

# Switch rates for orthogonally oriented kinetic-depth displays are correlated across observers

**Alexander Pastukhov**

Department of General Psychology and Methodology,  
University of Bamberg, Bamberg, Bavaria, Germany  
Forschungsgruppe EPÆG (Ergonomics,  
Psychological Aesthetics, Gestalt),  
Bamberg, Bavaria, Germany



**Philipp Kastrup**

Department of General Psychology and Methodology,  
University of Bamberg, Bamberg, Bavaria, Germany

**Isabel Friederike Abs**

Department of General Psychology and Methodology,  
University of Bamberg, Bamberg, Bavaria, Germany

**Claus-Christian Carbon**

Department of General Psychology and Methodology,  
University of Bamberg, Bamberg, Bavaria, Germany  
Forschungsgruppe EPÆG (Ergonomics,  
Psychological Aesthetics, Gestalt),  
Bamberg, Bavaria, Germany



When continuously viewing multistable displays, which are compatible with several comparably likely interpretations, perception perpetually switches between available alternatives. Prior studies typically report the lack of consistent individual switch rates across different displays. However, this comparison is based on an assumption that neural representations of physically identical displays are consistent across observers. Yet, given how different individuals are already at the level of the retina, it is likely that the difference in other relevant factors might mask the correlation. To address this issue, we compared switch rates in two kinetic-depth displays (KDE) that rotated around orthogonal axes (45° counterclockwise vs. 45° clockwise relative to the vertical). This ensured that dynamics of multistable perception was based on highly similar, but different and independent neural representations. We also included a Necker cube (NC) display as a control. We report that switch rates were correlated between two kinetic-depth effect displays, but not between either of the KDE and NC displays. This demonstrates that the usual lack of correlation may not be evidence for the lack of a shared pacesetter mechanism of multistable perception, but reflect other factors, such as differently modulated inputs to competing representations. In addition, we asked participants to speed-up or slow-down perceptual alternations and found that only the former ability was

correlated across different displays. This indicates that these two types of volitional control may differ in their use of attentional resources.

## Introduction

Certain stimuli are compatible with more than one comparably likely perceptual interpretation. For example, in a kinetic-depth effect (Wallach & O'Connell, 1953), perception of a rotating object fluctuates between two alternative directions of rotation (see Supplementary Movies S1 and S2). Similarly, the Necker cube (Necker, 1832) can be perceived in two different orientations (see Figure 1). This phenomenon, called multistable perception, exists across all sensory modalities including, for example, auditory (Denham & Winkler, 2006) and haptic (Liaci, Bach, Van Elst, Heinrich, & Kornmeier, 2016) perception. In all these cases, the perception of a continuously sensed stimulus endlessly switches between the alternative perceptual interpretations.

We still do not have the full understanding of the mechanisms behind the perceptual reversals, despite many potential explanations proposed over almost two hundred years of scientific studies of multistability

Citation: Pastukhov, A., Kastrup, P., Abs, I. F., & Carbon, C.-C. (2019). Switch rates for orthogonally oriented kinetic-depth displays are correlated across observers. *Journal of Vision*, 19(6):1, 1–13, <https://doi.org/10.1167/19.6.1>.

<https://doi.org/10.1167/19.6.1>

Received February 2, 2019; published June 3, 2019

ISSN 1534-7362 Copyright 2019 The Authors



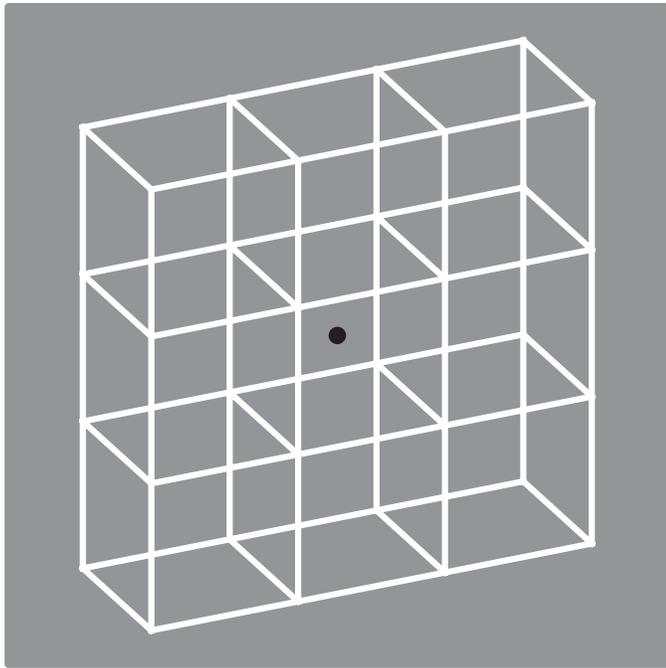


Figure 1. Necker cube lattice display used in the study.

