# Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception

**Jan Brascamp**[1,2,3,*], **Randolph Blake**[4,5], and **Tomas Knapen**[2,6]

[1]Helmholtz Institute and Division of Experimental Psychology, Utrecht University, Utrecht, The Netherlands [2]Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands [3]Department of Psychology, Michigan State University, East Lansing, MI 48823, U.S.A. [4]Department of Psychology and Vanderbilt Vision Research Center, Vanderbilt University, Nashville TN 37212, U.S.A. [5]Department of Brain and Cognitive Sciences, Seoul National University, Seoul, 151-742, Republic of Korea [6]Cognitive Psychology, Department of Behavioral and Movement Sciences, VU Amsterdam

## Abstract

The human brain's executive systems play a vital role in deciding and selecting among actions. Selection among alternatives also occurs in the perceptual domain, for instance when perception switches between interpretations during perceptual bistability. Whether executive systems also underlie this functionality remains debated, with known fronto-parietal concomitants of perceptual switches being variously interpreted as reflecting the switches' cause, or as reflecting their consequences. We developed a paradigm where the two eyes receive different inputs and perception demonstrably switches between these inputs, yet where switches themselves are so inconspicuous as to become unreportable, minimizing their executive consequences. Fronto-parietal fMRI BOLD responses that accompany perceptual switches were similarly minimized in this paradigm, indicating that these reflect the switches' consequences rather than their cause. We conclude that perceptual switches do not always rely on executive brain areas, and that processes responsible for selection among alternatives may operate outside of the brain's executive systems.

## Introduction

The intriguing phenomenon of bistable perception arises when an observer views a stimulus with several mutually exclusive perceptual interpretations: although the stimulus remains constant, perception fluctuates between interpretations. Fascination with this phenomenon no doubt originates from the arresting quality of experiencing bistable perception, but scientific interest has been spawned by the uniquely informative position of perceptual bistability regarding the relation between sensation and cognition. On the one hand bistable

*Corresponding author: brascamp@msu.edu.

perception is associated with characteristic activity patterns in relatively well understood visual brain regions[1–2], and its perceptual dynamics allow rigorous control using methods from psychophysics[3]. At the same time, other aspects of bistable perception implicate the involvement of complex cognitive functions like attention and action planning[e.g. 4–6]. Still, it remains unclear whether bistable perception is, at heart, a sensory phenomenon that arises from local processes within the visual brain[7–10], with cognitive factors playing a mere modulating role, or whether it is an expression of a processes also involved in guiding attention and response selection[2,11–12]. Arbitrating between these views is particularly difficult because perception, attention and behavioral responses are typically intertwined, as when viewers of perceptual bistability actively observe, and sometimes respond to, perceptual switches.

Recent debate focuses specifically on the challenge of interpreting blood-oxygen-level dependent (BOLD) signals measured using functional imaging around the time of perceptual switches. The right-lateralized fronto-parietal regions that show a switch-related elevation in BOLD signal overlap with a network implicated in attention and motor planning[13,14], prompting the interpretation that this network is causally involved in switches[15–17]. But recent studies have questioned this interpretation. One study[18] observed an equivalent BOLD signal elicited by attention-grabbing events that resembled spontaneous switches, but that were imposed externally, suggesting that the signal reflects attentional reorienting following the perceptual switch. Another study demonstrated the switch-related BOLD signal to be notably reduced if switches were task-irrelevant, and thus less actively attended[19], suggesting that attention and response planning importantly contribute.

These findings and the ensuing discussion make clear that a functional interpretation of switch-related BOLD signals is hindered by the complex of perceptual and cognitive events that accompany perceptual switches, with various authors listing as factors the initiation of the switch itself, attentional reorienting, response planning, self-monitoring and introspection[17–21]. Whereas some of these components can be minimized by rendering perceptual switches task-irrelevant[19], others would appear to require perceptual switches that remain altogether unattended or that are perfectly matched to some baseline condition in terms of perceptual experience and salience.

The above considerations raise a conundrum. To functionally interpret switch-related neural events, switches must be isolated from their attentional and behavioral consequences, but if perceptual switches do rely on neural events that also guide attention and behavior, then such an isolation may be unachievable even in principle. Indeed, when observers view stimuli that ordinarily provoke bistable perception but with attention withdrawn to a point where switches become unreportable, no evidence of any remaining switches is apparent[22–23].

Using a widely investigated form of perceptual bistability, binocular rivalry[1], we demonstrate here that a study of perceptual switches in isolation is possible. We developed a dichoptic viewing paradigm in which perception demonstrably switches between the two eyes' inputs – the hallmark signature of binocular rivalry – yet in which the switches themselves are rarely detected and, hence, often unreportable. From a theoretical

perspective, this demonstrates that one hitherto inherent component of perceptual switches, reportability, can be stripped away. From a practical standpoint, the paradigm minimizes the attention-related cascade that inevitably follows the salient perceptual switches of conventional paradigms, allowing a new assessment of switch-related BOLD responses that focuses on the switches' neural cause.

## Results

### Unreportable binocular rivalry switches

In binocular rivalry, conflict arises from the presentation of markedly different input to each eye (Figure 1**a**). Observers of such an arrangement often perceive only one eye's input at a time, with perception switching back and forth between the two during prolonged viewing, mirroring other forms of bistable perception[3,9,24–25]. In our study, each eye viewed a different sequence of quasi-randomly moving dots[cf. 26]. The two dot patterns were presented at corresponding retinal locations and both patterns had the same statistical properties (dot density, overall motion content). However, the positions and motion directions of the individual dots that comprised each pattern differed between eyes at all times, potentially causing binocular rivalry. In separate conditions the two patterns could furthermore differ in color (*different colors*; Figure 1**a**, top) or not (*same color*; Figure 1**a**, bottom). In the former condition, a perceptual switch between the dot patterns should be clearly visible as a color change. In the latter condition, however, a switch would entail a perceptual change between two statistically and chromatically identical, ever-changing dot patterns. We surmised that such a switch, if it occurred, might be so inconspicuous as to remain unreportable.

We performed preliminary experiments (Supplementary Fig. 1) to test whether these conditions cause binocular rivalry and, if they do, to investigate the conspicuity of perceptual switches. These experiments established that both conditions involve perceptual suppression of one eye's input at any time, as the dot density that observers perceived, estimated using a 2-interval categorization task, was close to that of a single eye's dot pattern rather than to that of both patterns superimposed. While this suggests binocular rivalry in both conditions, asking observers to press a key whenever they saw a perceptual switch, prompted an appreciable number of responses only in the *different colors* condition (see below for a quantitative analysis). These results together strengthened our supposition that the *same color* condition, by lacking a reliable cue to distinguish the two dot clouds, may involve unseen switches between the eyes' inputs. Our next experiment formally tested this notion, and provided an affirmative answer.

This experiment again included a *different colors* condition and a *same color* condition, and capitalized on known temporal regularities in bistable perception. Specifically, although switches during bistable perception occur at unpredictable intervals, the durations of multiple intervals experienced by a given observer jointly form a peaked distribution whose shape is reproducible and characteristic of that observer[24,27]. The idea of this experiment was to determine each observer's specific interval distribution based on the *different colors* condition, and then to verify that the same statistical regularities held in the *same color* condition. Such a correspondence would be expected if the *same color* condition involves

binocular rivalry, in spite of the near absence of perceptual switch reports for that condition in our pilot work.

But how can we determine the statistical properties of the alternation cycle without switch reports? A solution emerges from the realization that these stochastic properties can be expressed as temporal autocorrelation curves[23] rather than conventional interval distributions. For binocular rivalry these autocorrelation curves express the probability that the same eye's input is being perceived at two moments during the perceptual sequence, as a function of the time interval separating those moments. Under reasonable assumptions, percept duration distributions and autocorrelation curves are interchangeable and can be calculated from each other (Online Methods). To capitalize on autocorrelation curves in the *same color* condition, the quasi-random dot motion that characterizes our stimuli was interspersed with brief motion pulses: short periods during which one eye's dots all moved leftward while the other eye's dots moved rightward (Figure 1**b**). At each pulse the two directions were randomly assigned to the eyes, and observers reported which direction they perceived, allowing us to probe which eye's input dominated perception at those times. Autocorrelation curves were then constructed by calculating the probability that the same eye's input was perceived during two pulses as a function of the time interval between them.

Figure 1c, left, illustrates the approach by showing perceptual sequences of two observers (designated Obs. 2 and Obs 3.), reported in the *different colors* condition. Observer 2 appears to have longer percept durations on average than observer 3. Figure 1**c**, right, shows data for these observers in the *same color* condition, each marker corresponding to one motion pulse. Here observer 2 exhibits a stronger tendency than observer 3 for identical eye dominance across consecutive motion pulses. This is expected if the *same color* condition involves binocular rivalry, because observer 2 reported longer percept durations in the *different colors* condition. To make this comparison across conditions explicit, Figure 1**d** summarizes these observers' dominance sequences in terms of autocorrelation curves (left) and, equivalently, percept duration distributions (right), both derived from data such as those of Figure 1**c**. The across-condition correspondence in Figure 1**d** suggests that binocular rivalry occurs in both conditions. Figure 1**e**, left, extends this comparison to all 15 observers, showing the mean of each observer's inferred percept duration distribution for the *same color* condition (y-axis) and for the *different colors* condition (x-axis). Figure 1**e**, right, is analogous but displays distribution widths instead of means. The strong positive correlations and slopes close to unity indicate a close correspondence in temporal dynamics across conditions (see figure legend for statistics).

