The spatiotemporal dynamics of binocular rivalry: evidence for increased top-down flow prior to a perceptual switch

Nadine Dijkstra*,†, Marieke E. van de Nieuwenhuijzen and
Marcel A. J. van Gerven†

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands.

*Correspondence address. Radboud University, Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. E-mail: n.dijkstra@donders.ru.nl
†Nadine Dijkstra, http://orcid.org/0000-0003-1423-9277; Marcel A. J. van Gerven, http://orcid.org/0000-0002-2206-9098

Abstract

According to most theories, perceptual switching during binocular rivalry is caused by competition between the neural representations of the two input images. It remains unclear whether competition is resolved already at the early stages of visual processing and that information about the dominant percept is then fed forward to more high-level areas or whether competition is first resolved in high-level areas and then fed back to lower levels. This study aimed to dissociate between these theories by investigating the direction of information flow prior to a perceptual switch, using Granger causality on classifier output originating from occipital, temporal, parietal and frontal regions of interest. The results point toward increased top-down information flow between temporal and occipital areas before a switch in dominance. These findings do not support a low-level account of binocular rivalry but are in line with high-level and hybrid explanations.

Key words: perception; theories and models; contents of consciousness

To investigate the neural correlates of visual awareness, researchers have made ample use of bi-stable stimuli. These stimuli make it possible to disentangle activations solely responsible for sensory processing from those involved in visual awareness (for a review see: Sterzer *et al.*, 2009). One particularly popular paradigm is binocular rivalry. In this set-up, the subject is presented with a different stimulus to each eye. Instead of perceiving a mixture of the two stimuli, subjects typically only perceive one of the two images at a time. After a few seconds of perceiving one stimulus, the other stimulus becomes dominant. Thus, conscious perception alternates while physical stimulation remains stable (Blake and Logothetis, 2002).

According to the most accounts of binocular rivalry, the alternation between the two percepts is caused by competition between the neural representations of the two stimuli. The idea is that the neural representation of the currently dominant stimulus inhibits the representation of the nondominant stimulus. Over time, a combination of adaptation and noise causes the neural representation of the dominant stimulus to become weaker, eventually leading to a switch in dominance (Seely and Chow, 2011). Support for this idea comes from experiments showing large effects of changing stimulus characteristics, such as luminance and contrast, on dominance dynamics (Fahle, 1982; Kang, 2009). Furthermore, there is direct evidence that increases in adaptation result in decreases in subsequent dominance duration (Kang and Blake, 2010).

Several brain areas have been associated with different aspects of this process (Tong *et al.*, 2006). Fronto-parietal areas

have been implicated mostly in percept stabilization and attentional processes (Sterzer and Rees, 2008; Wilcke et al., 2009; Zaretskaya et al., 2010), whereas the actual neural representations of the input stimuli are mostly found in occipital and temporal visual areas (Britz et al., 2011; Haynes and Rees, 2005; Hsieh et al., 2012; Tong et al., 1998). Therefore, competition between the neural representations of the input images is most likely to happen along the ventral visual stream. However, it remains unclear at what level in the visual hierarchy this competition is resolved.

According to one view, competition happens in low-level, monocular areas of the visual cortex. In line with this idea, various neuroimaging studies have shown that, already in very early areas, activity reflects the dominant percept (Haynes and Rees, 2005; Tong and Engel, 2001; Wunderlich et al., 2005). In one representative study, functional magnetic resonance imaging (fMRI) was used to measure activity in the lateral geniculate nucleus (LGN) of the thalamus while participants were presented with a high contrast grating to one eye and a low contrast grating to the other eye. It was found that activity in the LGN increased when the high contrast grating was dominant and decreased when the low contrast grating was dominant (Wunderlich et al., 2005). These findings seem to imply that binocular rivalry is resolved already in the early stages of visual processing.

However, according to another view, competition between the two representations is resolved in more high-level, temporal areas and is then fed back via reentrant connections to early visual areas. The predictive coding account of binocular rivalry is in line with this view (Hohwy et al., 2008). According to this idea, competition takes place between high-level hypotheses about the incoming sensory input. Here, a strong prior that the world constantly changes causes the hypothesis of the currently dominant percept to lose strength over time, which explains the perceptual switch. In line with this idea, in contrast with human fMRI studies, animal studies have shown that most neurons in primary visual cortex represent the actual sensory input while neurons in high-level temporal cortex mostly reflect the dominant percept (Leopold and Logothetis, 1996).

The current study aims to dissociate between low-level and high-level explanations of binocular rivalry by investigating the direction of information flow prior to a switch in dominance. According to a low-level account, information about the dominant percept would first be present in early visual areas and then flow over time to more downstream areas. In contrast, according to a high-level account the direction of information flow would instead be more top-down, from high-level to low-level areas. The direction of information flow will be investigated using a combination of multivariate pattern analysis on source-reconstructed magnetoencephalography (MEG) measurements and Granger causality.

Methods

Subjects

Twenty (10 women) healthy subjects (mean age = 27.8 years, SD = 9.5), with normal or corrected-to-normal vision participated. One participant was excluded due to excessive head movement (more than 15 mm). Thus, in total 19 participants were included in the main analyses. All participants gave written informed consent to participate in the study. The study was approved by the local ethics committee (Commissie Mensgebonden Onderzoek regio Arnhem-Nijmegen).