(Wheatstone, 1838). At one extreme is an idea that perceptual ambiguity is resolved via a single central mechanism that initiates switching via top-down feedback (Leopold & Logothetis, 1999). It draws its inspiration from the strong influence of attention on multistable perception (Brascamp & Blake, 2012; Mitchell, Stoner, & Reynolds, 2004) and from observations that perceptual reversals are associated with transient activity in frontoparietal regions (Gelbard-Sagiv, Mudrik, Hill, Koch, & Fried, 2018; Weillhammer, Ludwig, Sterzer, & Hesselmann, 2014). At the other extreme is the idea that perceptual ambiguity is fully resolved via local neural circuits in the sensory regions themselves. It is supported by evidence that attention may not be necessary for multistability (Pastukhov & Braun, 2007) and questions the causal role of the frontoparietal regions in perceptual reversals (De Graaf, De Jong, Goebel, Van Ee, & Sack, 2011; de Jong et al., 2016; Knapen, Brascamp, Pearson, van Ee, & Blake, 2011). In addition, modeling shows that perceptual alternations, matching those of human observers, can be produced by a simple network with a specifically tuned level of noise, cross-inhibition, and self-adaptation (Laing & Chow, 2002).

One possible way to approach this question is by looking at the switch rate—the speed of perceptual reversals—that varies greatly both across different multistable displays (R. Cao, Pastukhov, Maurizio, & Braun, 2016; van Ee, 2005) and across observers (Medith, 1967). Specifically, switch rates for a single observer are compared across different displays and, if the switch rates are determined by a common high-level

neural mechanism, one would expect to observe the same relative switch rates across all displays. In other words, a “slow switcher” for the kinetic-depth effect should also be switching slower than most observers for the Necker cube, binocular rivalry, etc. This idea is supported by the report of an anatomical correlate of the perceptual rate that also appears to play a causal role in reversals (Kanai, Bahrami, & Rees, 2010), as well as a link to genetic factors (Shannon, Patrick, Jiang, Bernat, & He, 2011). However, most prior behavioral studies failed to find such clear relationship (Eysenck, Granger, & Brengelmann, 1957; George, 1936; Kondo et al., 2012; Medith, 1967; Pressnitzer & Hupé, 2006; Thurstone, 1944; van Loon et al., 2013; Washburn & Gillette, 1933). In the most recent work, Brascamp, Becker, and Hambrick (2018) found no correlation despite having more than 200 participants and multiple experimental conditions. Moreover, a metaanalysis performed in the same paper showed mostly the same lack of correlation. However, a recent study did report such correlation when comparing switch rates for kinetic-depth effect displays and auditory streaming (Wimmer et al., 2018). Similarly, T. Cao, Wang, Sun, Engel, and He (2018) reported that switch rates were correlated for certain subsets of stimuli, such as for various kinetic-depth effect displays. However, although these kinetic-depth effect displays differed in their shape, they rely on the same competing representation, as can be demonstrated both via an adaptation aftereffect (Nawrot & Blake, 1991; Pastukhov, Lissner, & Braun, 2014) and via a sensory memory of multistable perception (Maier, Wilke, Logothetis, & Leopold, 2003; Pastukhov, Füllekrug, & Braun, 2013).

This observed lack of consistent switch rates is mostly interpreted as evidence against the top-down or shared mechanisms of perceptual reversals. In contrast, this lack of correlation is easier to reconcile with the idea of local circuits (Laing & Chow, 2002) where the perceptual rate is determined by a complex interaction between self-adaptation (van Ee, 2009), cross-inhibition (van Loon et al., 2013), and neural noise (Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006). Here, correlated switch rates between the various displays can arise only if all three parameters are consistently similar across different parts of the sensory cortex or when different displays tap into the same key neural population.