As reasoned above, these are the predicted results if the *same color* condition involves normal binocular rivalry and thus preserves each observer's signature percept duration distribution. These results are not explained, however, by other a priori plausible hypotheses. In particular, the hypothesis that rivalry in the *same color* condition only happens during, but not between, the motion pulses, predicts estimated percept durations in the *same color* condition and the *different colors* condition to differ by orders of magnitude[28] (a prediction verified in Supplementary Fig. 2) and to be uncorrelated[29]. Similarly, if observers in the *same color* condition respond randomly, autocorrelation curves should be flat and near 0.5, unlike the ones we observe. Thus, these data corroborate the

conclusion that the *same color* condition supports binocular rivalry with similar dynamics as the *different colors* condition. Next, the six observers whose data are labeled in Figure 1**e**, performed an experiment that tested formally the degree of reportability of dominance switches in both conditions.

This experiment was again similar, but involved a dual task: motion pulses now prompted observers to report, not only which motion direction they perceived, but also whether a switch in eye dominance had happened during the interval that preceded the motion pulse (Figure 2**a**; same task for both conditions). We timed the interval duration between motion pulses such, that intervals with more than one switch were exceedingly rare, leaving only intervals with zero or one switch (see Online Methods). The dual task then allowed us to identify hits (correctly identified switch intervals), misses (misidentified switch intervals), correct rejections (correctly identified non-switch intervals), and false alarms (misidentified non-switch intervals).

Figure 2**b** displays the results across observers. The two leftmost panels show a large between-condition difference in hit rate and an unaltered, low false alarm rate, indicating much poorer detectability of switches in the *same color* condition than the *different colors* condition (two-tailed paired t-test, hit rates: $t(5)=9.06$, $p=2.7 \cdot 10^{-5}$); false alarm rates: $t(5)=.85$, $p=.43$). The rightmost two panels express the same data in terms of sensitivity ($d'$) and criterion[30]. Based on the sensitivity index, $d'$, detection of switches in the *same color* condition could not be distinguished from chance, demonstrating just how inconspicuous these switches are (two-tailed t-test, $t(5)=1.59$, $p=.17$). At the same time, the index does not strongly support the null hypothesis that these switches were altogether undetectable (Bayes factor = 1.1 in favor of the hypothesis $d'=0$), so a more conservative interpretation, based on hit rates, is that observers detected perceptual switches in the *same color* condition about 8 times less frequently than in the *different colors* condition (left panel).

The identification of a condition where the large majority of switches between perceptual interpretations goes unnoticed would seem to pose a serious challenge to the popular hypothesis that such switches are caused by a process involved in the control of attention and response selection. It also provides a unique opportunity to test this hypothesis by investigating the extent to which executive brain areas remain involved when switches, by remaining unnoticed, draw minimal attention and have no behavioral relevance. Our next experiment therefore investigated the fMRI BOLD correlates of switches in these conditions.

## Brain responses to reportable and unreportable switches

The same six observers each reported the direction of motion pulses for both conditions (Figure 3**a**) during multiple fMRI sessions. Analysis of the resulting autocorrelation data again confirmed the occurrence of binocular rivalry in the *same color* condition (per-observer correlation with *different colors* condition: slope=.63, $r^2=.94$, $F(1,5)=78.7$, $p=3.0 \cdot 10^{-4}$ for means, and slope=.81, $r^2=.98$, $F(1,5)=318.8$, $p=1.0 \cdot 10^{-5}$ for widths). We again timed the motion pulses so that each interval contained either zero or one switch (Online Methods), and contrasted these two cases to assess BOLD concomitants of switches between the eyes' views.

Before turning to the results, be reminded of the *same color* condition's objective to isolate BOLD signals that correspond to the cause of perceptual switches, while minimizing signals that arise as a consequence. Existing work has also attempted to isolate this neural cause, by the inclusion of so-called 'replayed switches': on-screen events designed to mimic perceptual switches, yet lacking a neural cause[15–18,31]. Indeed, the conclusion that fronto-parietal brain areas are causally involved in bistability rests importantly on the fact that these areas show a BOLD response to endogenous perceptual switches (arguably reflecting both cause and consequences), even after subtracting out their response to replayed switches (reflecting consequences only). As detailed below, our paradigm offers an opportunity to critically evaluate the validity of this subtraction approach, and for this reason our *same color* condition also included replayed switches. These were simultaneous color changes in both eyes' dot patterns (Figure 3**a**) that mimicked the perceptual change associated with switches in the *different colors* condition, yet without the neural cause. Before including replayed switches here, we verified that they did not induce switches in eye dominance (see Online Methods), so that there was minimal correlation between replayed switches and actual switches in our design (for further control analyses see below).

Focusing first on the *different colors* condition, an across-observer contrast between switch intervals and no-switch intervals indicates elevated switch-related BOLD activity across a similar set of cortical areas as implicated previously (Figure 3**b**). In other areas it suggests reduced switch-related activity. Such reduced activity (perhaps related to the so-called *default network*[32] and to recent brain connectivity findings[33]) is not traditionally emphasized in this context, and is examined more closely in Supplementary Fig. 3.

Using an independent localizer we defined regions of interest (ROIs) and grouped these into three partially overlapping assemblies (Figure 3**b**, bottom inset), one delineating the right-hemispheric fronto-parietal network that is often the focus of studies on perceptual bistability[2,15] (rhFP) and two corresponding to a proposed functional subdivision between a dorsal and a ventral attention network[13] (dAN and vAN, respectively). The contrast of Figure 3b does not suggest critical areas beyond these ROIs, nor do additional whole-brain analyses (Supplementary Fig. 4). Focusing on the three ROI assemblies, Figure 3c, orange bars, quantifies the response difference between switch-intervals and no-switch intervals for the *different colors* condition, showing a significant difference in each case (two-tailed t-test, dAN: $t(5)=7.5$, $p=6.7 \cdot 10^{-4}$; rhFP: $t(5)=4.7$, $p=5.4 \cdot 10^{-3}$; vAN: $t(5)=6.0$, $p=1.8 \cdot 10^{-3}$). This demonstrates that, although our stimulus and task plausibly put some ongoing demand on executive functions, additional responses were elicited by perceptual switches.

Regarding the *same color* condition, we can imagine two alternative lines of reasoning, each leading to specific predictions. One line of reasoning centers on the assumption, made in previous work, that a causal component can be isolated from the response to salient perceptual switches by subtracting out the response to replayed switches (see above). As switches in our *same color* condition were also designed to isolate this causal component, the prediction is that the BOLD response to *same color* switches should resemble the response to *different colors* switches minus the response to replayed switches. Figure 3**c** illustrates this by displaying each assembly's response to replayed switches (purple bars; difference between intervals that contain a replayed switch and intervals that do not) as well

as the prediction (dark gray bars), formed by subtracting this response from the *different colors* response. The non-zero outcome of this subtraction (two-tailed t-test, dAN: $t(5)=3.7$, $p=.01$; rhFP: $t(5)=4.0$, $p=.01$; vAN: $t(5)=4.5$, $p=6.3\cdot10^{-3}$) would traditionally be interpreted as evidence for a causal role of these assemblies, but this interpretation is not supported by a comparison to the actual switch-related responses measured in our *same color* condition. This *same color* response (green bars) consistently undershoots the prediction from the subtraction method (two-tailed paired t-test, dAN: $t(5)=2.2$, $p=.08$; rhFP: $t(5)=3.0$, $p=.03$; vAN: $t(5)=2.9$, $p=.03$), suggesting that this method over-estimates the causal component of the switch-related BOLD response, and may thus lead to erroneous inferences of causality, a suggestion consistent with earlier work that provides possible reasons for this over-estimation[18].

Do the measured *same color* responses provide any support for a causal role of these brain areas? A second, alternative line of reasoning arises from this question. In some brain areas the switch-related BOLD response may exclusively reflect the consequences of perceptual switches, and if the *same color* condition successfully minimizes these consequences, then we expect a minimal *same color* BOLD response in these areas. A quantitative prediction, assuming linearity, is that the response to *same color* switches should be proportional to these switches' reportability, or about 8 times as small as the *different colors* response (cf. Figure 2**b**). This prediction (Figure 3**c**, light gray bars) does match the actual *same color* responses measured in the three assemblies (two-tailed paired t-test, dAN: $t(5)=.14$, $p=.89$; rhFP: $t(5)=.50$, $p=.64$; vAN: $t(5)=.61$, $p=.57$), suggesting that no part of these assemblies' BOLD response reflects a neural cause of perceptual switches.

To test in even more detail whether the *same color* responses differ from the prediction based on reportability alone, we repeated this latter comparison using linear mixed models, which offer a more powerful approach in this case (see Online Methods). This again provided no evidence for a difference (dAN: $p=.68$; rhFP: $p=.35$; vAN: $p=.25$), and indeed identified these data as substantial evidence against such a difference (Bayes factors of 8.6, 6.2 and 4.9 in favor of the null hypothesis, respectively).

Figure 3**d** shows results for the individual ROIs identified previously, with bar colors corresponding to those of Figure 3**c**, and dark and light lines replacing the dark and light gray bars, respectively. For each ROI the pattern of results confirms the data of Figure 3**c**: the *same color* response undershoots the prediction based on subtracting the replay-related response (linear mixed model, $p<.05$ in all cases except rhPFC, rhAG and lhIPS) but is indistinguishable from the prediction based on reportability alone (linear mixed model, $p>.14$ in all cases).

