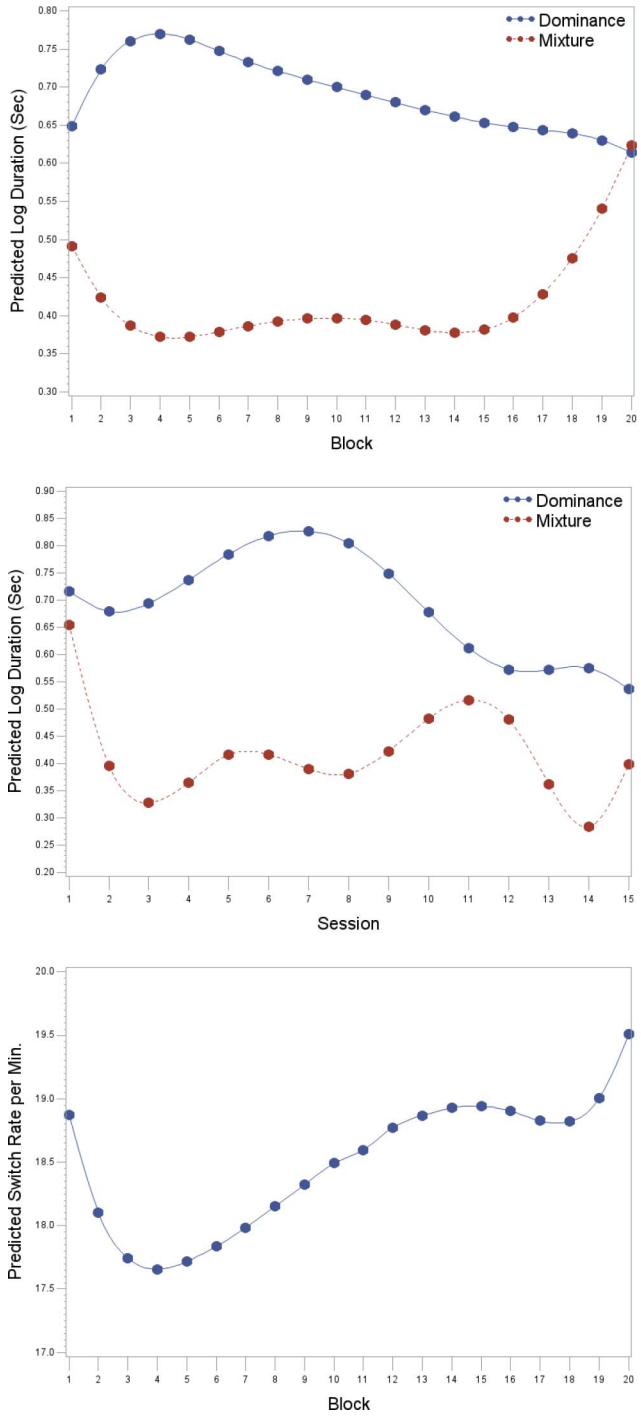

Figure 4. (Top) Log mean duration across blocks for dominance and mixture epochs. (Middle) Log mean duration across sessions for dominance and mixture epochs. (Bottom) Switch rate across blocks. Circles denote predicted values based on truncated polynomial cubic spline functions.

forming the ratio of the estimated variance for the random effect of subject to the sum of the variances of the random effect of subject and the residuals (i.e., we computed an intraclass correlation). For both dominance and mixture durations, there was a substantial proportion of variance due to subjects (dominance = 74%, mixture = 57%). In addition, likelihood ratio tests indicated that the variance parameters for subjects yielded by the LMM analyses were highly significant (both dominance and mixture ps < 0.0001).

Proportion and rate measures

The lower rows of Table 1 display the means across phases for switch rate and the proportion of return transitions and mixture epochs using the same averaging sequence described above for mean durations. Analyses of dominance switch rates indicated no significant effects of phase [omnibus, $F(2,60.71) < 1$, $p = 0.72$; planned contrast, $F(1, 53.13) < 1$, $p = 0.99$, all pairwise contrast ps $> .40$] or session, $F(6, 63.11) = 1.14$, $p = 0.35$). However, there was a highly significant effect for the block spline component, $F(6, 687.5) = 3.94$, $p < 0.001$. As shown in the bottom panel of Figure 4, after an initial early decline in switch rates from Trial 1 to Trial 4, there was a generally progressive increase through the final block. Although the absolute magnitude of the change across blocks is small, the effects were highly significant because of the strong individual differences in overall rates (an intraclass correlation indicated that 95% of the variance was due to differences among subjects) that led to extremely small standard errors when different blocks were compared. Thus, there was high power to detect even relatively moderate shifts in switch rates across blocks. There were no significant effects of phase [omnibus, $F(2, 39.59) < 1$, $p = 0.60$; planned contrast, $F(1, 38.63) < 1$, $p = 0.64$, all pairwise contrast ps > 0.40]; block, $F(6, 834.1) < 1$, $p = 0.76$; or session, $F(6, 43) = 1.31$, $p = .27$, on the proportion of return transitions. The GLMM random effects binomial model indicated no significant effects of phase on the proportion of mixtures: omnibus, $F(2, 64.47) < 1$, $p > 0.90$; planned complex comparison, $F(1, 60.84) < 1$; pairwise comparisons, all ps > 0.66. There were also no significant spline effects for block, $F(6, 726) = 1.22$, $p = 0.29$, or for session, $F(6, 63.56) = 1.65$, $p = 0.15$, on this measure.