However, this comparison between central versus local circuits is based on an implicit assumption that all other relevant factors, such as the inputs to the neural representations of the competing percepts, are identical across participants. This is a very important supposition as, for example, changes in the inputs have a dramatic effect on switch rates (Brascamp et al., 2006). Given how different individuals are already at the level

of the retina (Hofer, Carroll, Neitz, Neitz, & Williams, 2005) and the number of neural regions involved in producing that final representation, it is doubtful that the *absolute* level of the input signal will be comparable across observers. The relative effects of the strength of the inputs, such as in Brascamp et al. (2006), are, of course, robust and much easier to measure precisely because they are relative and use one of the conditions as a baseline. Thus, it is possible that the observed lack of correlation reflects merely the fact that the physically same display may produce different input strength for different observers, randomly shifting them within the group and breaking up the correlation.

Here, we sought to sidestep this issue and to replicate earlier studies by comparing two identical kinetic-depth displays that rotated around *orthogonal* axes. Specifically, the axes of rotation were 45° counterclockwise and 45° counterclockwise relative to the vertical (see Supplementary Movies S1 and S2). This arrangement means that neuronal ensembles that underpin representations of these two displays are highly similar, as they rely on similar neurons with the same selectivity and with receptive fields mapping the same spatial location within the same pathway throughout the visual cortex. Yet, due to orthogonal axes of rotation, the motion-sensitive neurons in these two ensembles belong to different independent subgroups, as the lack of response to the motion in the orthogonal direction starts already at the level of the primary visual cortex (Movshon & Newsome, 1996) and continues further into the dorsal pathway (Maunsell & Newsome, 1987), which is involved in representing the kinetic-depth effect (Brouwer & van Ee, 2007). This independence can also be demonstrated on a psychophysical level as the lack of interaction between orthogonally rotating displays, either via adaptation (Nawrot & Blake, 1993) or via sensory memory of multistable displays (Maier et al., 2003). Accordingly, our setup was designed to minimize the difference in the inputs and underlying neural representations. Thus, we were able to compare switching rates for stimuli that are represented by different, *independent* yet *highly similar* neural representations. In addition to the two kinetic-depth displays, we also included a Necker cube lattice display that served as a control condition (see Figure 1). As described above, prior work indicated that we should see little, if any, correlation between the Necker cube and kinetic-depth displays.

Finally, we also investigated whether participants' ability to control the switch rate is correlated across the three displays. Observers have a considerable degree of control over the switch rate (Brouwer & van Ee, 2006; Lack, 1969; Meredith & Meredith, 1962; Mossbridge, Ortega, Grabowecky, & Suzuki, 2013) although it differs across stimuli (Meng & Tong, 2004) and participants (Raemaekers, van der Schaaf, van Ee, &

van Wezel, 2009). This ability to control the perceptual rate relies on regions in the frontal cortex, which are not actively involved during passive viewing (De Graaf et al., 2011). Consistent with that, a recent study showed that the participants' ability for volitional control of multistable stimuli may share common mechanisms between different stimuli and modalities, even if the perceptual alternation rates during passive viewing are uncorrelated (Kondo, Pressnitzer, Shimada, Kochiyama, & Kashino, 2018). The task was to bias the perception towards one of the alternatives, and the ability to hold the dominance of a perceptual state was common across visual and auditory modality and was correlated with the ratio of inhibitory/excitatory neurotransmitters in the posterior parietal cortex. We sought to extend this work by looking at a different type of volitional control. To this end, we included two additional conditions that instructed participants to *speed-up* and *slow-down* perceptual alternations.

## Methods

### Participants

There were thirty-three participants (eight males, 25 females; ages 18–28 years), all of them undergraduates of the University of Bamberg. All procedures were in accordance with the national ethical standards on human experimentation and with the Declaration of Helsinki of 1975, as revised in 2008, and were approved by the University of Bamberg. All participants signed the informed consent prior to the experimental session. All subjects had normal or corrected-to-normal vision, normal color vision and, apart from the authors, were naïve to the purpose of the study. For their participation, observers received credits within the framework of a mandatory module of research participation in accordance with standards of University of Bamberg.

Participant BOF1999w was excluded from the analysis, as she produced no perceptual reports from the *Necker Cube* × *Slow down* condition. However, her data is included in the online dataset.