We considered two methodological explanations of these results. First, the *same color* condition's shorter mean percept duration (Figures 1**e**; 4**b**) could conceivably affect the amplitude of the estimated BOLD response. To assess this possibility we performed a simulation using data from the rhFP assembly (see Online Methods). Using deconvolution we first estimated across-observer BOLD response curves associated with perceptual switches in both conditions (Figure 4**a**), and confirmed that the areas under these curves differ by a factor of about 8 (Figure 4**b**, dots). To examine if percept duration differences

might account for this, we next created synthetic BOLD time series by convolving the *different colors* condition's estimated response curves with artifical event sequences, created while varying the simulated mean percept duration. We then assessed how the simulated mean percept duration influenced our ability to retrieve the switch-related response curve from these synthetic BOLD time series (using the same methods as used on the empirical data). As shown in Figure 4**b** (curve) the effect of simulated percept duration on the area under the retrieved curve was marginal in the relevant range, thus providing no explanation for the empirical results (Figure 4**b**, dots).

As a second methodological factor we considered the presence of replayed switches in the *same color* condition. This presence should not affect the estimated response to endogenous switches if one assumes different BOLD responses to combine linearly, but for further reassurance we performed an additional analysis. This 2-by-2 analysis assessed BOLD responses to four types of intervals in the *same-color* condition, categorized based on the presence or absence of endogenous switches, as well as the presence or absence of replayed switches. Although the analysis showed a main effect of replayed switches in all three ROI assemblies (repeated measures ANOVA, dAN: $F(1,5)=8.5$, $p=.03$; rhFP: $F(1,5)=8.1$, $p=.04$; vAN: $F(1,5)=10.0$, $p=.03$), it showed no main effect of endogenous switches (dAN: $F(1,5)=.37$, $p=.57$; rhFP: $F(1,5)=.73$, $p=.43$; vAN: $F(1,5)=1.6$, $p=.26$) and, importantly, no interaction (dAN: $F(1,5)=.22$, $p=.66$; rhFP: $F(1,5)=.66$, $p=.45$; vAN: $F(1,5)=1.36$, $p=.30$). This indicates that the estimated response to an endogenous switch is not affected by the presence of a replayed switch within the same interval. Indeed, when re-estimating the BOLD contrasts between intervals with and without a perceptual switch, this time including only intervals without a replayed switch (i.e. only two cells of the 2-by-2 analysis), the result does not differ from the original estimates of Figure 3**c** (paired t-test, dAN: $t(5)=-.12$, $p=.91$; rhFP: $t(5)=-.24$, $p=.82$; vAN: $t(5)=.25$, $p=.81$).

## Discussion

We designed a dichoptic display that causes perception to switch between the two eyes' views without those switches being noticed, showing that neural processes promoting perceptual bistability operate even when switches remain unreportable. Using this display, which minimizes the engagement of executive functions in response to perceptual switches, we estimated fMRI BOLD responses associated with the neural initiation of these switches. The results indicate that such responses in executive brain areas are overestimated by existing methods, and are altogether undetectable in our paradigm. Evidently, then, the rapid changes in neural response states that plausibly accompany perceptual switches[1] are not initiated by executive brain areas. Undetectable switches in perceptual states are compatible with empirical[8–10,34] and computational evidence[7,35] pointing to sensory areas of the brain as the loci for the neural processes govering perceptual switches. Several lines of evidence suggest that binocular rivalry and other forms of perceptual bistability may involve comparable neural processes[3,9–12,18]. It remains to be learned whether the present findings generalize to those other phenomena.

We focused on measures that index transient neural activity changes yoked to perceptual switches, but there are many other measures that suggest an executive contribution to the

perception of conflicting or ambiguous stimuli. For instance, the rate at which perception alternates when viewing these stimuli depends on the functional integrity of parietal[36–39] and frontal cortex[40, but see 41], and varies with differences in parietal anatomy[36,39,42]. Such findings, however, do not necessarily establish a causal link between perceptual switches and a general decision-making process, because they are also consistent with the notion that sensory regions give rise to perceptual switches, yet that their readiness to do so is influenced by non-sensory modulation of their response properties. Various authors have proposed such modulations in the form of top-down signals that predict sensory input[39] or that stabilize perception[44]. Such accounts might also explain why perceptual switches appear to involve an information flow from frontal to visual areas[16–17].

Switch-related fronto-parietal BOLD activity is reduced, but not eliminated, when dissociating switches from motor responses[19,43]. Our experiment reduced this activity to an undetectable level by furthermore removing awareness of the switches. A parsimonious conceptualization of these results frames awareness of sensory input as intimately related to the planning of motor actions, regardless of whether those actions are, in fact, executed[45–46]. In this view a perceptual change of which the observer is aware might be one that alters candidate motor plans or sensorimotor contingencies. This view also marries the present evidence against a driving role of fronto-parietal regions in perceptual switches to the notion that these regions do play a central role in visual awareness[21,43]: when viewing a conflicting or ambiguous stimulus, a switch in perception may arise within the visual system, but noticing the change may rely on brain regions dedicated to behavioral responses.

## Online Methods

### Observers

A total of 15 observers (seven female, age 23–36 years) participated in the psychophysical experiment of Figures 1, and six of them (one female) also participated in the experiments of Figures 2 and 3. The pilot experiments of supplementary Fig. 1 also includes six observers. All experiments were performed in agreement with the local ethics guidelines at Utrecht University and the University of Amsterdam. Two observers, coded *O2* and *O5* in the main text, were authors of this manuscript; the remaining observers were naive as to the purpose of the study.

### Stimuli

The rival stimuli in all experiments consisted of moving dots (radius=.08 degrees of visual angle (dva); density=138 dots/dva$^2$; speed=5.7 dva/s) presented within an annular aperture (inner radius=.15 dva, outer radius=1.25 dva) upon a dark background (minimum screen luminance). In the center of the aperture was a fixation mark (consisting of a white circular plateau, radius=.025 dva, surrounded by a Gaussian radial falloff to background luminance, σ=.03 dva). The stimulus was surrounded by a white ring (radius 2.9 dva) that was, in turn, surrounded by a pattern to aid fusion (random pixel values filling a square outline, side 7.0 dva). Stimuli were presented dichoptically, either on a CRT monitor viewed via a mirror stereoscope (psychophysics experiments) or projected using a DLP projector onto a screen viewed through prisms (fMRI/pupil experiment[47]). Everything was identical between the

two eyes except the centrally fixated dots that differed in directional content and, sometimes, color, differences that caused binocular rivalry.

With the exception of the periods during motion pulses (see below) the dots moved with a within-eye coherence of 0.4 except in the experiment where coherence was systematically varied (Supplementary Fig. 1), with 'coherence' indicating the proportion of dots moving in a single direction (signal dots) while the remaining dots all moved randomly (noise dots). Dot direction was determined per 300-ms interval: every 300 ms both the identity and the direction of signal dots within a given eye were randomly selected, as were the random directions of the remaining dots. The one constraint on signal dot direction was that one eye's signal dots always moved in a direction $+/-90°$ (randomly chosen) removed from the direction of the other eye's signal dots.

Motion pulses lasted 600 ms and were synchronized with the changes in dot motion described above (coinciding with two 300-ms intervals). At the start of a motion pulse dot speed remained unaltered but every dot's motion angle was randomly drawn from a uniform distribution (width 135°) centered on due leftward in one eye and on due rightward in the other (randomly assigned on each pulse). In addition, a black dot inside the fixation mark (radius=.013 dva) signaled the presence of a motion pulse throughout the 600-ms interval.

All dots in a given eye were either red-tinted or blue-tinted (randomly assigned on each trial). The luminance values of both tints, as well as that of white elements in the display, were approximately equal and based on the maximum blue output of the screen or projector. Specifically, per observer we used heterochromatic flicker photometry to estimate the red and white luminances equivalent to this maximum output, and created blue-tinted dots and red-tinted dots by blending a proportion $(1-p)$ of this white with a proportion $(p)$ of this blue or red, respectively. The value of $p$, the saturation, was determined per observer for both colors individually, using an experiment similar to the experiment illustrated in Figure 1b, but where dots were white and each pulse consisted, not of opposite directions of motion, but of blue being added to one eye's dots and red to the other eye's dots. Saturation levels were then set such that both colors were reported about equally often during these pulses, generally to a level of about $p=.3$ for both colors.

**Task**

Trials in the experiment reported in Figure 1 lasted 60 s. The *different colors* condition did not involve motion pulses, and observers reported switches in perception using three keys: one corresponding to each color and one corresponding to an incomplete mixture of both colors. The *same color* condition of that experiment involved motion pulses occurring randomly at one of four intervals ranging from 1.8 to 3.3 s (start to start). Observers reported motion direction using one of four keys that indicated, respectively, leftward motion, a mixture of motions dominated by leftward motion, a mixture dominated by rightward motion, and rightward motion. In the following, we will refer to the first and last of these options as 'complete' and to the other two as 'incomplete'.

One purpose of this experiment was to test whether rivalry dynamics during the *same eye* condition would be affected by replayed switches, as we planned to include such replayed

switches in our later fMRI sessions. For this reason the experiment included two versions of the *same color* condition: one without replayed switches and one where a replayed switch occurred midway 50% of the intervals between motion pulses (randomly selected). During a replayed switch, even though dots in both eyes all had the same color at any given moment, the color of all those dots simultaneously faded linearly from blue to red or vice versa over the course of 400 ms. Percept durations did not differ between these two versions of the *same color* condition in terms of mean (paired *t*-test, $t(14)=.25$; $p=.80$) or standard deviation (paired *t*-test, $t(14)=.005$; $p=.99$). Moreover, within the version where half of the intervals contained a replayed switch, those half of intervals did not differ from the other half in terms of the proportion of intervals that contained a genuine perceptual switch (repeated measures ANOVA, interval duration $F(3,42)=5.5$, $p=.003$; presence of replayed switch $F(1,14)=.19$, $p=.67$; interaction $F(3,42)=1.33$, $p=.28$). This indicates that replayed switches did not importantly affect rivalry dynamics, and data from both versions were pooled for Figure 1. Indeed, the correlation slopes reported there (.75 for means and .92 for widths; Figure 1e), would have been similar if only including the *same color* condition that did contain replayed switches (.76 and .94, respectively), or only the *same color* condition that did not (.76 and .91, respectively).