Experimental design

During the main experiment, subjects had to view superimposed images of a red face and a green building or a red building and a green face through red- and green-filtered anaglyph glasses. The glasses had a green filter covering the right eye and a red filter covering the left eye such that the right eye was only exposed to the green image and the left eye only to the red image. Stimuli were surrounded by a white border and contained a white fixation cross which were transmitted through both filters to increase percept stabilization. There were two different face images and two different building images leading to eight different stimuli (two faces × two buildings × two colors). Images depicted famous faces (Emma Watson and Brad Pitt) and buildings (the Taj Mahal and the Notre Dame), obtained from the World Wide Web. All images were corrected for on screen luminance with the SHINE toolbox for MATLAB (Willenbockel et al., 2010). Presentation of stimuli was done using Presentation software (Version 9.13, www.neurobs.com). Stimuli were presented at a size of 2.6 cm and a distance of 75 cm via an LCD projector located outside the magnetically shielded room and were back-projected onto a translucent screen inside the magnetic room via two front-silvered mirrors.

One stimulus was shown for 30 seconds during which participants had to indicate their dominant percept by means of a button press each time a switch occurred. Participants were instructed to withhold their key press until one image gained (near) complete dominance (such as to exclude mixed percepts). Which button corresponded to which percept was randomized over trials and presented on screen for 1 second prior to the stimulus presentation. After the stimulus was shown, a blank screen with a fixation cross was presented for 3 seconds followed by the next stimulus (Fig. 1). In one block this was repeated eight times, after which the participant had a short break during which they could relax. The break ended when the participant pressed a button. The experiment started with a practice session in which stimuli were counterbalanced, such that all stimuli were shown to each participant but the order of presentation was randomized over participants. After the practice block the participant had time to ask questions. Following the practice block, the main experiment consisted of
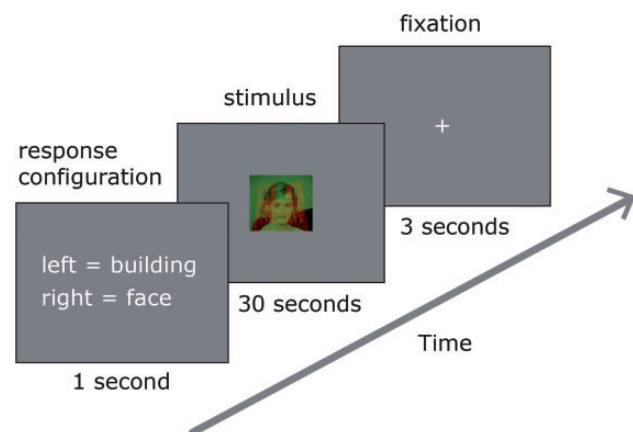

Figure 1. Study design. Stimuli were viewed through red- and green-filtered anaglyph glasses. Prior to stimulus presentation, the configuration of the buttons was presented. Each stimulus was presented for 30 seconds during which participants continuously indicated their current dominant percept by means of a button press. After this a fixation cross was presented for 3 seconds.

eight blocks in which the stimuli were randomized. The total experimental time was approximately 45 minutes (depending on the length of the self-paced breaks).

MEG acquisition

MEG data were recorded using a 275-sensor whole-head system (CTF Systems Inc., Port Coquitlam, Canada) at a sampling frequency of 1200 Hz. Due to malfunction, data from two sensors (MRF66 and MLC11) were not recorded. The MEG acquisition took place in a dimmed, magnetically shielded room. Before the experiment began, participants were instructed to minimize head movement and try to blink only when there was no stimulus on screen. During the experiment, head movement was monitored continuously via three coils, one in both ears and one in the nasion, using a real-time head localizer (Stolk *et al.*, 2013). During breaks, it was checked whether head movement exceeded 5 mm. In this case, participants were instructed to move back to their initial head position by the experimenter who had access to live video feedback of the head position relative to the initial position. Eye movements and blinks were monitored with a continuous bipolar electrooculogram (EOG) for later offline artifact rejection (see "Preprocessing" section). Vertical EOG was measured with two electrodes: one below and one above the left eye. Horizontal EOG was measured with one electrode to the left of the left eye and one electrode to the right of the right eye. The ground electrode was placed on the left mastoid.

Preprocessing

Data were analyzed using MATLAB version 8.1.0, R2013a (The Mathworks Inc, Natic, MA) and FieldTrip, an open-source MATLAB toolbox for the analysis of neuroimaging data (Oostenveld *et al.*, 2011). Trials were defined as measurements ranging from 2 seconds before until 1 second after the button press. Trials containing artifacts resulting from SQUID jumps or muscle contractions were automatically rejected. Before further processing, data were downsampled to 300 Hz sampling frequency to reduce memory and CPU load. Removal of EOG artifacts was performed by first applying ICA on the data and then removing the components that showed the highest correlation with the EOG channels. Trials were also visually inspected on eye blinks and kurtosis over channels, ensuring that trials with very high variance between channels (kurtosis above 15) were removed.

Beamforming

Source-level time courses were reconstructed for every trial with an LCMV beamformer (Van Veen *et al.*, 1997). This method creates a spatial filter that optimizes the signal coming from a given source while suppressing signals coming from other sources. As a head model, the single shell model described by Nolte (2003) was used. Individual grids with a resolution of 10 mm were computed based on T1-weighted MRI data of each participant acquired with a 1.5 T whole body scanner (Siemens Magnetom Avanto, Siemens, Erlangen, Germany). Alignment of the MEG and MRI data was based on vitamin E markers which marked the same location in the ears as the fiducial ear coils during the MEG measurement. For later comparison between subjects, a template grid in Montreal Neurological Institute (MNI) space was inverse-warped to subject-specific coordinates based on the subject's T1 image.

Occipital, temporal, parietal and frontal regions of interest (ROIs) were defined using an anatomical atlas in MNI space on the inverse-warped subject-specific templates. For the occipital ROI, 78 grid points were defined, for the temporal ROI, 204 grid points, for the parietal ROI, 64 grid points and for the frontal ROI, 296 grid points (Fig. 2).