Nonparametric tests of other summary indices

We also assessed between-phase differences in distributional features of rivalry beyond means. Table 2 shows the results of omnibus Friedman tests, pairwise Wilcoxon signed-rank tests, and a comparison between the follicular phase and the pooled mid-luteal and late-luteal phases on sample medians, variances, skew, and kurtosis. None of the tests performed on dominance duration measures indicated any significant effects or even notable trends (all ps > 0.30). Similarly, Wilcoxon

Measure	Follicular vs. mid-luteal	Follicular vs. late-luteal	Mid-luteal vs. late-luteal	Follicular vs. average of mid- and late-luteal	Omnibus test
Dominance					
Median	$T^+ = 53$, $p = 0.64$	$T^+ = 37$, $p = 0.59$	$T^+ = 38$, $p = 0.64$	$T^+ = 53$, $p = 0.64$	$\chi^2_2 = 0.62$, $p = 0.74$
Variance	$T^+ = 45$, $p = 1.00$	$T^+ = 31$, $p = 0.34$	$T^+ = 33$, $p = 0.41$	$T^+ = 37$, $p = 0.59$	$\chi^2_2 = 1.08$, $p = 0.58$
Skew	$T^+ = 49$, $p = 0.84$	$T^+ = 41$, $p = 0.79$	$T^+ = 48$, $p = 0.89$	$T^+ = 49$, $p = 0.84$	$\chi^2_2 = 0.15$, $p = 0.93$
Kurtosis	$T^+ = 43$, $p = 0.89$	$T^+ = 46$, $p = 1.00$	$T^+ = 53$, $p = 0.64$	$T^+ = 50$, $p = 0.79$	$\chi^2_2 = 0.15$, $p = 0.93$
Signed K-S D_{Max}	$T^+ = 57$, $p = 0.46$	$T^+ = 30$, $p = 0.31$	$T^+ = 34$, $p = 0.45$	$T^+ = 53$, $p = 0.64$	—
Mixture					
Median	$T^+ = 66$, $p = 0.17$	$T^+ = 51$, $p = 0.74$	$T^+ = 31$, $p = 0.56$	$T^+ = 51$, $p = 0.74$	$\chi^2_2 = 0.83$, $p = 0.66$
Variance	$T^+ = 56$, $p = 0.50$	$T^+ = 37$, $p = 0.59$	$T^+ = 32$, $p = 0.38$	$T^+ = 21$, $p = 0.09$	$\chi^2_2 = 2.00$, $p = 0.37$
Skew	$T^+ = 27$, $p = 0.22$	$T^+ = 19$, $p = 0.07$	$T^+ = 35$, $p = 0.50$	$T^+ = 12$, $p = 0.02$	$\chi^2_2 = 2.46$, $p = 0.29$
Kurtosis	$T^+ = 25$, $p = 0.17$	$T^+ = 20$, $p = 0.08$	$T^+ = 37$, $p = 0.59$	$T^+ = 16$, $p = 0.04$	$\chi^2_2 = 3.85$, $p = 0.15$
Signed K-S D_{Max}	$T^+ = 73$, $p = 0.06$	$T^+ = 50$, $p = 0.79$	$T^+ = 20$, $p = 0.08$	$T^+ = 69$, $p = 0.11$	—

Table 2. Results of Wilcoxon signed-rank and Friedman tests of between-phase differences in distributional features. *Notes*: $N = 13$. Two-tailed exact Wilcoxon signed-rank tests were used for pairwise comparisons. T+ equals the sum of the ranked differences with positive signs. The Friedman test was used to conduct an omnibus hypothesis of equality of ranks across phases. Values of the large-sample chi-square approximation to the Friedman statistic are displayed. Conclusions about statistical significance were identical when observed test statistics were compared with exact critical values of the Friedman statistic. K-S = Kolmogorov-Smirnov. For the K-S measure, Wilcoxon signed-rank tests were performed on the signed D_{Max} values that were computed for each pairing of phases for each participant.

signed-rank tests performed on the signed K-S D_{Max} values failed to indicate any systematic differences between the phases (all ps > 0.30; see Table 2). Friedman and Wilcoxon tests of between-phase differences in mixtures and Wilcoxon tests of signed K-S D_{Max} values indicated only two significant effects, between the follicular phase and the average of the mid-luteal and late-luteal phases on skew ($p = 0.02$) and kurtosis ($p = 0.04$). These results indicate that, relative to the follicular distribution, the averaged mid-luteal and late-luteal distributions were slightly more skewed to the right and more peaked. These results should, however, be regarded with caution given the number of statistical tests performed in the distributional analyses. Forty-eight statistical tests are reported in Table 2, and had multiplicity corrections been imposed (e.g., via Bonferroni or step-down Bonferroni procedures), the results of these two contrasts would not have been statistically significant.

Bayes factors for effect size intervals

Across dependent measures, the analyses summarized above indicate no significant effects for phase on dominance epochs, which were the focus of our primary hypotheses. Reliance on the conventional null hypothesis testing framework, however, has limitations. Null findings can occur because of limited power. In turn, a major determinant of power is sample size, and, in the present context, although there are a large number of observations per participant, the total number of participants is 13. Two related limitations of the reliance on null hypothesis tests are that (a) they fail to indicate the relative strength of the evidence for and against null and alternative hypotheses, and (b) they fail to indicate the magnitude of effects or the relative likelihood that effect sizes are within a given range versus outside that range. Although the proportion of variance measures reported above that were based on the LMM analyses indicated that the effects of phase on rivalry dynamics were quite small, particularly in the case of dominance durations, we felt compelled to include additional assessments that went beyond the traditional null hypothesis testing framework and addressed these issues.