For the experiment of Figure 2, involving an analysis of detection performance, trials again lasted 60 s, and the interval duration between motion pulses was fixed at a value where the proportions of intervals with either zero or one perceptual switch were both expected to lie close to 50%, whereas intervals with more than one switch were expected to be rare (see section 'Behavioral data analysis' for details). As in the main experiments, motion direction could be reported as 'complete' or 'incomplete', and only intervals flanked by two reports of complete motion were included in the analysis.

For the fMRI experiment, trials lasted 90 s and motion pulses could again be separated by four possible interval durations. For each observer these were selected on the basis of that individual's rivalry dynamics, in order to minimize the occurrence of more than one perceptual switch in any given interval (total range 1.2–4.5 s start to start; see section 'Behavioral data analysis' for details). Replayed switches in the *same color* condition were presented at intervals drawn from a uniform distribution ranging from 1/2 to 3/2 times the observer's average percept duration, thus ensuring that the rates of replayed switches and actual switches were similar. The fMRI experiment included a third condition besides the *same color* condition and *different colors* condition described in the main text. This third condition served as an independent functional localizer of voxels from which to extract data for the area-specific analyses of Figures 3 and 4 (see Fig. 4). The condition involved the same stimulus sequence as the *different colors* condition, but each motion pulse prompted the observer to report, using one of four keys, whether zero, one, two or more perceptual switches had happened during the preceding interval (the latter two options were rarely used). In other words, perceptual switches here were both visible and task-relevant, ensuring a strong switch-related BOLD response. The BOLD contrast between switch intervals and no-switch intervals for this condition (Fig. 4**a**) yielded a result very similar to the analogous contrast for the *different colors* condition, shown in Figure 3**b**. The data shown in Figures 3**c**, 3**d** and 4 were extracted from voxels identified per observer by a significant outcome of

this contrast in the third condition. Each observer completed several sessions of the fMRI experiment, amounting to about 200 minutes of data per observer.

The experiment of Supplementary Fig. 1**a** involved 25-s trials without motion pulses. Six observers responded whenever they noticed a perceptual switch. As a test of the observers' tendency to provide unwarranted responses, the experiment also included a randomly intermixed condition where both eyes viewed the same stimulus (fusion), precluding perceptual switches. For the data of Supplementary Fig. 1**b** the same observers viewed pairs of intervals (2 s) separated by a blank (1 s). During one interval (rivalry) each eye viewed a different stimulus as in our main experiment, and during the other interval both eyes' stimuli were the same (fusion; the order was randomized). Observers reported which interval appeared to contain a higher density of dots, and the dot density of the non-rivalry stimulus was varied according to a Quest procedure[48] to find the physical density where the rivalry display and the fusion display were reported to contain the highest density equally often. This physical dot density could reasonably vary from the density of a single eye's pattern during binocular rivalry (y=1 in Fig. 1**b**), to twice that value if both patterns were perceptually superimposed (y=2 in Fig. 1**b**).

The experiment of Supplementary Fig. 2 was identical to Figure 1's *same color* condition except that the stimulus was removed from the screen during the intervals between motion pulses. This matches a well-characterized category of paradigms that involve intermittent presentation of a binocular rivalry stimulus, typically leading to repeated perception of the same eye's input across many presentations[28,49–50]. In a second control experiment that we aimed to include, rivalry between motion pulses was ruled out by temporarily making the stimulus fusible. This experiment yielded essentially only reports of mixed motion during the pulses, precluding an analysis centered on autocorrelations. Nevertheless, the high incidence of mixed perception, in itself, is inconsistent with the main text's *same color* condition, further corroborating the conclusions drawn from Supplementary Fig. 2.

## fMRI data acquisition

After giving written consent to participate, observers were placed supine in a Philips Achieva MRI scanner at the Spinoza Center of the University of Amsterdam. fMRI data were acquired using a standard SENSE EPI sequence with an isotropic resolution of 3 mm and a TR time of 2 s, covering practically the entire cortex with 37 slices and 80 by 80 in-plane voxels. Flip angle was 76.1° and TE was 27.63 ms. Data from the first five TRs of each scan were discarded to minimize T1 saturation effects. During scanning we recorded the pupil of one eyes at 1000 Hz using an Eyelink 1000 infrared eye tracker (SR Research Ltd., Ontario, Canada). Each observer completed three to four sessions of the same experiment, amounting to an average number of 36 scans per observer for both the *same color* condition and the *different colors* condition.

## Behavioral data analysis

When based on motion direction reports, autocorrelation curves were calculated after excluding intervals flanked by one or more reports of incomplete motion, because the question of whether a perceptual switch occurred cannot be answered reliably based on

reports of both eyes' motions mixed together. Converting a given, observed, percept duration distribution to the corresponding autocorrelation curve was done by simulation, randomly drawing many consecutive durations from the distribution and calculating the autocorrelation in the resulting simulated percept sequence. This approach is valid as long as the sequential dependence between consecutive binocular rivalry percept durations is negligible, so that individual percept durations can be treated as random samples from a stationary distribution. Existing work indeed shows this sequential dependence to be small[9,51–52]. To convert a given autocorrelation curve to the corresponding percept duration distribution we created an extensive bank of hypothetical autocorrelation curves, each constructed in the above-described fashion by drawing from a gamma distribution, but using different distribution parameters for each hypothetical autocorrelation curve. To find the duration distribution corresponding to a given empirical autocorrelation curve, we then searched the bank for the hypothetical curve that fit the empirical curve best.

Both for the experiment of Figure 2 and for the fMRI experiment of Figure 3, we tailored the interval durations between motion pulses per observer, to minimize the occurrence of intervals with more than one perceptual switch. This calculation again depended on the assumption that each percept duration is a random sample from a stationary distribution (see above), which allowed us to determine the probability that a randomly placed interval within the perceptual sequence contains a given number of perceptual switches. It is important to note that establishing a percept duration distribution on the basis of autocorrelation data does not require any assumptions regarding the number of switches separating two sampling moments (i.e. the autocorrelation curve, by indexing the probability that two moments yield matching eye dominance, quantifies the probability that the number of intervening switches is even; not whether that number is 0, 2, 4, etc.). Conversely, however, the percept duration distribution does predict the number of switches in a given interval. For the fMRI experiment we again verified post-hoc, on the basis of percept duration distributions estimated during the experiment, the proportion of intervals that inadvertently contained more than one switch. For each condition-observer combination, this proportion was smaller than 1%.

### fMRI data analysis

All fMRI preprocessing, including coregistration and motion compensation, was done using FSL[53], and we used Freesurfer[54] for conversion to the surface for visualization and delineation of brain areas. The general linear models used to produce the per-area results were implemented using FSL's FEAT tool, after which per-scan parameter estimates were extracted from functionally defined voxels (see above) that we separated into individual brain areas by manual delineation based on anatomical landmarks. For illustrations of across-observer contrasts (Figure 3**b**, Supplementary Figs. 3 and 4) each observer's per-scan FEAT results were entered into an across-scan fixed effects analysis using FSL's FEAT tool, and the resulting per-observer data were then converted to an average surface for a random-effects analysis across observers using Freesurfer's mri_glmfit.

For the *different colors* condition, the model included regressors for stimulus onset and offset, motion pulses, key presses, and intervals between motion pulses. These latter events

were modeled in a number of separate regressors, depending on whether a switch occurred during the interval, but also depending on the number of reports of incomplete motion that flanked the interval. Specifically, just like the behavioral results (see above), all fMRI results only concern intervals that were flanked by two reports of complete motion. This avoided the uncertainty caused by incomplete motion reports when it comes to inferring the occurrence of switches at the analysis stage, and it also minimized the potential influence of any BOLD signal associated with the difficulty of selecting a behavioral response when perception is mixed. The model also included regressors for two event types identified in the eye traces we recorded, namely blinks and microsaccades[55]. Both had a weak tendency to be more numerous in intervals that contained a perceptual switch than in intervals that did not[cf.56–57]. All events were modeled as instantaneous, except interval events, which were modeled as boxcars that had the durations of the associated intervals and amplitudes that were scaled down to retain a norm of 1. This scaling is consistent with the notion that a switch, if it occurs during an interval, does not fill the entire interval but takes a particular time that is independent of interval duration.

For the *same color* condition we used the same model, but added regressors to accommodate the replayed switches. Specifically, the model used to estimate the BOLD contrast between intervals with and without a replayed switch (Figures 3**c** and 3**d**, purple bars) included regressors for intervals with and without such a switch, again separated on the basis of the number of flanking reports of exclusive perception. To optimally estimate the contrast between intervals with and without an endogenous eye-dominance switch in the *same color* condition, we assessed that contrast using a model where replayed switches were not modeled as intervals, but as instantaneous events, aligned with the actual moments of replayed switches. The precise treatment of replayed switches turned out not to matter, however, as shown by the 2×2 analysis described in the main text. These were obtained from a model in which we modeled four kinds of intervals: ones with neither an endogenous switch nor a replayed switch, ones with either of those switches, and ones with both.