### Apparatus

Stimuli were generated using custom code and PsychoPy library (Peirce, 2008). They were displayed on an EIZO CG245W screen (<https://www.eizoglobal.com/support/db/products/model/CG245W>), with a size of the visible area being 52.0 × 32.4 cm operating at a resolution of 1920 × 1200 pixels and a refresh rate of 60 Hz. With a viewing distance of 65 cm, we obtained single pixels subtending 0.023° of visual angle. The

experimental room was dimly lit. Chin and forehead rests were used to stabilize the viewing position. Observers responded using the arrow keys of a computer keyboard.

### Kinetic-depth effect displays

The kinetic depth effect (KDE) stimulus was an orthographic projection of a sphere, which consisted of 400 white, semitransparent dots distributed randomly on the surface. The radius of the sphere was  $4^\circ$ , whereas the size of the individual dots was  $0.15^\circ$  of visual angle. The sphere rotated around the axis that was tilted  $45^\circ$  either clockwise (KDE-CW) or counterclockwise (KDE-CCW). The rotation speed was 0.25 Hz. The sphere was presented at fixation.

### Necker lattice display

The Necker cubes (NC) display consisted of a  $3 \times 3$  lattice of white wire cubes, which spanned  $1^\circ$  each (hence, total width  $3^\circ \times 3^\circ$ ). Please see Figure 1. The display was presented at fixation.

### Procedure

During each block, which lasted for one minute, participants viewed a continuously presented bistable display and reported on their perception using a keyboard. Specifically, they were instructed to press *left* and *right* arrow keys for as long as they saw the corresponding percept. Periods with no key presses indicated unclear perception.

An experimental session consisted of 45 experimental blocks. The first fifteen blocks were always *Passive viewing* condition. In two additional conditions, participants were asked to *speed-up* or *slow-down* perceptual switching as much as possible. Accordingly, the next fifteen blocks were randomly chosen to be either Speed-up or Slow-down condition. The last fifteen blocks were the complementary condition, Slow-down or Speed-up, respectively.

Every three blocks contained all three displays with shuffled presentation order. To minimize the adaptation aftereffect for KDE displays, for each subject one orientation was picked as a starting one and the two displays were always presented in that fixed order within the block triplet. For example, if the counterclockwise-tilted KDE was chosen as the first KDE display, possible triplets were KDE-CCW – KDE-CW – NC, KDE-CCW – NC – KDE-CW, and NC – KDE-CCW – KDE-CW. When the KDE display was used on two consecutive blocks, the instructions screen was

preceded by a compulsory 15-s long break in order to minimize any potential adaptation aftereffect.

### Statistical analysis

Statistical analysis was performed in R (version 3.5.1).

Before performing the main analysis, we cleaned data up by excluding blocks with no reports (1.05% of all blocks) and blocks with unclear perception taking up more than 25% of the total block duration ( $\sim 2\%$  of all blocks). In addition, we corrected potentially erroneous responses. We identified them as very brief return transitions—periods of unclear perception with the same clear perceptual state before and after it. A return transition was deemed too short if it was shorter than a median response time at the block onset for that individual observer. Once erroneous key releases were identified, they were excluded from the time-series and the preceding and succeeding clear percepts were merged into a single dominance phase.

For the correlation analysis, we opted to use geometric rather than the arithmetic mean. The former is better suited for asymmetric distributions, such as a gamma-like distribution of dominance phase durations. Our preliminary results indicated that it was also a more conservative measure, as it produced lower correlation coefficients. However, the difference was minor and quantitative rather than qualitative. The same analysis and plots but using the arithmetic mean are available in the online repository.

In addition to the *uncorrected*  $p$  and Bayes factor values, we computed distributions of Pearson correlation coefficient values using both bootstrapping (Davison & Hinkley, 1997) and Bayesian sampling (Morey & Rouder, 2015). Given that the strength of the correlation is the central question of the manuscript, we felt that this elaborative, multiviewed analysis is more justified than a simple report of individual point-estimates. Accordingly, for each comparison, we plotted the two distributions of  $\rho$  values and reported the 95% CI as well as the proportion of samples above zero.

Please note that because both methods rely on random sampling, we artificially fixed reported values by seeding the random number generator for each comparison using randomly picked numbers. The seeding did not make a qualitative difference to the results and can be commented out in the analysis code in the online repository. For bootstrapping, we used 2,000 repetitions; for Bayesian analysis 10,000 iterations.

Because the Bayesian analysis samples the posterior distribution, it can lead to  $\rho$  values above one. This is a purely mathematical side effect of stochastic sampling via the Markov chain Monte Carlo method used by the

R package *BayesFactor* that may overestimate the slope.