As an alternative approach, we computed Bayes factors (BFs; Kass & Raftery, 1995) for the dominance duration data. BFs allow us to estimate the relative strength of the evidence for and against a given hypothesis and are analogous to likelihood ratios. Because of the computational complexity involved in quantifying BFs, we aggregated the log-transformed data for each subject across the blocks and sessions of a given phase and computed the mean dominance and mixture durations per subject per phase. We then used the framework of between-phase pairwise t tests to compute measures of effect size and BFs using the framework outlined by Morey, Rouder, and colleagues (Morey & Rouder, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009) and its software implementation in the R package BayesFactor (Morey & Rouder, 2015). We computed BFs for interval-level hypotheses that specified ranges for the value of Cohen's (1988) effect size measure δ, defined as $\delta = (\mu_1$

	Comparison		
Measure	Follicular vs. mid-luteal	Follicular vs. late-luteal	Follicular vs. average of mid- and late-luteal
Cohen's *d*	0.24	−0.07	0.10
Bayes factors			
H_0: $-0.10 < \delta < 0.10$	3.09	4.47	4.29
H_0: $-0.20 < \delta < 0.20$	3.83	6.14	5.79
H_0: $\delta = 0$	2.62	3.48	3.37

Table 3. Bayes factors for between-phase pairwise *t* tests on dominance durations. *Notes*: $N = 13$. Cohen's *d* is computed as $(\bar{x}_1 - \bar{x}_2)/\hat{s}_{dif}$, where $\hat{s}_{dif}$ = the standard deviation of the within-subject differences between phases. Because we computed Bayes factors as the ratio of the likelihoods for the null hypothesis relative to the alternative hypothesis, values greater than 1 indicate support for the null hypotheses.

$-\mu_2)/\sigma$, where σ = the population standard deviation of the difference between paired observations. The likelihood that an effect size measure equals exactly 0 or any given point value is exceeding low. For that reason, we computed two different interval-level BFs that yielded more meaningful and realistic assessments of the magnitude of effects. The corresponding null hypotheses specified that the range of the effect sizes for between-phase effects was either "small" or "very small." Specifically, we computed BFs for two separate null hypotheses: H_0: $-0.20 < \delta < 0.20$, with the alternative hypothesis that $\delta < -0.20$ or $\delta > 0.20$, and, H_0: $-0.10 < \delta < 0.10$, with the alternative hypothesis that $\delta < -0.10$ or $\delta > 0.10$. These values were chosen based on the conventional rule of thumb (Cohen, 1988) that $\delta = \pm 0.20$ indicates a small effect size. We computed BFs for these ranges specifying that δ is distributed as a Cauchy random variable with scale factor $= 2\sqrt{2}$. The Jeffreys prior (Jeffreys, 1961) was specified for the distribution of the variances. Recall that our strongest original predictions concerned dominance durations and focused on the difference between the follicular phase and the mid-luteal and late-luteal phases. Accordingly, for the dominance data, we computed BFs for the pairwise contrasts between follicular and mid-luteal and follicular and late-luteal phases and for a pairwise contrast that compared the follicular phase to the average of mid- and late-luteal phases (computed for each participant).

Table 3 presents the Bayes factor results for dominance durations. Because we computed our BFs as the ratio of the likelihoods for the null hypothesis relative to the alternative (see, e.g., Jarosz & Wiley, 2014), values greater than 1 indicate relative support for the null hypotheses that $\delta = 0$, that $-0.10 < \delta < 0.10$, and that $-0.20 < \delta < 0.20$. Additionally shown are the sample estimates of δ, computed as $d = (\bar{X}_1 - \bar{X}_2)/s_d$, where s_d = the sample standard deviation of the within-subject difference scores computed between a pair of phases. Several features of the results are notable. First, the sample estimates of δ (= d) are small for all comparisons (see Table 3). Some methodologists prefer to express d values for within-subjects data as the ratio of the differences between means to the average of the within-condition standard deviations (i.e., there is no reduction in the computed standard deviation because of correlated observations characteristic of within-subjects data; for a discussion, see Dunlap, Cortina, Vaslow, & Burke, 1996). In the present context, if d's were expressed in that manner, they would be approximately half the size shown in Table 3 and thus quite small on the whole. Second, all BFs computed were greater than 1 and thus in the direction supporting the null hypothesis given the way that we computed BFs. Although criteria for BF evaluations vary somewhat (e.g., Jeffreys, 1961; Kass & Raftery 1995), Raftery (1995) proposed that BFs between 1 and 3 offer weak support for a target hypothesis and that BFs between 3 and 10 offered what he termed "positive" support. From this perspective, the obtained BFs clearly offer positive support for the null hypothesis that the effect size is within a small range ($-0.20 < \delta < 0.20$) and, on the whole, offer positive support for the null hypothesis that the effect size is very small ($-0.10 < \delta < 0.10$), although the BF value for the follicular versus mid-luteal comparison is quite close to threshold. Table 3 also shows BFs for the point null hypothesis that the population effect size is 0. These values indicate a degree of support that is, on average, on the borderline between weak and positive support. We should reiterate, though, our belief that the point null suffers in realism and applicability relative to interval null hypotheses.