Classification

In order to investigate the neural representations of the dominant percept, we used multivariate pattern analysis to reveal stimulus-specific information. Classification of the dominant percept was performed separately in the four ROIs on the amplitude of the source-reconstructed signals. Classification was done on the averaged signal of every five sample points to induce smoothness while still maintaining a high temporal resolution (60 Hz). An elastic net logistic regression algorithm was used for classification (Friedman *et al.*, 2010). Given training data, this algorithm maximizes the log-likelihood, penalized by the elastic net penalty:

$$p_\alpha(b) = \sum_{j=1}^{p} \left(\frac{1}{2}(1-\alpha)b_j^2 + \alpha|b_j| \right),$$

where b is the vector of regression coefficients. This penalty term combines ridge and lasso regularization through a mixing parameter α. That is, when $\alpha = 0$, $p_\alpha(b)$ reduces to a ridge penalty
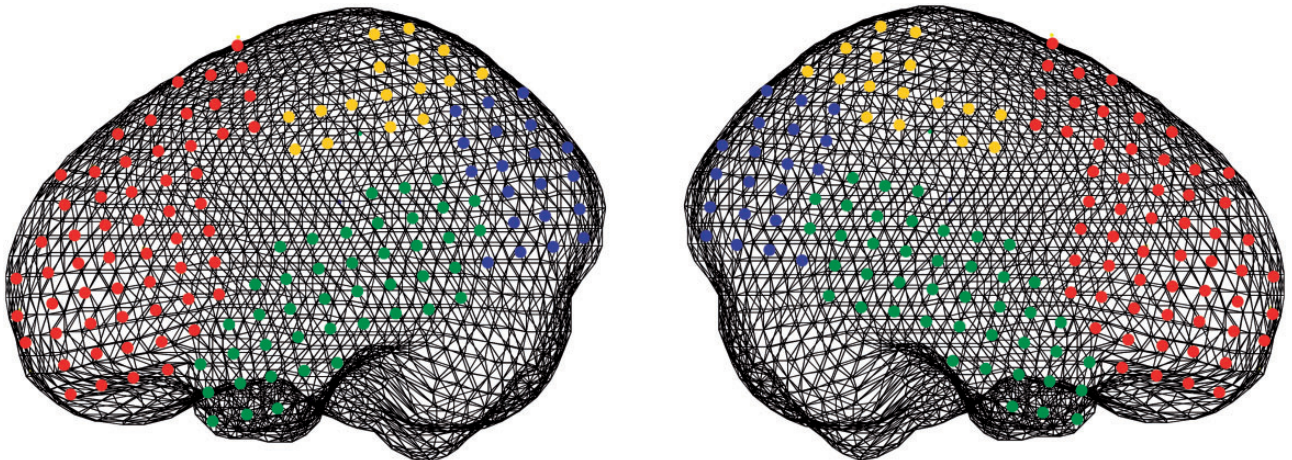

Figure 2. Locations of the beamformed occipital (blue), temporal (green), parietal (yellow) and frontal (red) sources in one subject.

and when $\alpha = 1$, $p_\alpha(b)$ reduces to a lasso penalty. In the current classification analysis, the mixing parameter was set to $\alpha = 0.01$, leading to both sparse and smooth vectors of regression coefficients. The influence of the elastic net penalty is controlled by a regularization parameter λ, which was optimized using a nested cross-validation procedure. In case the training set did not contain an equal number of face and building trials, the numbers were balanced by randomly sampling trials from the training set. Prior to classification, the input data was standardized relative to the mean and standard deviation of the training set.

Classifier performance was validated using 5-fold cross-validation, ensuring that the classifier was always tested on data that it was not trained on. Furthermore, by imposing a sparsity constraint, the elastic net algorithm performed feature selection by setting a large number of features that were not necessary for classification to zero. The computations were run on a distributed computing cluster with cores whose clock rate ranged between 2.0 GHz and 3.6 GHz. Classifier accuracy was quantified in terms of proportion of correctly classified trials. Note that the class assignment for a trial was determined by the subject's response at the button press during that trial.

Granger causality

The basic idea of Granger Causality (GC) is that a time series **Y** G-causes another time series **X** if **Y** precedes **X** and if the past of **Y** conveys information about the future of **X** beyond information already contained into the past of **X** itself (Granger, 1988). Recently, GC has been applied to local field potentials in monkeys to dissociate between bottom-up and top-down mechanisms (Bastos et al., 2014; van Kerkoerle et al., 2014). Here, similarly, Granger causality was used to assess whether information about the dominant percept flowed from early visual to more high-level areas (bottom-up) or from high-level to hierarchically lower-level areas (top-down) over time. Because we were mainly interested in processes leading up to a switch, only data from before the button press were used. That is, the time window ranging from minus two to zero seconds relative to button press. Furthermore, because we had more than two ROIs, we used multivariate Granger causality which assesses the causality between two ROIs conditional on the time series of the other ROIs in the set.

For each pair of regions, the pairwise-conditional GC statistic was calculated, which is defined as the log-likelihood ratio

$$\mathcal{F}_{\mathbf{Y} \to \mathbf{X}} = \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|},$$

where Σ_{xx} is the residual covariance matrix of the full model which predicts **X** using the past of **Y**, the past of **X** itself and the past of the other variables in the subset and Σ'_{xx} is the residual covariance matrix of the reduced model in which **X** is predicted with only the past of **X** itself and the past of the other variables in the subset. This value could therefore be seen as the amount of evidence in favor of **Y** causing **X**. All GC analyses were performed using the Multivariate Granger Causality (MVGC) Toolbox by Barnett and Seth (2014).