When comparing distributions of dominance times we used either a two-sample Bayesian  $t$  test or a two-sample  $t$  test via Monte Carlo permutation (Fellows, 2012). In both cases, we used default parameters suggested by the functions' authors. For the permutation ANOVA, we used R package *lmPerm* (version 2.1.0).

## Data availability

All data files and the analysis code are available under Creative Commons Attribution 4.0 International Public License at [osf.io/t2p5q](https://osf.io/t2p5q) (DOI: 10.17605/OSF.IO/T2P5Q) or at [github.com/alexander-pastukhov/correlated-switching-rate-in-kdes](https://github.com/alexander-pastukhov/correlated-switching-rate-in-kdes).

## Results

### Passive viewing condition

The purpose of the passive viewing condition was to compare our results with prior work. Specifically, after computing the geometric mean of dominance phase durations for every display and observer, we computed a Pearson correlation coefficient for every pair of displays. Figure 2 contains the results with the left column showing individual mean durations and the right column depicting distributions generated via either bootstrapping or Bayesian analysis plus the summary statistics.

We found that the mean dominance phase durations were strongly correlated for the two kinetic-depth effect (KDE) displays ( $\rho \approx 0.6$ ), but not between either of KDE displays and the Necker cube (NC). In the latter cases, Bayes factors below one gave additional support to the null-hypothesis (independence). In the former case, the means were not only correlated, but also statistically similar for most observers. Only four out of 32 distributions were different based on corrected  $p$  values computed via a two-sample permutation  $t$  test (marked by circles with a black outline in Figure 2A) and seven out of 32 based on Bayes Factor computed via a Bayesian two-sample  $t$  test (please refer to the online repository for details). In contrast, the mean dominance times were statistically different for most participants when comparing either KDE display with NC (20 out of 32 distributions were different based on corrected  $p$  values and 24 out of 32 based on Bayes Factor).

We also compared the coefficient of variation across the displays (results not shown but available in the online repository). As for the geometric mean, we

observed a strong correlation between two KDE displays ( $\rho \approx 0.53$ ) but not between either of KDE displays and NC.

To summarize, for the passive viewing condition we observed highly correlated and mostly similar dominance phase times between the two KDE displays, but consistent with prior work, we were not able to detect correlations between one of the two KDEs and NC displays.

### Volitional control conditions

Following the passive view condition, participants were instructed to either Speed-Up or Slow-Down perceptual switching for the following fifteen blocks (blocks 15 to 30) and to attempt to do the opposite for the last fifteen blocks (blocks 31 to 45). The participants were moderately successful in their volitional control (see Figure 3), and their ability to change the rate of perceptual alternations depended only on the instructions but neither on the display nor the condition order (see Table 1). However, they were significantly better in speeding perceptual switches up than in slowing them down:  $t(190) = -2.6$ ,  $p = 0.012$ ,  $BF = 3.3$ ,  $R_m^2 = 0.032$ ,  $R_c^2 = 0.032$ .

For the Speed-Up condition (Figure 4), we observed a strong correlation for changes in average dominance phase durations relative to the passive viewing condition not only for the two KDE displays but also between the two KDE and the NC displays. The latter correlation is weaker ( $\rho \approx 0.3$ – $0.4$ ) and is only marginally statistically significant. However, both bootstrap and Bayesian analysis indicate that this moderate correlation is likely to be genuine.

In contrast, for the Slow-Down condition relative changes in dominance phase durations were correlated only for the KDE displays (see Figure 5). In addition, the bootstrapped distribution was multimodal, likely reflecting the influence of two outliers.

## Discussion

The main goal of our study was to compare perceptual switch rates across participants for two kinetic-depth effect (KDE) displays and a Necker cube (NC) display. We found that consistent with prior work, switch rates were uncorrelated between KDE and NC displays during the passive viewing. However, we observed a very strong correlation between the switch rates between the two KDE displays. Moreover, these switch rates were not just correlated but