Discussion

Based on the link between individual variability in occipital GABA concentrations and binocular rivalry (van Loon et al., 2013), and fluctuations of GABA concentrations in occipital cortex associated with the healthy menstrual cycle (Epperson et al., 2002), we tested the extent to which rivalry dynamics fluctuated with the phases of a healthy menstrual cycle. According to reciprocal inhibition models of rivalry and the higher

concentrations of the main inhibitory transmitter in the brain, GABA, during the follicular phase compared with the luteal phase in the menstrual cycle, we expected longer periods of perceptual dominance and slower alternations over time. Contrary to this prediction, however, our results disclosed no reliable change across menstrual phases in the temporal dynamics of dominance percepts, despite the fact that we used several different analytic approaches (LMM, GLMM, nonparametric) and assessed several different types of dependent measures (e.g., means as well as other distributional features of dominance and mixtures, proportion of return transitions). This conclusion holds for both omnibus and pairwise comparisons for dominance durations and even for those analyses that were not multiplicity corrected. In addition, BFs indicated that it is likely that differences on dominance durations between the follicular and the mid- and late-luteal phases are, at best, small in magnitude. We also failed to find between-phase differences on mixture durations except for two effects on skew and kurtosis that would not have survived multiplicity corrections.

From the outset, we want to stress that our findings do not contradict van Loon et al.'s (2013) finding that GABA levels in the occipital cortex covary with individual differences in rivalry rate. We, like they, found reliable differences in rivalry rate (via percept durations), but we did not measure resting GABA concentrations in our study group. Before considering possible implications of our findings, we first want to dispel four possible objections to our study and its failure to find a relationship between menstrual phase and rivalry dynamics.

First, one might argue that the displays or the task we used were inappropriately designed for reliably assessing rivalry dynamics. We find this highly unlikely, for essentially identical methods have been used successfully in our lab for other purposes. Moreover, the distributions of dominance durations and the reliable individual differences in rivalry rates measured within our sample of participants replicate the pattern of results found in many earlier studies, including those explicitly designed for studying trait-level variability across individuals (Carter et al., 2005; Miller et al., 2010; Nagamine, Yoshino, Miyazaki, Takahashi, & Nomura, 2008; Shannon, Patrick, Jiang, Bernat, & He, 2011; van Loon et al., 2013).

Second, results from a previous study led us to wonder whether the stability of binocular eye alignment in our participants might show subtle but systematic differences during the follicular and luteal phases of the menstrual cycle, which, in turn, can affect our measures of rivalry dynamics. Specifically, in one experiment of the study by van Loon et al. (2013), participants (all males) were administered lorazepam, a benzodiazepine that putatively potentiates $GABA_A$ receptors at the dose levels given to those participants. Their plan was to learn whether potentiated GABA would slow rivalry alternations, but unfortunately, it proved impossible to measure binocular rivalry alternations because this short-acting pharmacological agent disrupted stable binocular eye alignment, a prerequisite for measuring binocular rivalry. This observation led us to wonder whether there was a tendency for eye position to vary more during the follicular phase when GABA is purportedly the highest (as it was in the van Loon et al. study because of lorazepam).

To examine that possibility, we used results from the dichoptic alignment procedure performed on each participant immediately before each test session to index the stability of eye position over sessions. Recall that each person used a method of adjustment procedure to ensure that the monocular fusion frames were appropriately aligned for the two eyes so as to produce the impression of a stable, single visual object. This adjustment was performed three times in succession, with the x/y offsets of the dichoptic targets jittered to new positions before each adjustment. The prediction here is that participants would have more difficulty maintaining stable eye alignment during the follicular phase, where GABA was reportedly higher (Epperson et al., 2002). If this were the case, we would expect greater variability across repeated alignment trials in the follicular phase compared with the mid- and late-luteal phases of the menstrual cycle. That was not what we observed, however, as indicated by the eye position result in Figure 5. Thus, we are disinclined to attribute our null results for rivalry dynamics over menstrual phase to eye instability. Moreover, we wonder whether the failure of van Loon et al.'s (2013) participants to maintain eye alignment when given lorazepam might be attributable to disruption in the oculomotor system—double vision is one of the possible side effects listed for this drug.

Third, one might argue that changes in occipital GABA concentrations throughout the menstrual phase are not sufficient in magnitude to evoke measurable changes in binocular rivalry dynamics. We find this unlikely based on previous work. From the work of Lunghi and colleagues (2015), we know that short-term patching of one eye temporarily reduces GABA concentrations by approximately 8% and, at the same time, slows binocular rivalry alternations by approximate 20%. In the study that motivated our experiment, Epperson et al. (2002) reported an approximately 30% drop in occipital GABA concentration from the follicular to luteal phase during the menstrual cycle of healthy females, a drop surely sufficient to affect rivalry dynamics based on rivalry results from the Lunghi et al. (2015) study. It is worth noting that a subsequent study by Epperson et al. (2005) found no fluctuations in GABA over the phases of the menstrual cycle in