Results

Behavioral results

On average there was a switch every 15.39 seconds ($SD = 6.01$), leading to a total of 168 trials on average over participants

($SD = 87.45$). Since participants were instructed not to respond when they perceived a mixed percept, it could be the case that between two subsequent button presses the participant experienced a mixed percept for any duration. Therefore, we do not know the actual dominance durations per percept. For each stimulus, the percentage that it was reported as dominant out of the total number of trials, averaged over participants, is depicted in Fig. 3. The face stimuli were more often reported as being dominant ($M = 58.85\%$, $SD = 5.29$) than the building stimuli ($M = 41.15\%$, $SD = 5.29$). Furthermore, within the face category, the Brad Pitt image ($M = 33.24\%$, $SD = 6.07$) was more often dominant than the Emma Watson image ($M = 25.61\%$, $SD = 6.36$). There were no effects of gender on these distributions. Concerning eye dominance, there was no significant difference between the percentage of left-eye dominant trials ($M = 44.57\%$, $SD = 18.69$) and right-eye dominant trials ($M = 55.43\%$, $SD = 18.69$).

Classification analysis

It has been shown that the time it takes to press the button after a perceptual switch has occurred is approximately 500 milliseconds (Sandberg et al., 2013). Therefore, the actual switch in dominancy happens around 500 milliseconds before the button press. In Fig. 4A the average classification accuracy around this time, between 750 milliseconds and 250 milliseconds before the button press, is depicted per ROI for the individual participants. Average classification accuracy for this period was significantly above chance over participants in the occipital ($M = 0.515$, $SD = 0.019$; $t(18) = 3.35$, $P < 0.01$) and in the temporal ($M = 0.513$, $SD = 0.020$; $t(18) = 2.80$, $P < 0.01$) ROIs but not in the parietal ($M = 0.505$, $SD = 0.021$; $t(18) = 1.01$, $p > 0.1$) and frontal ROIs ($M = 0.505$, $SD = 0.016$; $t(18) = 1.47$, $P > 0.05$). Note that average accuracies are very low due to averaging over all subjects and time points of which only a subset is expected to contain signal.

As can be seen, there are great individual differences in classification accuracy. Furthermore, there is also a large variation within participants, which indicates that classification accuracy varied over time. In Fig. 4B, the classification accuracy is shown over the entire trial for the participant with the highest average accuracy. From these results alone, it was not possible to infer in which brain area relevant information about the dominant percept was first present. To be able to infer the direction of information flow, we subsequently employed Granger causality analysis.
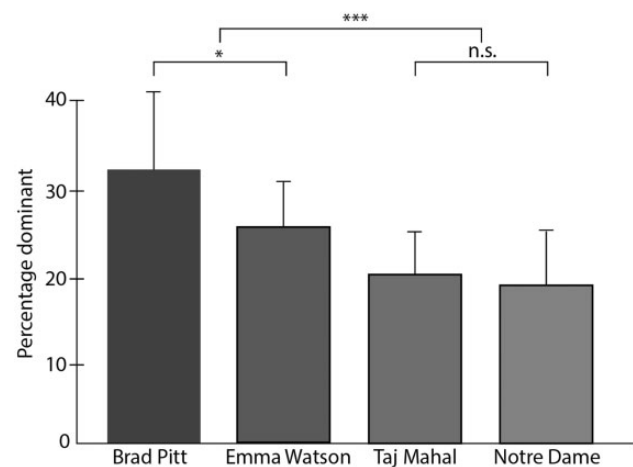

Figure 3. Behavioral results. Average percentage of dominance for the four different stimuli. *$P < 0.01$, ***$P < 0.0001$.
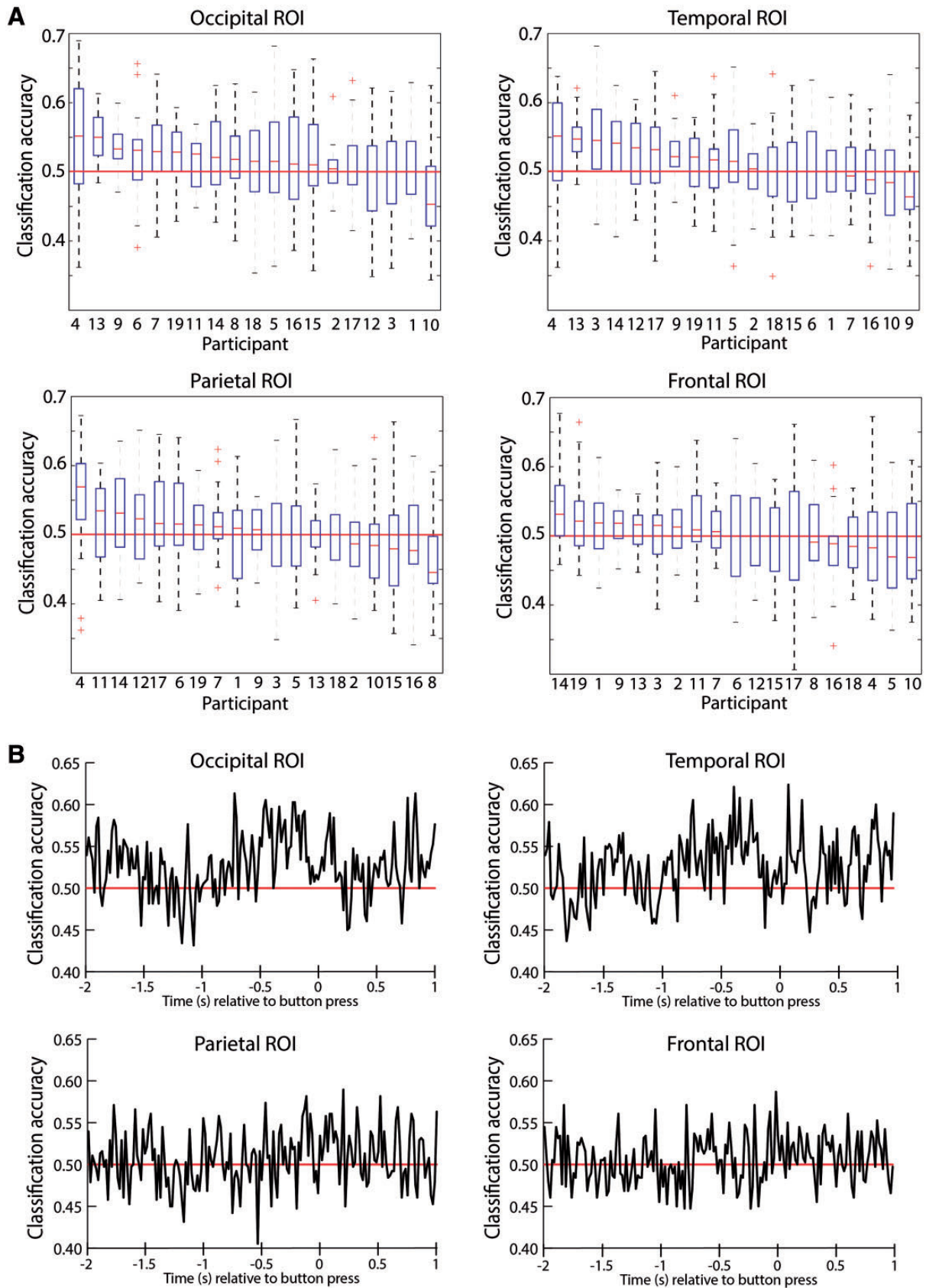
Figure 4. **(A)** Classification accuracy of individual participants during the 500 milliseconds around the switch in the different ROIs. For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers are the most extreme points that are not yet outliers and the plusses indicate outliers. **(B)** Classification accuracy over time per ROI for the participant with the highest average accuracy.