For the deconvolution analysis of Figure 4**a** we first followed the same preprocessing steps and per-observer general linear model approach as used for the main analyses of Figure 3, but left the regressors for intervals between motion pulses out of the model. Deconvolution was then performed per observer on the residuals of this analysis step, which should no longer contain the response components associated with the events included in the general linear model, while still containing the interval-related response components. The y-axis units of Figure 4**a** are z-score units of this residual signal. In the deconvolution analysis, as in the main analysis, intervals flanked by complete motion pulses were modeled as distinct from intervals flanked by one or two incomplete motion pulses, and all intervals were modeled as boxcar regressors that had the durations of the associated intervals and amplitudes that were scaled down to retain a norm of 1. Because these boxcar regressors caused considerable collinearity in the design matrix, we used ridge regression implemented with the Python Scikit-learn class RidgeCV[58] at standard settings (which selects complexity parameter $\alpha$ from among the values .1, 1 and 10 based on cross validation). The curves of Figure 4**a** show the estimated across-observer response to intervals with one switch (flanked

by complete motion pulses) minus the estimated response to intervals with zero switches (again flanked by complete motion pulses).

For the simulations of Figure 4**b** we first calculated an average percept duration distribution across observers based on the fMRI experiment's *different colors* condition. To this end we constrained a gamma distribution to the across-observer averaged mean percept duration, and to the across-observer averaged standard deviation in percept duration (calculated per observer and then averaged). For each simulation setting we adjusted the resulting across-observer distribution by scaling down the mean as well as the standard deviation. For Figure 4**b** the standard deviation was scaled down more slowly than the mean, in accordance to the empirical finding that the *same color* condition's mean percept duration during the fMRI sessions was 0.74 that of the *different colors* condition, whereas the corresponding standard deviations differed only by a factor 0.87. We also performed simulations where mean and standard deviation were scaled at the same rate, thus maintaining the same distribution shape, and obtained similar results. To create synthetic BOLD time series (of a length equal to the average length of the per-observer *same-color* BOLD time series) we first convolved Figure 4**a**'s orange curve with a time series of switch moments randomly generated using the relevant gamma distribution. We then added the result to a randomly selected section of the residuals that had remained after performing deconvolution on an observer's empirical data, in order to match the noise level in the simulations to that in the actual data. To perform deconvolution on the synthetic BOLD time series we created the same four regressors as used for deconvolving the empirical data: intervals with and without a switch, and flanked by either zero or more incomplete motion pulses. To this end we calculated the across-observer average of the four interval durations that could separate consecutive motion pulses, and lay a random sequence of motion pulses, each consecutive pair separated by one of these four durations, alongside each randomly generated sequence of perceptual switches. Each interval between motion pulses was classified as containing either zero switches if the number of switch moments between the midpoints of flanking motion pulses was even, or a single switch if this number was odd (thus matching the procedure followed with the empirical data, and allowing misclassification of intervals that contain more than one switch). Moreover, an interval was classified as flanked by incomplete motion pulses if at least one switch moment coincided with either of the flanking motion pulses. The curve of Figure 4**b** was calculated by, for each simulated mean percept duration, dividing the area under the deconvolved switch-related curve by the corresponding area obtained in simulations that used the unscaled across-observer duration distribution.

## Statistical testing

Most statistical testing relied on t-tests and repeated measures ANOVAs. For every reported t-test we verified normality of the test data (Kolmogorov-Smirnov $p>.05$). For most repeated measures ANOVAs sphericity testing did not come into play because the factors had only two levels, the only exception being the ANOVA reported in the Online Methods paragraph 'Task', which included four levels of interval duration. In this case sphericity was not violated (Mauchly's test $p>.54$).

For Figure 1e we performed a bootstrap analysis on the data of each individual observer, to test whether the resemblance between that observer's autocorrelation curve in the *same color* condition and his/her curve in the *different colors* condition was larger than expected by chance. On each iteration we randomly shuffled the sequence of eye dominance states (left, mostly left, mostly right, right) associated with motion pulse reports during the *same color* condition, and calculated the (sum of squares) distance between the resulting autocorrelation curve and the autocorrelation curve of the *different colors* condition. These simulated sequences, in other words, maintained the overall proportion of eye dominance reports that a given observer gave in response to motion pulses, but assumed that these reports occurred in random order. We then calculated the proportion of iterations that yielded a distance that was smaller than the observed distance between the two conditions' actual autocorrelation curves. For each observer the reported *p*-value indicates this proportion, which can be interpreted as the probability that the observed resemblance would be observed by chance. For each observer the bootstrap procedure involved a total of 500 randomly reshuffled sequences, putting a lower limit on the measurable *p*-value at .002.

The Bayes factor associated with Figure 2 was calculated using the function *ttestBF* of R's BayesFactor package[59].

The linear mixed models associated with Figures 3**c** and 3**d** were implemented using R. The models included a random intercept across observers, and included 'prediction vs. actual observation' as a fixed effect. Although the BOLD parameter estimates that underlie these figures were extracted for each scan individually (see above), the associated t-tests reported in the main text were (necessarily) run on per-observer average values across all scans, and therefore did not take into account any measure of within-observer reliability. The main reason for the added power of the linear mixed model approach in this case is that observers ran many scans for each condition (36 on average) and that this approach takes into account each per-scan BOLD contrast individually, thus capitalizing on the power gained by the great number of per-observer repetitions. *p*-values were obtained by implementing the models using the *lmer* function of R's lme4 package[60], and then comparing models with and without the fixed effect using the *anova* function. The associated Bayes factors were calculated by implementing the models using the *lmBF* function of R's BayesFactor package[59] and again comparing models with and without the fixed effect.

All figures that illustrate fMRI results in terms of whole-brain contrasts (Figure 3**b**, Supplementary Figs. 3 and 4) are based on across-observer random-effects analyses performed using Freesurfer's mri_glmfit, without correction for multiple comparisons.

No statistical methods were used to pre-determine sample sizes but rather our sample sizes were based on precedents in existing literature. In particular, the number of observers for the fMRI experiments was initially chosen to be four, on the grounds that, first, existing studies in this field had obtained robust results using similar sample sizes[15,18] and, second, the difficulty of the in-scanner task favored a design that capitalized on a large number of per-observer repetitions and a small number of experienced observers. The observer number was later increased to six after referees expressed concern about statistical power. The number of observers for the behavioral experiments was based on the anticipation that an observer-to-

observer correlation in mean percept duration would play an important role (Figure 1**e**). Previous work had shown robust correlations of this type using similar participant numbers[10].

We did not employ randomization in our experiment, in the sense that our within-observer design did not involve randomized assignment to groups, and the experiment was not blind, in the sense that the experimenter was aware which condition was being performed.

A supplementary methods checklist is available.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## Bibliography main text

1. Blake R, Logothetis N. Visual competition. Nature Reviews Neuroscience. 2002; 3(1):13–21. [PubMed: 11823801]

2. Sterzer P, Kleinschmidt A, Rees G. The neural bases of multistable perception. Trends in Cognitive Sciences. 2009; 13(7):310–318. [PubMed: 19540794]

3. Brascamp J, Klink PC, Levelt WJM. with a contribution from. The 'laws' of binocular rivalry: 50 years of Levelt's propositions. Vision Research. 2015; 109:20–37. [PubMed: 25749677]

4. Maruya K, Yang E, Blake R. Voluntary action influences visual competition. Psychol. Sci. 2007; 18(12):1090–1098. [PubMed: 18031417]

5. Dieter KC, Tadin D. Understanding attentional modulation of binocular rivalry: a framework based on biased competition. Front. Hum. Neurosci. 2011; 5(155):1–12. [PubMed: 21283556]

6. Paffen CLE, Alais D. Attentional modulation of binocular rivalry. Front. Hum. Neurosci. 2011; 5(105):1–10. [PubMed: 21283556]

7. Wilson H. Minimal physiological conditions for binocular rivalry and rivalry memory. Vision Research. 2007; 47(21):2741–2750. [PubMed: 17764714]

8. Alais D, Cass J, O'Shea R, Blake R. Visual Sensitivity Underlying Changes in Visual Consciousness. Current Biology. 2010; 20(15):1362–1367. [PubMed: 20598538]

9. Pastukhov A, Braun J. Cumulative history quantifies the role of neural adaptation in multistable perception. Journal of Vision. 2011; 11(10):12–12. [PubMed: 21931128]

10. van Loon A, Knapen T, Scholte H, St John-Saaltink E, Donner T, Lamme V. GABA Shapes the Dynamics of Bistable Perception. Current Biology. 2013; 23(9):823–827. [PubMed: 23602476]

11. Leopold D, Logothetis N. Multistable phenomena: changing views in perception. Trends in Cognitive Sciences. 1999; 3(7):254–264. [PubMed: 10377540]

12. Einhäuser W, Stout J, Koch C, Carter O. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(5):1704–1709. [PubMed: 18250340]

13. Corbetta M, Patel G, Shulman G. The reorienting system of the human brain: from environment to theory of mind. Neuron. 2008; 58(3):306–324. [PubMed: 18466742]

14. Silver M, Kastner S. Topographic maps in human frontal and parietal cortex. Trends in Cognitive Sciences. 2009; 13(11):488–495. [PubMed: 19758835]

15. Lumer E, Friston K, Rees G. Neural correlates of perceptual rivalry in the human brain. Science. 1998; 280(5371):1930–1934. [PubMed: 9632390]

16. Sterzer P, Kleinschmidt A. A neural basis for inference in perceptual ambiguity. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(1):323–328. [PubMed: 17190824]

17. Weilnhammer V, Ludwig K, Hesselmann G, Sterzer P. Frontoparietal Cortex Mediates Perceptual Transitions in Bistable Perception. Journal of Neuroscience. 2013; 33(40):16009–16015. [PubMed: 24089505]

18. Knapen T, Brascamp J, Pearson J, Van Ee R, Blake R. The Role of Frontal and Parietal Brain Areas in Bistable Perception. Journal of Neuroscience. 2011; 31(28):10293–10301. [PubMed: 21753006]