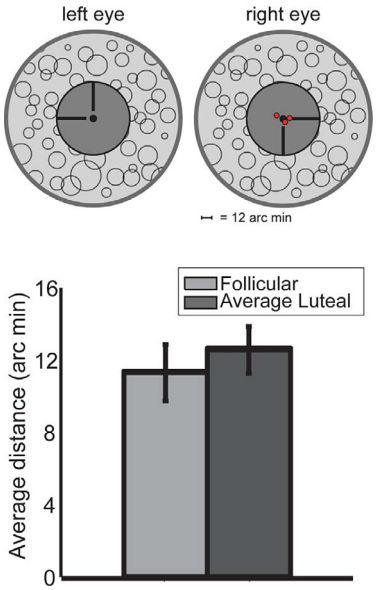

Figure 5. Eye alignment procedure and results. (Top) Prior to each session, the observer performed three successive alignment estimates, adjusting the x/y position of the right-eye image until it appeared stably superimposed on the left-eye image (i.e., a single, stable percept in which the pairs of vertical and horizontal nonius markers were aligned; note that nonius lines were included during the adjustment phase only, not during rivalry tracking trials). To estimate the variability over these three successive adjustments, we performed an additive translation in the X/Y positioning in both eyes so that the left eye position was centered on (0,0). The translated $X_{translation}$/$Y_{translation}$ position in the right eye thus reflected the relative positioning between the eyes to achieve a stable fusion. We calculated the absolute distance of each $X_{translation}$/$Y_{translation}$ for each participant and session-specific right fixation, determined by the centroid of the three right-eye alignment trials. The three red dots mark the relative right eye placement in three alignment trials for one participant whose results coincide with the group average. The scale bar underneath the right-eye display signifies 12 arc min, the average deviation among the three settings averaged over observers and sessions. (Bottom) The difference between the averaged distance in the follicular phase and the average of the luteal phases was not statistically significant, $t(12) = 1.05$, $p = 0.31$. Error bars represent *SEM*.

healthy females who were chronic smokers (i.e., smoked 20–40 cigarettes a day for at least one year). In the prescreening portion of our study, all of our participants reported being nonsmokers, so our failure to find rivalry fluctuations during the menstrual cycle cannot be chalked up to this nicotine-related impact on GABA.

Fourth, one might question whether our study had sufficient power to detect between-phase differences on rivalry duration measures. To address this issue more directly, we conducted power analyses for dominance durations (the major focus of our predictions) using a procedure outlined by Stroup (2002, 2013) for LMM designs and using an additional procedure for repeated-measures ANOVA designs (Maxwell & Delaney, 2004). We performed several different power calculations within the LMM framework because there are alternative approaches to calculating effect sizes in our complex LMM models. These calculations indicated that we had sufficient power to detect pairwise differences between phases that corresponded to a medium effect size (Cohen's $\delta = .50$). Power estimates to detect $\delta = .50$ varied from .63 to .90. These values are not surprising because of the large number of observations per participant and because of the stable individual differences observed across phases. Recall that individual differences among subjects accounted for a high proportion of the total variability in the data. In any repeated measures design (whether analyzed via the LMM approach or more traditional methods), the magnitude of the correlations among the levels of a within-subjects factor is a major determinant of power (e.g., Maxwell & Delaney, 2004). As a result, even studies with fewer than 20 participants can have adequate to substantial power, which we think is the case for our study.

Moreover, we also assessed how many participants would be required to reveal a significant main effect for phase on dominance durations had we observed the precise means, standard deviations, and effect sizes used in the present study and used the same statistical model for analyses. We used two different approaches to compute power (e.g., Maxwell & Delaney, 2004; Stroup 2002, 2013), and both indicated that minimal sample sizes equal to at least 100 would be necessary to have adequate power. Although power analyses based on extant data have clear limitations, these computations suggest such a small effect that doubling or even tripling the sample size would be highly unlikely to yield significant between-phase differences in dominance durations.

In addition to the power analyses, three additional sources of evidence indicate that our results are more likely due to small effects rather than insufficient power to detect nontrivial effects: (a) the very small proportions of variance on dominance and mixture durations accounted for by phase, (b) BFs indicating that small effect sizes were more likely than not, and (c) the fact that we clearly had sufficient power to detect significant—and, in some cases, highly significant—effects of block and session on several measures.

Having rejected these alternative, methodologically based arguments for our findings, we are instead led to believe that our results reveal a more complex relationship among GABA, hormonal fluctuations in the menstrual cycle, and the inhibitory processing

involved in binocular rivalry. For one thing, Epperson and colleagues measured GABA levels only in the occipital cortex. Yet we know that some other cortical regions exhibit patterns of GABA fluctuation during the menstrual cycle that differ from those seen in occipital cortex. Prefrontal regions, for example, are reported to exhibit larger concentrations of GABA during ovulation, compared with follicular and luteal phases (De Bondt, De Belder, Vanhevel, Jacquemyn, & Parizel, 2015), but in the anterior cingulate cortex, GABA remains invariant during the menstrual cycle (Harada, Kubo, Nose, Nishitani, & Matsuda, 2011). Thus, when using binocular rivalry as an index of perception-related inhibitory activity, one should take into account GABA fluctuations in cortical regions other than the occipital cortex. After all, cortical responses in frontoparietal regions of the brain have also been correlated with rivalry alternations (Frässle, Sommer, Jansen, Naber, & Einhäuser, 2014; Knapen, Brascamp, Pearson, van Ee, & Blake, 2011; Lumer, Friston, & Rees, 1998).