Granger causality analysis

We wanted to use GC analysis to reveal the direction of information flow about the dominant percept between the different ROIs. There were three candidate time series to use as input for the GC analysis: the source-reconstructed signal amplitudes, the classification accuracy traces and the probability traces. The classification accuracy traces reflect the proportion of correctly classified trials over time (ROI × time), whereas the probability traces reflect the probability of classifying a given trial as belonging to the true class for each time point (ROI × trials × time). The latter can be seen as a more detailed measure of classification accuracy. Since it was unclear a priori which would be most informative, a simulation study was conducted.

In the simulation, source amplitudes were generated for two ROIs as a 1-Hz sinusoidal signal with an amplitude of 3 and a phase of 0 sampled at 60 Hz for 3 seconds (similar to the measured signal). The signal of the second ROI was lagged (circularly shifted) by five samples (83.3 milliseconds) relative to the first ROI. We created a dataset consisting of 50 trials in condition 1 that contained the (lagged) signal in each source and 50 trials in condition 2 that did not contain any signal in each source. To each trial, irrespective of the condition, zero mean Gaussian noise with unit variance was added. Classification was done per time point in each ROI using regularized logistic regression. GC analysis was done separately on the source amplitudes, on the classification accuracy traces and on the probability traces. The simulation was repeated 200 times. Pairwise GC values for both directions from all analyses are depicted in Fig. 5.

The results of the simulation indicate that GC analysis of the probability traces is most suitable to reveal the true underlying connectivity structure: this analysis showed most evidence in favor of the true direction relative to the false direction. This could be explained by the fact that in comparison to the classification accuracies, the probability traces contain more information because they contain the probabilities of the two classes per trial, and not only the decision of the classifier averaged over trials. In comparison to the source amplitudes, the probability traces contain representation-specific information while the source amplitudes also contain task-irrelevant information and therefore contain more noise relative to the signal of interest. Thus, for the main experiment, the pairwise Granger causality was calculated between the empirical probability traces of the different ROIs before the button press for all participants.
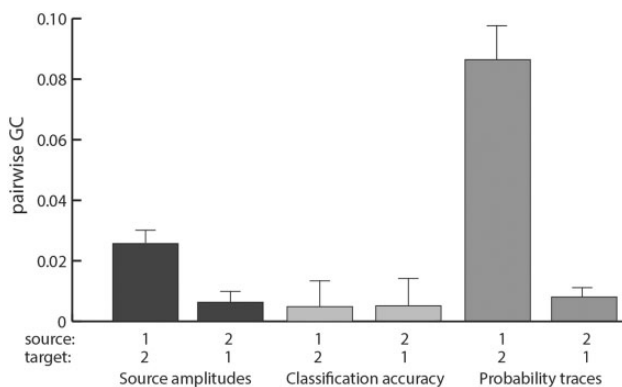

Figure 5. Results of the simulation using GC to infer directionality from source amplitudes, classification accuracy traces and probability traces. The true direction was defined to be from 1 to 2. Error bars indicated standard deviation over the 200 repetitions.

The most stringent assumption of GC is that the analyzed time series should be stationary (Granger, 1988). We investigated this with the KPSS test which tests the null hypothesis that a given time series is stationary (Kwiatkowski et al., 1992). After correction for false positive rate, none of the used time series were significantly nonstationary. Thus there was no indication that stationarity was violated.

The average pairwise-conditional GC estimates for all combinations of ROIs are shown in Fig. 6A. The best model order was always one sample point, which in the current experiment indicates a temporal lag of 16.7 milliseconds between the two signals. To quantify differences in direction of coupling between ROIs, paired samples t-tests were performed over participants on the GC values. There was a significant difference between the GC for the temporal-to-occipital ($M = 9.65e\text{-}04$, $SD = 9.07\text{-}04$) and the occipital-to-temporal ($M = 6.29e\text{-}04$, $SD = 7.79\text{-}04$) directions ($t(18) = 3.14$, $P < 0.01$), with more evidence for top-down directionality. Furthermore, the full model, predicting the time course of the one ROI by including the past of the other ROI, was significantly better than the reduced model in 14 out of 19 participants for the top-down GC and in 10 out of 19 participants for the bottom-up GC. All comparisons of the direction of coupling between the other ROIs were nonsignificant, indicating that there was no dominant direction of information flow between these pairs of ROIs. Therefore, further analyses focused on the coupling between occipital and temporal ROIs.