19. Frässle S, Sommer J, Jansen A, Naber M, Einhäuser W. Binocular rivalry: Frontal activity relates to introspection and action but not to perception. Journal of Neuroscience. 2014; 34(5):1738–1747. [PubMed: 24478356]

20. Zaretskaya N, Narinyan M. Introspection, attention or awareness? The role of the frontal lobe in binocular rivalry. Frontiers in Human Neuroscience. 2014; 8:1–2. [PubMed: 24474914]

21. Safavi S, Kapoor V, Logothetis N, Panagiotaropoulos I. Is the frontal lobe involved in conscious perception? Front. Psychol. 2014; 5:1–2. [PubMed: 24474945]

22. Zhang P, Jamison K, Engel S, He B, He S. Binocular rivalry requires visual attention. Neuron. 2011; 71:362–369. [PubMed: 21791293]

23. Brascamp JW, Blake R. Inattention abolishes binocular rivalry: perceptual evidence. Psychol. Sci. 2012; 23(10):1159–1167. [PubMed: 22933458]

24. Brascamp J, van Ee R, Pestman W, van den Berg A. Distributions of alternation rates in various forms of bistable perception. Journal of Vision. 2005; 5(4):287–298. [PubMed: 15929652]

25. Klink P, Van Ee R, Van Wezel R. General validity of Levelt's propositions reveals common computational mechanisms for visual rivalry. PLoS ONE. 2008; 3(10):e3473. [PubMed: 18941522]

26. Blake R, Zimba L, Williams D. Visual motion, binocular correspondence and binocular rivalry. Biological Cybernetics. 1985; 52(6):391–397. [PubMed: 4052502]

27. Levelt, WJM. On binocular rivalry. Soesterberg, The Netherlands: Institute for Perception RVO-TNO; 1965.

28. Leopold D, Wilke M, Maier A, Logothetis N. Stable perception of visually ambiguous patterns. Nat. Neurosci. 2002; 5(6):605–609. [PubMed: 11992115]

29. Brascamp J, Pearson J, Blake R, Van den Berg A. Intermittent ambiguous stimuli: implicit memory causes periodic perceptual alternations. Journal of Vision. 2009; 9(3):3–3. [PubMed: 19757942]

30. Green, D.; Swets, J. Signal detection theory and psychophysics. New York: Wiley; 1966.

31. Tong F, Nakayama K, Vaughan J, Kanwisher N. Binocular rivalry and visual awareness in human extrastriate cortex. Neuron. 1998; 21(4):753–759. [PubMed: 9808462]

32. Buckner R, Andrews-Hanna J, Schacter D. The brain's default network. Anatomy, function and relevance to disease. Ann. N.Y. Acad. Sci. 2008; 1124:1–38. [PubMed: 18400922]

33. Baker D, Karapanagiotidis T, Coggan D, Wailes-Newson K, Smallwood J. Brain networks underlying bistable perception. NeuroImage. 2015; 119:229–234. [PubMed: 26123379]

34. Blake R, Sobel K, Gilroy L. Visual motion retards alternations between conflicting perceptual interpretations. Neuron. 2003; 39(5):869–878. [PubMed: 12948452]

35. Moreno-Bote R, Rinzel J, Rubin N. Noise-induced alternations in an attractor network model of perceptual bistability. Journal of Neurophysiology. 2007; 98(3):1125–1139. [PubMed: 17615138]

36. Kanai R, Bahrami B, Rees G. Human parietal cortex structure predicts individual differences in perceptual rivalry. Current Biology. 2010; 20:1626–1630. [PubMed: 20727757]

37. Carmel D, Walsh V, Lavie N, Rees G. Right parietal TMS shortens dominance durations in binocular rivalry. Current Biology. 2010; 20(18):R799–R800. [PubMed: 20869603]

38. Zaretskaya N, Thielscher A, Logothetis N, Bartels A. Disrupting Parietal Function Prolongs Dominance Durations in Binocular Rivalry. Current Biology. 2010; 20(23):2106–2111. [PubMed: 21093263]

39. Kanai R, Carmel D, Bahrami B, Rees G. Structural and functional fractionation of right superior parietal cortex in bistable perception. Current Biology. 2011; 21(3):R106–R107. [PubMed: 21300270]

40. Ricci C, Blundo C. Perception of ambiguous figures after focal brain lesions. Neuropsychologia. 1990; 28(11):1163–1173. [PubMed: 2290491]

41. De Graaf T, De Jong M, Goebel R, Van Ee R, Sack A. On the functional relevance of frontal cortex for passive and voluntarily controlled bistable vision. Cerebral Cortex. 2011; 21(10):2322–2331. [PubMed: 21385836]

42. Watanabe T, Masuda N, Megumi F, Kanai R, Rees G. Energy landscape and dynamics of brain activity during bistable perception. Nature Communications. 2014; 5:4765.

43. Lumer E, Rees G. Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96(4):1669–1673. [PubMed: 9990082]

44. Kloosterman N, Meindertsma T, Hillebrand A, Van Dijk B, Lamme V, Donner T. Top-down modulation in human visual cortex predicts the stability of a perceptual illusion. J. Neurophys. 2014; 113(4):1063–1076.

45. Cotterill R. On the mechanism of consciousness. Journal of Consciousness Studies. 1997; 4(3): 231–247.

46. O'Regan J, Noë A. A sensorimotor account of vision and visual consciousness. The Behavioral and Brain Sciences. 2001; 24(5):939–1031. [PubMed: 12239892]

## Methods-only bibliography

47. Schurger A. A very inexpensive MRI-compatible method for dichoptic visual stimulation. Journal of neuroscience methods. 2009; 177(1):199–202. [PubMed: 18973774]

48. Watson A, Pelli D. QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 1983; 33(2):113–120. [PubMed: 6844102]

49. Pastukhov A, Braun J. A short-term memory of multi-stable perception. Journal of Vision. 2008; 8(13):7–7. [PubMed: 19146337]

50. Pearson J, Brascamp J. Sensory memory for ambiguous vision. Trends in Cognitive Sciences. 2008; 12(9):334–341. [PubMed: 18684661]

51. Fox R, Herrmann J. Stochastic properties of binocular rivalry alternations. Perc. Psychophys. 1967; 2:432–436.

52. Van Ee R. Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: evidence from serial correlations in perceptual bistability. J. Opt. Soc. Am. A. 2009; 26(12):2612–2622.

53. Smith S, Jenkinson M, Woolrich M, Beckmann C, Behrens T, Johansen-Berg H, Bannister P, De Luca M, Drobnjak I, Flitney D, Niazy R, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady J, Matthews P. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004; 23(Suppl 1):S208–S219. [PubMed: 15501092]

54. Dale A, Fischl B, Sereno M. Cortical surface-based analysis. I. segmentation and surface reconstruction. Neuroimage. 1999; 9:179–194. [PubMed: 9931268]

55. Rolfs M. Microsaccades: small steps on a long way. Vision Research. 2009; 49:2415–2441. [PubMed: 19683016]

56. Van Dam L, Van Ee R. Retinal image shifts, but not eye movements per se, cause alternations in awareness during binocular rivalry. Journal of Vision. 2006; 6:1172–1179. [PubMed: 17209727]

57. Kalisvaart J, Goossens J. Influence of retinal image shifts and extra-retinal eye movement signals on binocular rivalry alternations. PLoS ONE. 2013; 8(4):e61702. [PubMed: 23593494]

58. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay

E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.

59. Morey R, Rouder J. *BayesFactor*: An R package for computing Bayes factor for a variety of psychological research designs (available on the Comprehensive R Network). 2012