Another factor is the complex dynamic between GABA and menstrual hormones and their metabolites. The multifaceted relationship between GABA and menstrual hormones is made evident in the reported finding that despite differences in premenstrual symptomology and occipital GABA concentrations between females with a healthy menstrual cycle and those with premenstrual dysphoric disorder (PMDD), ovarian hormone levels are not significantly different between groups (Bäckström et al., 2003). Moreover, despite fluctuations of group average concentrations of ovarian hormones and GABA across the menstrual cycle, a correlation between their respective levels or with the severity of premenstrual symptomology across individuals remains elusive (De Bondt et al., 2015). Instead, some have hypothesized that abnormal symptomology associated with PMDD is not the consequence of the overall concentration but rather some abnormal sensitivity to hormones and their metabolites within the central nervous system (Bäckström et al., 2014; Barth, Villringer, & Sacher, 2015; Huo et al., 2007). One mechanism by which sensitivity can vary is at the receptor level. For instance, progesterone and its metabolites facilitate GABAergic transmission by increasing receptor affinity for GABA, whose binding opens chloride gated channels and increases neural inhibition (Deligiannidis et al., 2013; Lan & Gee, 1994; Rupprecht, 1997; van Wingen et al., 2008). Considering the interaction between progesterone and GABA in isolation, the increased presence of progesterone in the luteal phase could be expected to increase neural inhibition. In the context of reciprocal inhibition in binocular rivalry, the luteal phases should then have longer durations of exclusive dominance and shorter mixtures. This prediction is in direct opposition to the predictions made when considering only lowered occipital GABA concentration during the luteal phase compared with the follicular phase. When considering the interaction between progesterone and GABAergic receptors and the changes in GABA in the menstrual cycle together, we are led to wonder whether GABA concentration is down-regulated during the luteal phase to maintain a level of homeostasis that counteracts greater receptor sensitivity in the presence of progesterone. The balancing of GABA in the presence of progesterone would therefore maintain the level of inhibition within an individual throughout the menstrual cycle and thereby explain the invariance of binocular rivalry dynamics in each menstrual phase. The presence of a neurochemical regulatory process that acts along the temporal scale of the menstrual cycle could also explain the present contradictory findings and the reported relationship between perceptual alternations and transient manipulations of GABA, either through short-acting pharmacological manipulations using, for example, lorazapam (van Loon et al., 2013) or eye patching (Lunghi et al., 2015). Essentially, rapid inhibitory state changes in GABA might still alter perceptual alternation in the absence of any change in an opposing excitatory homeostat such as progesterone.

Finally, the relationship between the menstrual cycle and binocular rivalry is also complicated when considering that ovarian hormones can act on multiple receptor types other than those reactive to GABA (e.g., Gulinello, Gong, Li, & Smith, 2001; Sumner & Fink, 1998; Weiland, 1992; Woolley, Weiland, McEwan, & Schwartzkroin, 1997). In sum, during the menstrual cycle, there is a delicate balance of hormones that have diverse interactions with neurotransmitter receptors throughout various brain regions. Thus, it is important to adopt a broad and systemic perspective when attempting to link GABA with perceptual phenomena such as binocular rivalry.

Keywords: binocular rivalry, menstrual phase, GABA, dominant and mixed percepts

Acknowledgments

Kevin Dieter helped with programing software for executing the experiment and analyzing the data. This research was supported by National Institutes of Health (NIH) Research Grant R21EY022752, by NIH Core Grant EY022751, and by Clinical and Translational Science Award No. UL1TR000445 from the National Center for Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent official views of the

National Center for Advancing Translational Sciences or the National Institutes of Health.

Commercial relationships: none.
Corresponding author: Jocelyn L. Sy.
Email: jocelyn.l.sy@vanderbilt.edu.
Address: Department of Psychology, Vanderbilt University, Nashville, TN, USA.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723, doi:10.1109/TAC.1974.1100705.

Akk, G., Shu, H.-J., Wang, C., Steinbach, J. H., Zorumski, C. F., Covey, D. F., & Mennerick, S. (2005). Neurosteroid access to the $GABA_A$ receptor. *Journal of Neuroscience*, *25*, 11605–11613.

Bäckström, T., Andreen, L., Birzniece, V., Björn, I., Johansson, I. M., Nordenstam-Haghjo, M., … Zhu, D. (2003). The role of hormones and hormonal treatments in premenstrual syndrome. *CNS Drugs*, *17*, 325–342.

Bäckström, T., Haage, D., Lofgren, M., Johansson, I. M., Nyberg, S., Ossewaarde, L., … van Wingen, G. (2014). Allopregnanolone and mood disorders. *Progress in Neurobiology*, *113*, 88–94.

Barth, C., Villringer, A., & Sacher, J. (2015). Sex hormones affect neurotransmitters and shape the adult female brain during hormonal transition periods. *Frontiers in Human Neuroscience, 9*, 37, doi:10.3389/fnins.2016.00276.

Blake, R., & Logothetis, N. (2002). Visual competition. *Nature Reviews: Neuroscience*, *3*, 13–21, doi:http://dx.doi.org/10.1038/nrn701.

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

Brascamp, J. W., Klink, P. C., & Levelt, W. J. (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Research*, *109*, 20–37.

Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R. H. A. H., & van den Berg, A. V. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *Journal of Vision*, *6*(11):8, 1244–1256, doi:10.1167/6.11.8. [PubMed] [Article]

Cao, D., Zhuang, X., Kang, P., Hong, S. W., & King, A. C. (2016). Acute alcohol drinking promotes piecemeal percepts during binocular rivalry. *Frontiers in Psychology*, *7*, 489, doi:10.3389/fpsyg.2016.00489.

Carter, O. L., & Pettigrew, J. D. (2003). A common oscillator for perceptual rivalries? *Perception*, *32*, 295–305.

Carter, O. L., Pettigrew, J. D., Hasler, F., Wallis, G. M., Liu, G. B., Hell, D., & Vollenweider, F. X. (2005). Modulating the rate and rhythmicity of perceptual rivalry alternations with the mixed 5-HT2A and 5-HT1A agonist psilocybin. *Neuropsychopharmacology*, *30*, 1154–1162, doi:http://dx.doi.org/10.1038/sj.npp.1300621.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Bondt, T., De Belder, F., Vanhevel, F., Jacquemyn, Y., & Parizel, P. M. (2015). Prefrontal GABA concentration changes in women—Influence of menstrual phase, hormonal contraceptive use, and correlation with premenstrual symptoms. *Brain Research*, *1597*, 129–138.