To take into account the reaction time after a switch has occurred, we also performed the GC analysis on the probability traces up to 500 milliseconds before the button press (i.e. the time window of −2 to −0.5 seconds relative to button press). This analysis showed similar results with significantly more evidence for top-down directionality ($M = 10.43e\text{-}04$, $SD = 11.31e\text{-}04$) than for bottom-up directionality ($M = 8.81e\text{-}04$, $SD = 13.09e\text{-}04$; $t(18) = 2.29$, $P < 0.05$) between occipital and temporal ROIs.

To further investigate whether the increase in top-down information flow was specific to the activity leading up to a perceptual switch, GC analysis was performed on sliding time windows of 500-milliseconds each throughout the trial. The onset of each next window was six sample points (100 milliseconds) after the onset of the previous window. The results are shown in Fig. 6B. The average GC values for each time window are plotted at the central time point of that window. The shaded area represents the 95% confidence interval over participants. Paired samples t-tests were performed at each time window to compare GC for top-down directionality with GC for bottom-up directionality. There was significantly more evidence for top-down GC in the windows of −1.8 to −1.3 ($t(18) = 2.44$, $P < 0.05$) and −0.9 to −0.4 ($t(18) = 2.25$, $P < 0.05$) seconds relative to button press. After this, there was significantly more evidence for bottom-up directionality in the window of −0.5 to 0 ($t(18) = -2.40$, $P < 0.05$) seconds relative to button press. However, these p-values were uncorrected for multiple comparisons. After correction none of the differences remained significant. Therefore, these results should be interpreted with caution.

To explore whether the same top-down pattern between occipital and temporal ROIs could be observed in the source-reconstructed signal, we also performed GC on the source amplitude time courses of all trials before the button press. Again, the model order was always one, which in this case indicates a temporal lag of 3 milliseconds. This analysis revealed on average more occipital to temporal coupling than temporal to occipital coupling, but this difference was not significant ($P = 0.073$). We did observe significantly more parietal to temporal coupling than temporal to parietal coupling ($t(18) = 4.29$, $P < 0.001$) and more
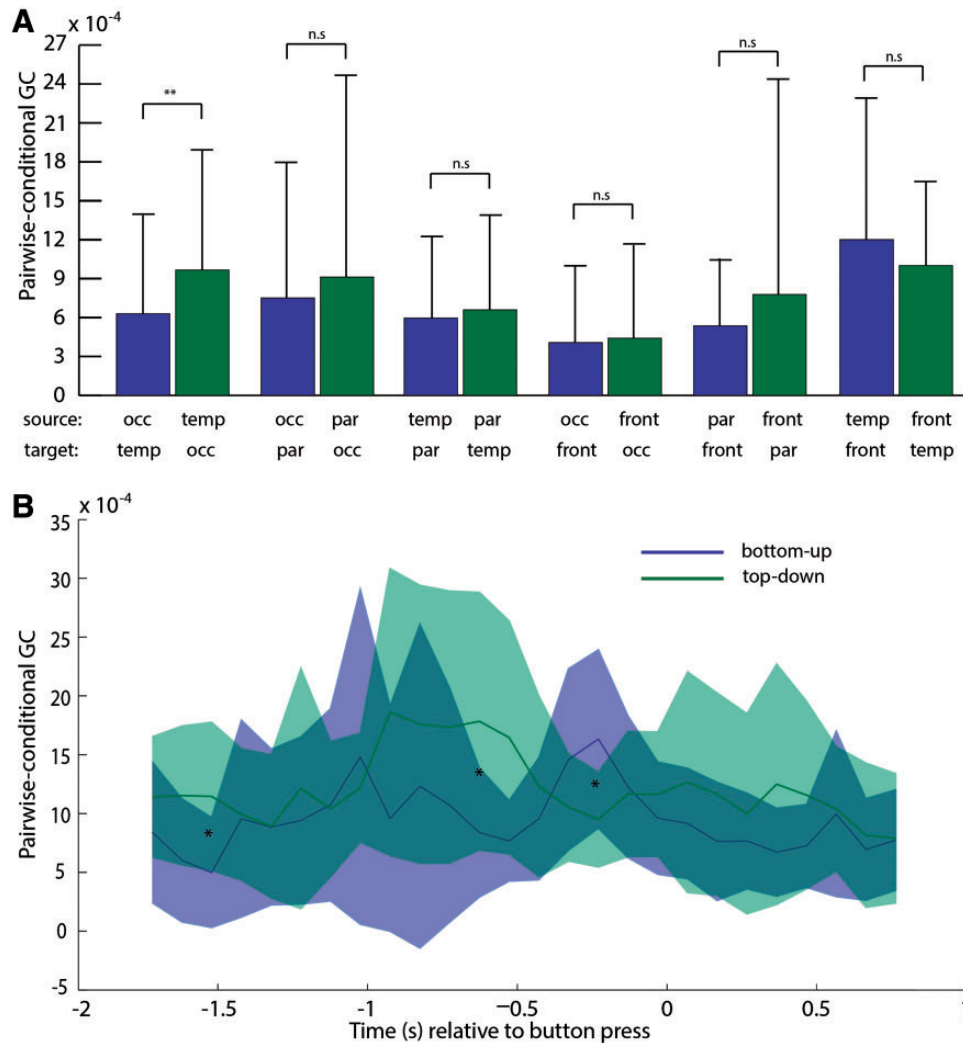
Figure 6. (A) Average pairwise-conditional GC for all pairs of ROIs. **P < 0.01, n.s. = nonsignificant. (B) Average pairwise GC for the two directions for different time windows. Shaded area represents the 95% confidence interval. *P < 0.05, uncorrected.

parietal to frontal coupling than frontal to parietal coupling ($t(18) = 2.83$, $P < 0.05$) although the latter did not survive correction for multiple comparisons.