60. Bates D, Maechler M, Bolker B, Walker S. *lme4*: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014 http://CRAN.R-project.org/package=lme4.
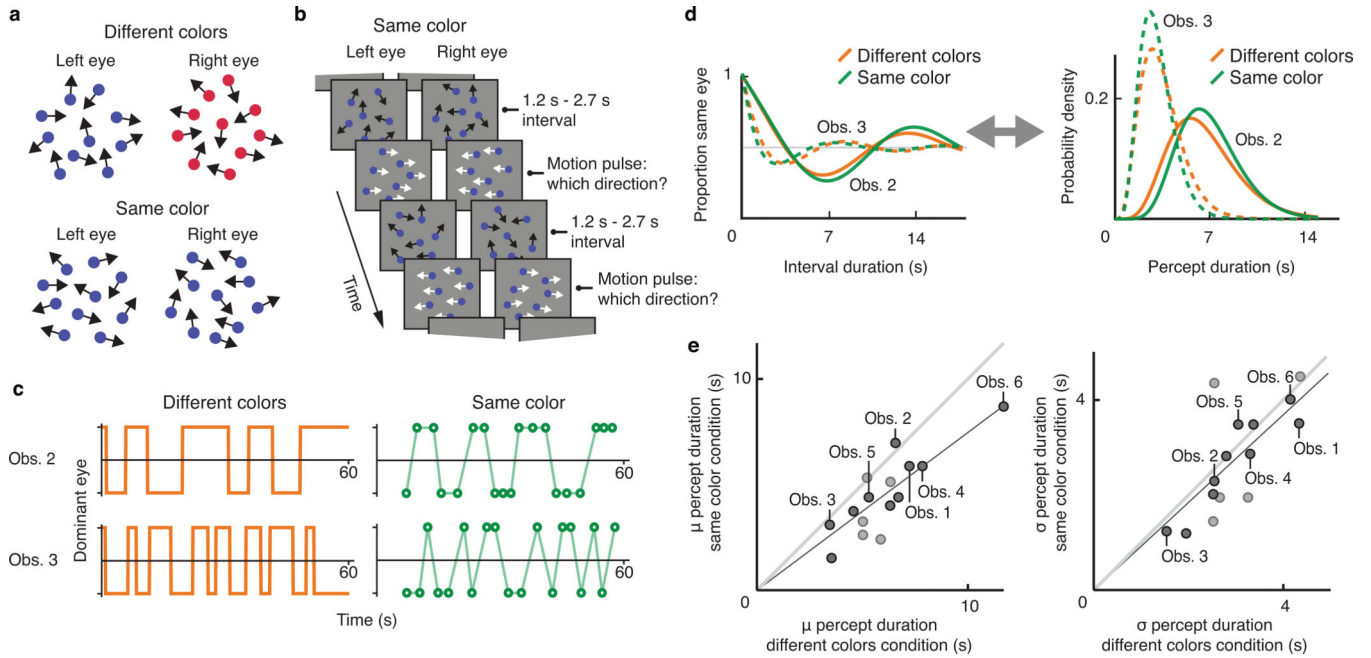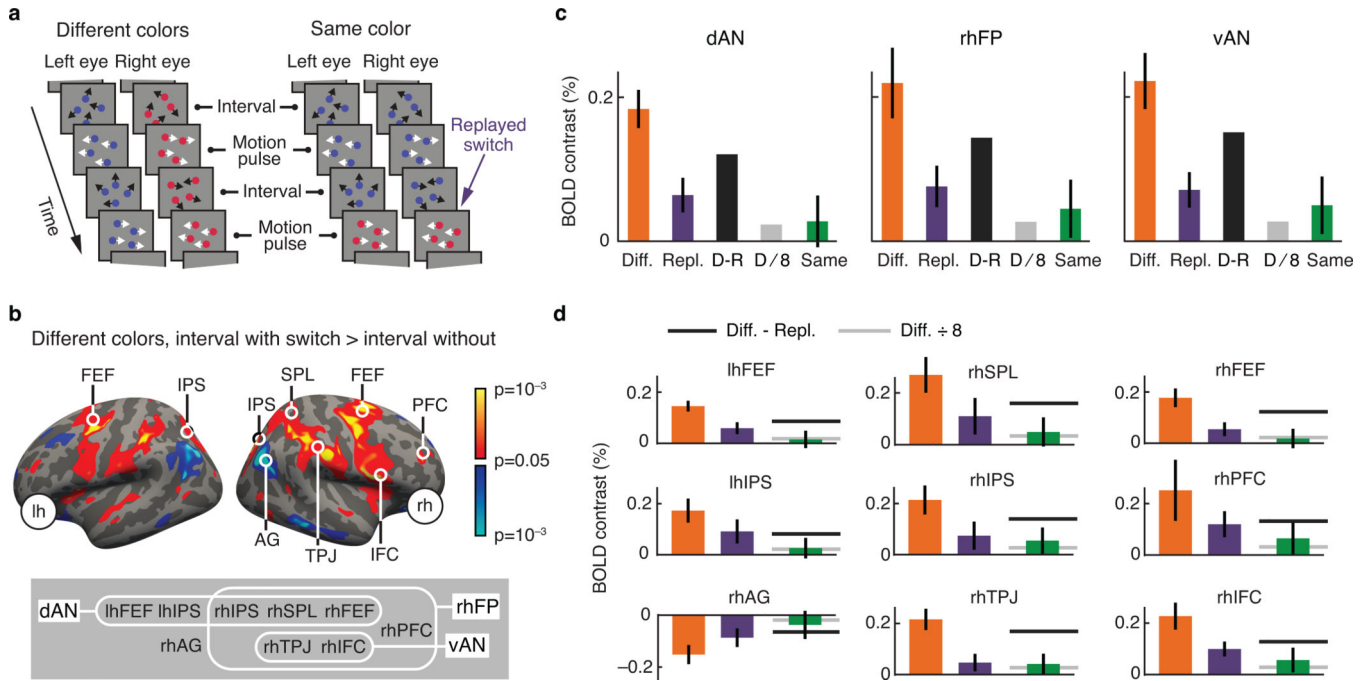
**Figure 1.**