Deligiannidis, K. M., Sikoglu, E. M., Shaffer, S. A., Frederick, B., Svenson, A. E., Kopoyan, A., … Moore, C. M. (2013). GABAergic neuroactive steroids and resting-state functional connectivity in postpartum depression: A preliminary study. *Journal of Psychiatric Research*, *47*, 816–828, doi:10.1016/j.jpsychires.2013.02.010.

Dieter, K., & Blake, R. (2015). Sensory eye dominance varies within the visual field. *Journal of Vision*, *15*(12):268, doi:10.1167/15.12.268. [Abstract]

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170–177.

Epperson, C. N., Haga, K., Mason, G. F., Sellers, E., Gueorguieva, R., Zhang, W., … Krystal, J. H. (2002). Cortical γ-aminobutyric acid levels across the menstrual cycle in healthy women and those with premenstrual dysphoric disorder. *Archives of General Psychiatry*, *59*, 851–858.

Epperson, C. N., O'Malley, S., Czarkowski, K. M., Gueorguieva, R., Jatlow, P., Sanacora, G., … Mason, G. F. (2005). Sex, GABA, and nicotine: The impact of smoking on cortical GABA levels across the menstrual cycle as measured with proton magnetic resonance spectroscopy. *Biological Psychiatry*, *57*, 44–48.

Fox, R., & Herrmann, J. (1967). Stochastic properties of binocular rivalry alternations. *Perception & Psychophysics*, *2*, 432–436.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *Journal of Neuroscience*, *34*, 1738–

1747, doi:http://dx.doi.org/10.1523/JNEUROSCI.4403-13.2014.

Freyberg, J., Robertson, C. E., & Baron-Cohen, S. (2015). Reduced perceptual exclusivity during object and grating rivalry in autism. *Journal of Vision*, *15*(13):11, 1–12, doi:10.1167/15.13.11. [PubMed] [Article]

Gulinello, M., Gong, Q. H., Li, X., & Smith, S. S. (2001). Short-term exposure to a neuroactive steroid increases alpha4 GABA(A) receptor subunit levels in association with increased anxiety in the female rat. *Brain Research*, *910*, 55–66, doi:10.1016/S0006-8993(01)02565-3.

Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, *30*, 2696–2707.

Hampson, E. (1990). Variations in sex-related cognitive-abilities across the menstrual-cycle. *Brain Cognition*, *14*, 26–43, doi:10.1016/0278-2626(90)90058-V.

Hancock, S., Gareze, L., Findlay, J. M., & Andrews, T. J. (2012). Temporal patterns of saccadic eye movements predict individual variation in alternation rate during binocular rivalry. *i-Perception*, *3*, 88–96.

Harada, M., Kubo, H., Nose, A., Nishitani, H., & Matsuda, T. (2011). Measurement of variation in the human cerebral GABA level by in vivo MEGA-Editing proton MR spectroscopy using a clinical 3T instrument and its dependence on brain region and the female menstrual cycle. *Human Brain Mapping*, *32*, 828–833.

Harville, D. A., & Jeske, D.R. 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, *87*, 724–731.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*, 687–701, doi:http://dx.doi.org/10.1016/j.cognition.2008.05.010.

Huo, L., Straub, R. E., Roca, C., Schmidt, P. J., Shi, K., Vakkalanka, R., ... Rubinow, D. R. (2007). Risk for premenstrual dysphoric disorder is associated with genetic variation in ESR1, the estrogen receptor alpha gene. *Biological Psychiatry*, *62*, 925–933, doi:10.1016/j.biopsych.2006.12.019.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, *7*, 2–9.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kang, M. S., & Blake, R. (2010). What causes alternations in dominance during binocular rivalry? *Attention, Perception, & Psychophysics*, *72*, 179–186, doi:http://dx.doi.org/10.3758/APP.72.1.179.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Keele, L. J. (2008). *Semiparametric regression for the social sciences*. New York: Wiley.

Kenward, M. G., & Roger, M. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983–987.

Klink, P. C., Brascamp, J. W., Blake, R., & van Wezel, R. J. A. (2010). Experience-driven plasticity in binocular vision. *Current Biology*, *20*, 1464–1469, doi:http://dx.doi.org/10.1016/j.cub.2010.06.057.

Knapen, T., Brascamp, J., Pearson, J., van Ee, R., & Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *Journal of Neuroscience*, *31*, 10293–10301, doi:http://dx.doi.org/10.1523/JNEUROSCI.1727-11.2011.

Lan, N. C., & Gee, K. W. (1994). Neuroactive steroid actions at the GABAA receptor. *Hormones and Behavior*, *28*, 537–544.

Lehky, S. R. (1988). An astable mulitvibrator model of binocular rivalry. *Perception*, *17*, 215–228.

Lehky, S. R., & Blake, R. (1991). Organization of binocular pathways: Modeling and data related to rivalry. *Neural Computation*, *3*, 44–53.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, *3*, 254–264, doi:http://dx.doi.org/10.1016/S1364-6613(99)01332-7.

Levelt, W. J. M. (1965). On binocular rivalry. Soesterberg, The Netherlands: Institute for Perception RVO-TNO.

Levin, J. R., Serlin, R. C., & Seaman, M. A. (1994). A controlled, powerful multiple-comparison strategy for several situations. *Psychological Bulletin*, *115*, 153–159.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*, 1930–1934, doi:http://dx.doi.org/10.1126/science.280.5371.1930.

Lunghi, C., Emir, U. E., Morrone, M. C., & Bridge, H. (2015). Short-term monocular deprivation alters GABA in the adult human visual cortex. *Current Biology*, *25*, 1496–1501, doi:http://dx.doi.org/10.1016/j.cub.2015.04.021.