Recent animal studies have shown that activity in certain oscillatory bands, especially alpha and beta, may reflect interareal top-down communication (Bastos *et al.*, 2014; van Kerkoerle *et al.*, 2014). To explore this idea we have also conducted a spectral GC analysis. However, our results did not reveal any directionality in specific frequency bands.

Discussion

The current study aimed to investigate the direction of information flow about the dominant percept prior to a perceptual switch during binocular rivalry. First, information about the dominant percept was quantified in temporal, occipital, parietal and frontal ROIs as the classification accuracy based on MEG measurements. This analysis showed that even though there were substantial individual differences in accuracy, both occipital and temporal areas contained relevant information about the dominant percept, while classification was near chance in both

parietal and frontal ROIs. This is in line with previous fMRI and MEG studies that found representations of the dominant percept in early visual and temporal areas (Leopold and Logothetis, 1996; Tong *et al.*, 1998; Haynes and Rees, 2005; Sandberg *et al.*, 2013; Wang *et al.*, 2013).

To quantify the direction of information flow between the different ROIs, multivariate GC was applied to the probability traces before the button press. This analysis revealed significantly more evidence for the temporal-to-occipital direction than for the occipital-to-temporal direction. To take into account the time it takes to respond after a perceptual switch has occurred, GC was also applied to the probability traces up to 500 milliseconds before button press. This analysis showed similar results, with more evidence for temporal-to-occipital directionality. Furthermore, a subsequent exploratory analysis revealed that the stronger top-down GC was specific for the time windows until 500 milliseconds before the button press. Taken together, these results indicate that prior to a perceptual switch, information about the dominant percept flows mainly from high-level to low-level visual areas over time.

This finding is in line with high-level explanations of binocular rivalry, stating that competition between the two stimuli

is first resolved in high-level areas and information about the dominant percept is then fed back to hierarchically lower areas over time (Mumford, 1991; Rao and Ballard, 1999; Kersten *et al.*, 2004; Hohwy *et al.*, 2008). One of these theories is the predictive coding account of binocular rivalry put forward by Hohwy *et al.* (2008). In light of this theory, the increased top-down information flow prior to a switch could be seen as reflecting the sending of a prediction about the incoming input to sensory areas. Furthermore, our findings also indicate that after this top-down coupling there is a relative increase in bottom-up information flow. This could be interpreted as confirming the prediction in the incoming sensory data. However, as mentioned before, these findings were nonsignificant after correction for multiple comparisons and should therefore be interpreted with caution.

Furthermore, even though there was more evidence for top-down flow, the main GC analysis also revealed significant evidence for bottom-up flow in 10 out of 19 participants. Hierarchical/hybrid models of binocular rivalry may therefore better explain the present findings. According to these models, competition between the two stimuli takes place at all levels of the visual hierarchy. The idea is that competition may already favor one percept in early visual areas, which then transmit this information to higher-areas. Feedback from higher-areas is then necessary to strengthen this representation and to eventually produce perceptual dominance (for a review, see Tong *et al.*, 2006). According to hybrid models, high-level and low-level areas are differentially involved depending on the complexity of the used stimuli (Blake *et al.*, 2014). Considering the relatively complex stimuli used in the current study, namely buildings and faces, hybrid models would predict the competition to take place more in higher-level areas and to feed back to lower areas over time. The current design cannot distinguish with certainty between top-down and hybrid models. To this end, future studies that compare the direction of information flow between binocular rivalry with complex stimuli as well as simple stimuli are necessary.

In this study, we used Granger causality on probability traces from different ROIs to reveal which area contained information about the dominant percept first. To our knowledge this approach has not been used before. To test its validity, a simulation study was conducted which showed that GC on probability traces was able to reveal the true direction between two time series. Furthermore, we did not find any indication that our data were nonstationary, which supports the soundness of our approach. To explore whether the same top-down GC pattern was present in the source amplitude time series, we also performed GC analysis on these signals. This analysis did not show increased temporal-to-occipital coupling but revealed increased parietal-to-temporal coupling. In contrast to the probability traces, the source amplitude signals mostly contain activity that is not relevant for the current dominant percept but is common to every trial. Therefore, this increase in parietal-to-occipital coupling likely reflects processing that is generally involved during binocular rivalry such as percept stabilization (Sterzer *et al.*, 2009; Wilcke *et al.*, 2009; Zaretskaya *et al.*, 2010). However, since we do not have an appropriate nonrivalry baseline, we cannot draw any conclusions about the function of this coupling. Future research focusing on time-domain GC during rivalry and control is necessary to address this question. The same holds for the absence of frequency specific top-down signals in our data; future research using an appropriate baseline is necessary to explore this idea further.

There are a number of potential limitations to the current study. First, in some participants, at some time points, classification accuracy was very low. If classification accuracy is below chance it is unclear what the probability traces reflect other than noise. This could result in spurious GC patterns. However, it is unlikely that this is the entire cause of our findings since we found consistent differences in direction of coupling on the group level and this pattern was also confirmed with an additional sliding time window analysis. Still, future studies should focus on reaching higher classification accuracies by for example adopting a more objective measure of timing of perceptual switches such as pupil dilation (Frässle *et al.*, 2013). Another possible confounding factor of the employed analyses is that the temporal ROI contained more sources (i.e. more features) than the occipital ROI. This could mean that due to the fact that the classifier had access to more information in the temporal ROI, the SNR in this area was higher, allowing classification accuracy to increase earlier than in the occipital ROI. This in turn could explain the GC results. However, this would mean that the ROI with the highest number of sources would always be the source of information flow. The GC results of coupling between the other ROIs show that this is not the case. For example, parietal-to-occipital GC was higher than occipital-to-parietal GC even though the occipital ROI contains more sources (78) than the parietal ROI (68). To completely rule out this explanation, future studies could use searchlight methods to employ ROIs of equal size (Kriegeskorte *et al.*, 2006).