Psychophysical experiment investigating stochastic properties of the perceptual sequence. **a**. Schematic illustration of the visual stimuli used throughout all experiments in this study. **b**. Trial sequence in the *same color* condition of the first psychophysical experiment. c. Examples of perceptual sequences as reported by two individual observers, designated Obs. 2 and Obs. 3, in the *different colors* condition (left, orange) and the *same color* condition (right green). **d**. Autocorrelation curves (left) and percept duration distributions (right) for the same two observers in the *different colors* condition (orange) and in the *same color* condition (green), inferred from data such as shown in **c**. **e**. Left panel: 15 individual observers' estimated mean percept duration in the *different colors* condition (x-axis) and in the *same color* condition (y-axis). Linear regression indicates a strong positive correlation: slope=.75, $r^2$=.96, $F(1,14)$=313, $p<1\cdot10^{-6}$. The two different dot shades represent observers whose autocorrelation data in the *same color* condition are (dark dots) or are not (light dots) significantly different, on an individual-observer basis, from what would be predicted from random button presses (bootstrap analysis; dark dots: $p$=.004, .008 or $p<$.002, the latter being the bootstrap analysis' smallest measurable $p$ value; light dots: $p$=.07, .14, .22, .28 and .64); see Online Methods). Right panel: same as left but depicting estimated standard deviations of the percept duration distributions. Linear regression again indicates a strong correlation: slope=.92, $r^2$=.94, $F(1,14)$=284, $p<1\cdot10^{-6}$.
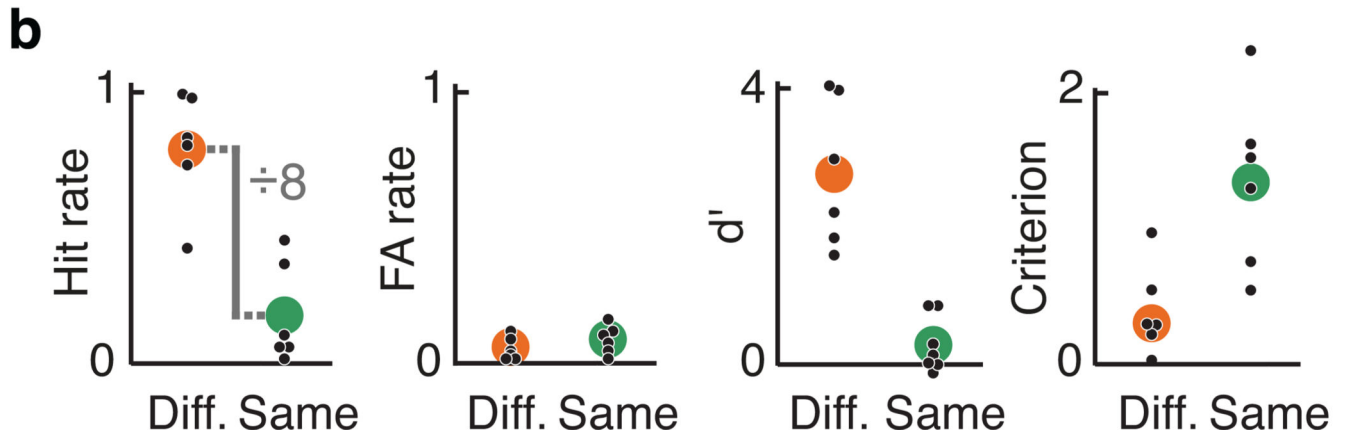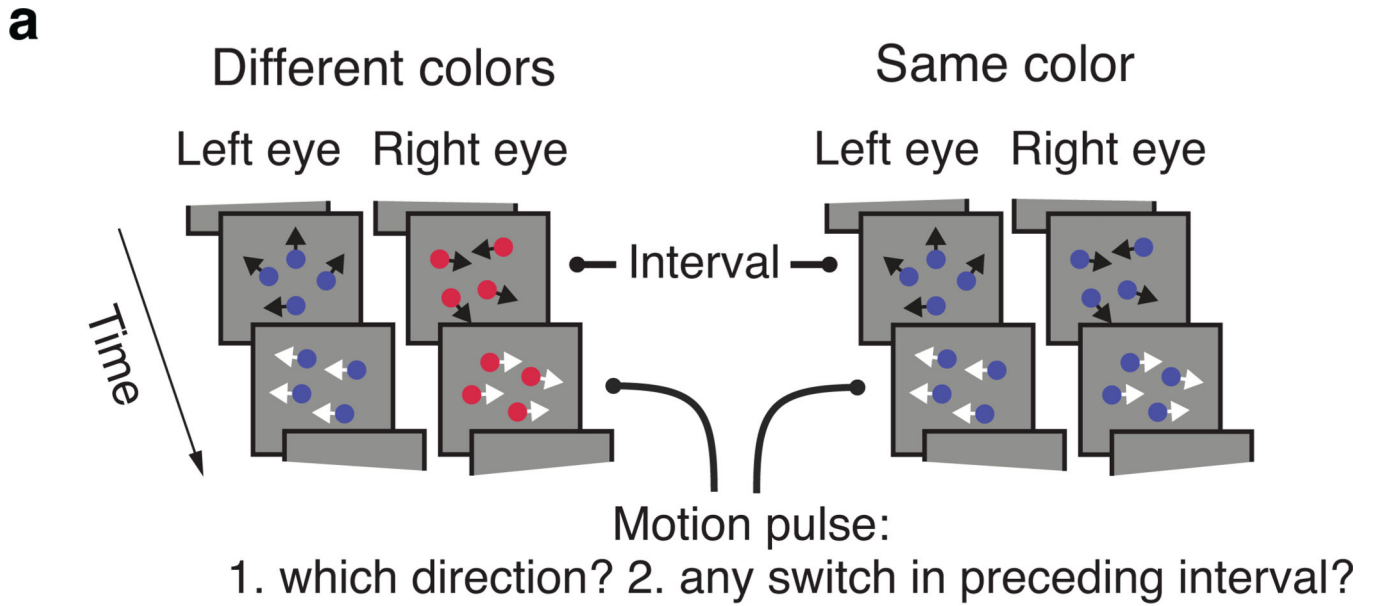
**Figure 2.**
Psychophysical experiment assessing the detectability of perceptual switches. **a**. Trial sequence. Both conditions now contained motion pulses, and six observers were required to provide a two responses to each motion pulse. **b**. Signal detection analysis of detectability of perceptual switches in the *different colors* condition ('Diff', orange) and in the *same color* condition ('Same', green), in terms of hit rate (leftmost panel) and false alarm rate (FA rate, second panel) and, equivalently, in terms of sensitivity (d', third panel) and criterion (right panel). Hit rates differs significantly between conditions (two-tailed paired t-test, $t(5)=9.06$, $p=2.7 \cdot 10^{-5}$), but false alarm rates do not (two-tailed paired t-test, $t(5)=.85$, $p=.43$). d' does not significantly differ from 0 (two-tailed t-test, $t(5)=1.59$, $p=.17$) but the hypothesis that it equals zero is not strongly supported either (Bayes factor = 1.1 in favor of the hypothesis $d'=0$).
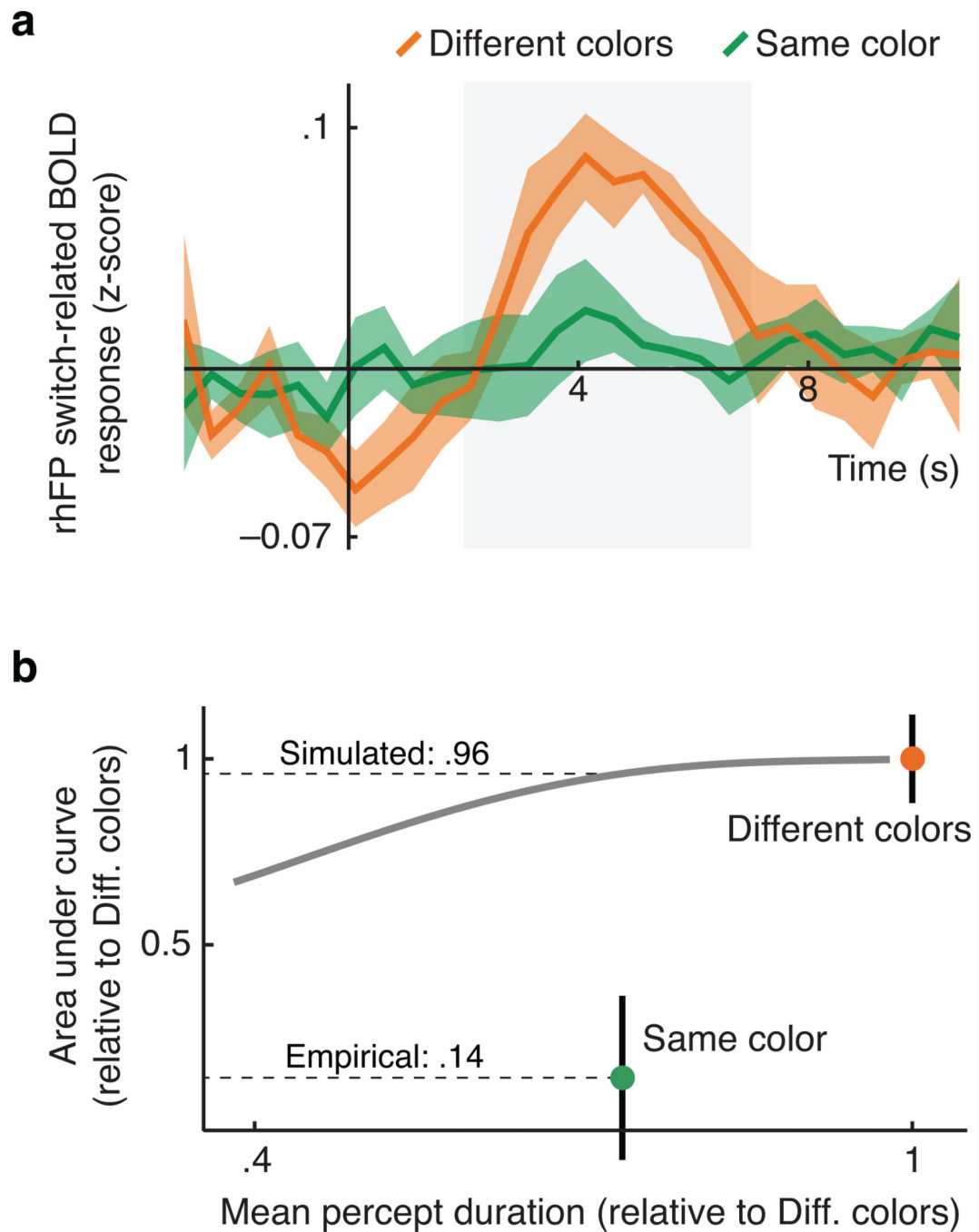
**Figure 3.**
Functional imaging results: general linear model. **a**. Both conditions again contained motion pulses. The *same color* condition also contained 'replayed switches': instances where both eyes' dot patterns simultaneously faded from one color to the alternative color. **b**. fMRI BOLD contrast across six observers for the *different colors* condition, between intervals inferred to contain a perceptual switch and intervals inferred to contain no perceptual switch. lh=left hemisphere, rh=right hemisphere, FEF=frontal eyefields, IPS=intraparietal sulcus, SPL=superior parietal lobule, AG=angular gyrus, TPJ=temporoparietal junction, IFC=inferior frontal cortex, PFC=prefrontal cortex. Inset: grouping of regions of interest (ROIs) within three overlapping assemblies. dAN=dorsal attention network, rhFP=right hemisphere fronto-parietal cortex, vAN=ventral attention network. **c**. BOLD contrasts across six observers, extracted from each of the three ROI assemblies, between intervals with an without a perceptual switch in the *different colors* condition (*Diff.*, orange; significant in each case: two-tailed t-test, dAN: $t(5)=7.5$, $p=6.7\cdot10^{-4}$; rhFP: $t(5)=4.7$, $p=5.4\cdot10^{-3}$; vAN: $t(5)=6.0$, $p=1.8\cdot10^{-3}$), between intervals with and without a replayed switch in the *same color* condition (*Repl.*, purple), and between intervals with and without a perceptual switch in the *same color* condition (*Same*, green). The two gray bars in each panel represent predictions for the *same color* contrast, based on two alternative lines of reasoning (D=*different colors* contrast; R=*replayed switches* contrast; see text for details). The *same color* data consistently differ from the *D-R* prediction (two-tailed paired t-test, dAN: $t(5)=2.2$, $p=.08$; rhFP: $t(5)=3.0$, $p=.03$; vAN: $t(5)=2.9$, $p=.03$), but are not distinguishable from the *D/8* prediction (two-tailed paired t-test, dAN: $t(5)=.14$, $p=.89$; rhFP: $t(5)=.50$, $p=.64$; vAN: $t(5)=.61$, $p=.57$; and linear mixed model, dAN: $p=.68$, Bayes factor $K=8.6$; rhFP: $p=.35$, Bayes factor $K=6.2$; vAN: $p=.25$, Bayes factor $K=4.9$). **d**. The same three BOLD contrasts as in **c**, now for individual regions of interest (ROIs). Gray lines reflect the same two predictions as the gray bars in c (*Diff.*=*different colors* contrast; *Repl.*=replayed

switches contrast). The *same color* response undershoots the prediction *D-R* prediction for nearly all ROIs (linear mixed model, lhFEF: $p=.02$; rhSPL: $p=5.7 \cdot 10^{-3}$; rhFEF: $p=2.6 \cdot 10^{-4}$; lhIPS: $p=.25$; rhIPS: $p=.03$; rhPFC: $p=.05$; rhAG: $p=.48$; rhTPJ: $p=6.5 \cdot 10^{-5}$; rhIFC: $p=.04$) but is indistinguishable from the *R*/8 prediction in all cases (linear mixed model, lhFEF: $p=.86$; rhSPL: $p=.32$; rhFEF: $p=.97$; lhIPS: $p=.57$; rhIPS: $p=.39$; rhPFC: $p=.21$; rhAG: $p=.67$; rhTPJ: $p=.46$; rhIFC: $p=.14$). Error bars in all panels are standard errors of the sample mean across observers.

**a**



**b**



**Figure 4.**

Functional imaging results: deconvolution, and simulation of the influence of percept duration. **a**. BOLD time courses across six observers in the right hemisphere fronto-parietal ROI assembly, in response to perceptual switches in the *different colors* condition (orange) and in the *same color* condition (green). The gray shaded area indicates the time range (2 s to 7 s) used for determining the area under each curve, depicted in the next panel. **b**. Area under the switch-related BOLD response curve as a function of mean percept duration. The colored dots show the empirical results for the two conditions: the *same color* condition

yielded an area 0.14 times that obtained in the *different colors* condition. The gray curve shows the area obtained from synthetic BOLD data, created using the empirical responses from the *different colors* condition but assuming shorter percept durations. An increase in percept duration only marginally reduces the area estimate (to a factor 0.96 in the relevant range), and therefore cannot explain the *same color* condition's empirical result. Error bars in both panels are standard errors of the sample mean across observers.