Maki, P. M., Rich, J. B., & Rosenbaum, R. S. (2002). Implicit memory varies across the menstrual cycle: estrogen effects in young women. *Neuropsychologia*, *40*, 518–529, doi:10.1016/s0028-3932(01)00126-9.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.

Maxwell, S. E., & Delaney, H.D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). New York: Taylor & Francis.

Mihic, S. J. (1999). Acute effects of ethanol on GABAA and glycine receptor function. *Neurochemistry International*, *35*, 115–123.

Miller, S. M., Gynther, B. D., Heslop, K. R., Liu, G. B., Mitchell, P. B., Ngo, T. T., ... Geffen, L. B. (2003). Slow binocular rivalry in bipolar disorder. *Psychological Medicine*, *33*, 683–692.

Miller, S. M., Hansell, N. K., Ngo, T. T., Liu, G. B., Pettigrew, J. D., Martin, N. G., & Wright, M. J. (2010). Genetic contribution to individual variation in binocular rivalry rate. *Proceedings of the National Academy of Sciences, USA*, *107*, 2664–2668, doi:http://dx.doi.org/10.1073/pnas.0912149107.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*, 406–419.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. R package version 0.9.12-2. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nagamine, M., Yoshino, A., Miyazaki, M., Takahashi, Y., & Nomura, S. (2008). Effects of selective 5-HT1A agonist tandospirone on the rate and rhythmicity of binocular rivalry. *Psychopharmacology*, *198*, 279–286.

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

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–196). Cambridge, MA: Blackwell.

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., & Baker, C. I. (2013). Slower rate of binocular rivalry in autism. *Journal of Neuroscience*, *33*, 16983–16991, doi:http://dx.doi.org/10.1523/JNEUROSCI.0448-13.2013.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting null hypotheses. *Psychonomic Bulletin & Review*, *16*, 225–237.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge, UK: Cambridge University Press.

Rupprecht, R. (1997). The neuropsychopharmacological potential of neuroactive steroids. *Journal of Psychiatric Research*, *31*, 297–314, doi:10.1016/S0022-3956(96)00060-X.

Said, C. P., Egan, R. D., Minshew, N. J., Behrmann, M., & Heeger, D. J. (2013). Normal binocular rivalry in autism: Implications for the excitation/inhibition imbalance hypothesis. *Vision Research*, *77*, 59–66.

Said, C. P., & Heeger, D. J. (2013). A model of binocular rivalry and cross-orientation suppression. *PLoS Computational Biology*, *9*, e1002991, doi:10.1371/journal.pcbi.1002991.

Schmid, F., & Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping—Theory and empirical application. *Computational Statistics and Data Analysis*, *50*, 1583–1596.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin*, *110*, 577–586.

Shannon, R. W., Patrick, C. J., Jiang, Y., Bernat, E., & He, S. (2011). Genes contribute to the switching dynamics of bistable perception. *Journal of Vision*, *11*(3):8, 1–7, doi:10.1167/11.3.8. [PubMed] [Article]

Smith, S. S., Waterhouse, B. D., Chapin, J. K., & Woodward, D. J. (1987). Progesterone alters GABA and glutamate responsiveness: A possible mechanism for its anxiolytic action. *Brain Research*, *400*, 353–359, doi:10.1016/0006-8993(87)90634-2.

Smith, S. S., Waterhouse, B. D., & Woodward, D. J. (1988). Locally applied estrogens potentiate glutamate-evoked excitation of cerebellar Purkinje cells. *Brain Research*, *475*, 272–282.

Stroup, W. W. (2002). Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agriculture, Biological, and Environmental Statistics*, *70*, 491–511.

Stroup, W. W. (2013). *Generalized linear mixed models: Modern concepts, methods and applications*. Boca Raton, FL: CRC Press.

Sumner, B. E., & Fink, G. (1998). Testosterone as well as estrogen increases serotonin2A receptor mRNA and binding site densities in the male rat brain.

Molecular Brain Research, *59*, 205–214, doi:10.1016/S0169-328X(98)00148-X.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*, 502–511.

Tononi, G., & Edelman, G. M. (2000). Schizophrenia and the mechanisms of conscious integration. *Brain Research Reviews*, *31*, 391–400.

van Loon, A. M., Knapen, T., Scholte, H. S., St. John-Saaltink, E., Donner, T. H., & Lamme, V. A. F. (2013). GABA shapes the dynamics of bistable perception. *Current Biology*, *23*, 823–827, doi: http://dx.doi.org/10.1016/j.cub.2013.03.067.

van Wingen, G. A., van Broekhoven, F., Verkes, R. J., Petersson, K. M., Bäckström, T., Buitelaar, J. K., & Fernández, G. (2008). Progesterone selectively increases amygdala reactivity in women. *Molecular Psychiatry*, *13*, 325–333, doi:10.1038/sj.mp.4002030.

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness of fit in generalized nonlinear mixed-effects models. *Biometrics*, *52*, 572–587.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.

Weiland, N. G. (1992). Glutamic acid decarboxylase messenger ribonucleic acid is regulated by estradiol and progesterone in the hippocampus. *Endocrinology*, *131*, 2697–2702.

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences, USA*, *100*, 14499–14503, doi:http://dx.doi.org/10.1073/pnas.2333622100.

Wolfinger, R. D., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computing and Simulation*, *4*, 233–243.

Woolley, C. S., Weiland, N. G., McEwen, B. S., & Schwartzkroin, P. A. (1997). Estradiol increases the sensitivity of hippocampal CA1 pyramidal cells to NMDA receptor-mediated synaptic input: Correlation with dendritic spine density. *Journal of Neuroscience*, *17*, 1848–1859.