In conclusion, by using a novel combination of different techniques, the current study indicates that during binocular rivalry, prior to a perceptual switch, information about the dominant percept flows mainly from downstream to upstream areas over time. Even though these findings can be explained by both high-level and hybrid models, it is clear that they cannot be explained by a purely bottom-up account of binocular rivalry. In order to further dissociate between high-level and hybrid models, future studies investigating the direction of information flow in different binocular rivalry settings are necessary.

Data availability

Data are available upon request.

Acknowledgements

References

Bastos, AM, Vezoli, J, Bosman, CA *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 2014; **85**: 390–401.

Blake, R, Brascamp, J, Heeger, DJ. Can binocular rivalry reveal neural correlates of consciousness? *Philos Trans R Soc Lond Ser B Biol Sci* 2014;369: 20130211.

Blake, R, Logothetis, NK. Visual competition. *Nat Rev Neurosci* 2002; **3**: 13–21.

Britz, J, Pitts, MA, Michel, CM. Right parietal brain activity precedes perceptual alternation during binocular rivalry. *Human Brain Map* 2011; **32**: 1432–42.

Fahle, M. Binocular rivalry: suppression depends on orientation and spatial frequency. *Vis Res* 1982; **22**: 787–800.

Frässle, S, Sommer, J, Naber, M *et al*. Neural correlates of binocular rivalry as measured in fMRI are partially confounded by observers' active report. *J Vis* 2013;**13**: 937.

Friedman, J, Hastie, T, Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software* 2010; **33**: 1–22.

Granger, CWJ. Some recent development in a concept of causality. *J Econometrics* 1988; **39**: 199–211.

Haynes, J-D, Rees, G. Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol CB* 2005; **15**: 1301–7.

Hohwy, J, Roepstorff, A, Friston, K. Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 2008; **108**: 687–701.

Hsieh, PJ, Colas, JT, Kanwisher, NG. Pre-stimulus pattern of activity in the fusiform face area predicts face percepts during binocular rivalry. *Neuropsychologia* 2012; **50**: 522–9.

Kang, M-S. Size matters: a study of binocular rivalry dynamics. *J Vis* 2009; **9**: 17.1–11.

Kang, M-S, Blake, R. What causes alternations in dominance during binocular rivalry? *Atten Percept Psycho* 2010; **72**: 179–86.

Kersten, D, Mamassian, P, Yuille, A. Object perception as Bayesian inference. *Ann Rev Psychol* 2004; **55**: 271–304.

Kriegeskorte, N, Goebel, R, Bandettini, P. Information-based functional brain mapping. *Proc Natl Acad Sci USA* 2006; **103**: 3863–8.

Kwiatkowski, D, Phillips, P, Schmidt, P *et al*. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J Econom* 1992;**54**: 159–78.

Leopold, DA, Logothetis, NK. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 1996; **379**: 549–53.

Mumford, D. On the computational architecture of the neocortex - I. The role of the thalamo-cortical loop. *Biol Cybern* 1991;**65**: 135–45.

Nolte, G. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics Med Biol* 2003;**48**: 3637–52.

Oostenveld, R, Fries, P, Maris, E *et al*. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011; 1–9.

Rao, RP, Ballard, DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 1999;**2**: 79–87.

Sandberg, K, Bahrami, B, Kanai, R *et al*. Early visual responses predict conscious face perception within and between subjects during binocular rivalry. *J Cogn Neurosci* 2013; **25**: 969–85.

Seely, J, Chow, CC. Role of mutual inhibition in binocular rivalry. *J Neurophysiol* 2011; **106**: 2136–50.

Sterzer, P, Kleinschmidt, A, Rees, G. The neural bases of multistable perception. *Trends Cogn Sci* 2009; **13**: 310–8.

Sterzer, P, Rees, G. A neural basis for percept stabilization in binocular rivalry. *J Cogn Neurosci* 2008; **20**: 389–99.

Stolk, A, Todorovic, A, Schoffelen, JM *et al*. Online and offline tools for head movement compensation in MEG. *NeuroImage* 2013; **68**: 39–48.

Tong, F, Engel, SA. Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 2001; **411**: 195–99.

Tong, F, Meng, M, Blake, R. Neural bases of binocular rivalry. *Trends Cogn Sci* 2006; **10**: 502–11.

Tong, F, Nakayama, K, Vaughan, JT *et al*. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 1998; **21**: 753–59.

van Kerkoerle, T, Self, MW, Dagnino, B *et al*. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc Natl Acad Sci* 2014; **111**: 14332–41.

Van Veen, BD, van Drongelen, W, Yuchtman *et al*. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans BioMed Eng* 1997; **44**: 867–80.

Wang, MG, Arteaga, D, He, BJJ. Brain mechanisms for simple perception and bistable perception. *Proc Natl Acad Sci USA* 2013; **110**: E3350–E3359.

Wilcke, JC, O'Shea, RP, Watts, R. Frontoparietal activity and its structural connectivity in binocular rivalry. *Brain Res* 2009; **1305**: 96–107.

Willenbockel, V, Sadr, J, Fiset, D *et al*. Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods* 2010; **42**: 671–84.

Wunderlich, K, Schneider, KA, Kastner, S. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci* 2005; **8**: 1595–602.

Zaretskaya, N, Thielscher, A, Logothetis, NK *et al*. Disrupting parietal function prolongs dominance durations in binocular rivalry. *Curr Biol* 2010;**20**: 2106–11.