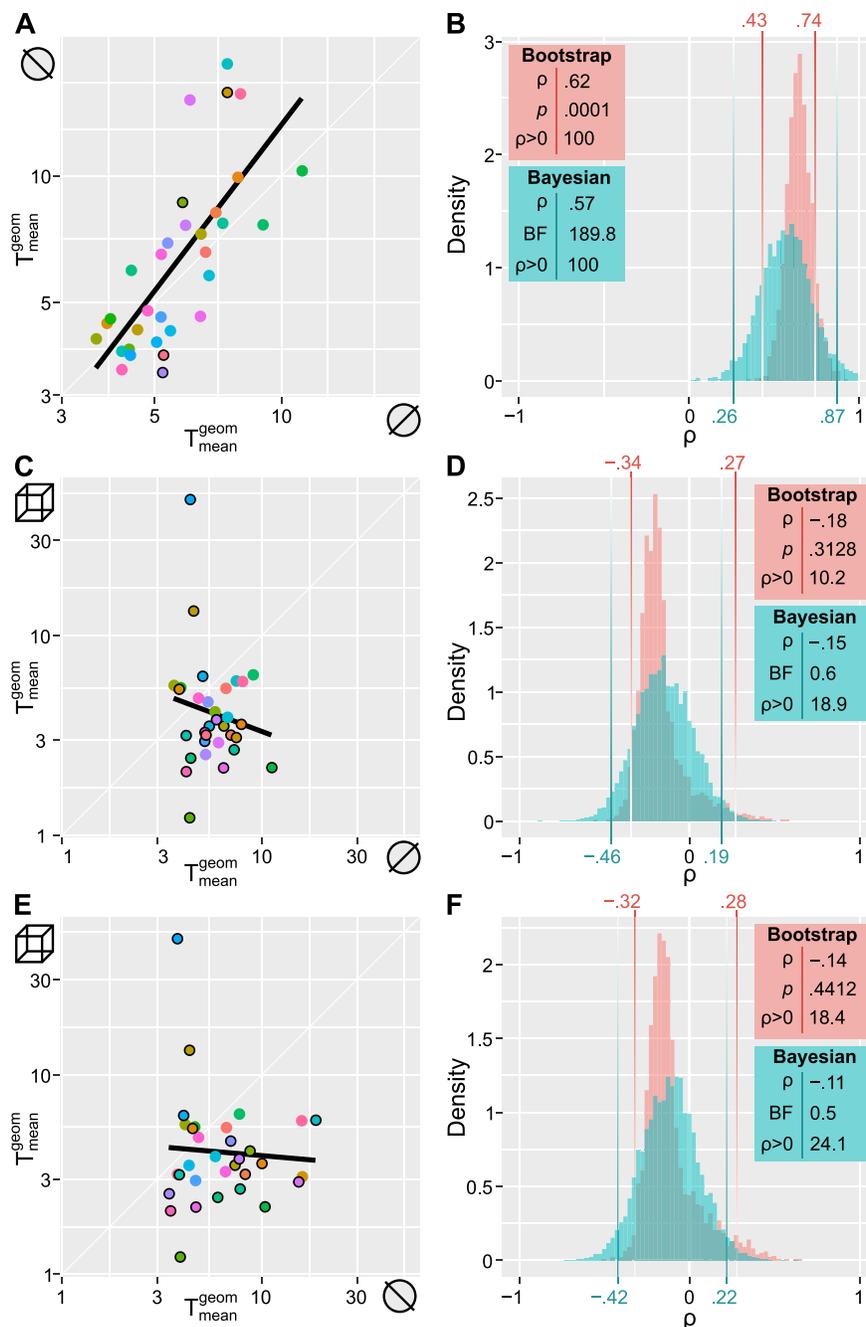


Figure 2. Pairwise correlations between kinetic-depth effect displays whose axis of rotation was 45° counterclockwise ⊙ and 45° clockwise ⊙, and Necker cube lattice ⊠. Left column, (A), (C), and (E): geometric means for dominance phase durations. Colors denote individual observers. The straight black line shows the linear regression fit. Circles with a black outline mark observers for whom mean dominance durations for the two displays were significantly different. Right column (B), (D), and (F): Distributions of correlation coefficients computed using bootstrap and Bayesian methods. Numbers above and below indicate limits for the 95% CI for, respectively, bootstrap and Bayesian methods. Insets show the correlation coefficient ( $\rho$ ), statistical significance (either uncorrected  $p$  value or Bayes Factor), and percentage of samples above the zero mark ( $\rho > 0$ ).

statistically very similar for most participants. In addition, we compared the participants' ability to speed-up and slow-down perceptual alternations. We report that only the former was correlated not only for KDE displays but also between KDE and NC displays.

### Correlations in perceptual switch rates

Our results for the passive view condition showed that it is possible to observe correlated switch rates across multistable displays, even if they rely on

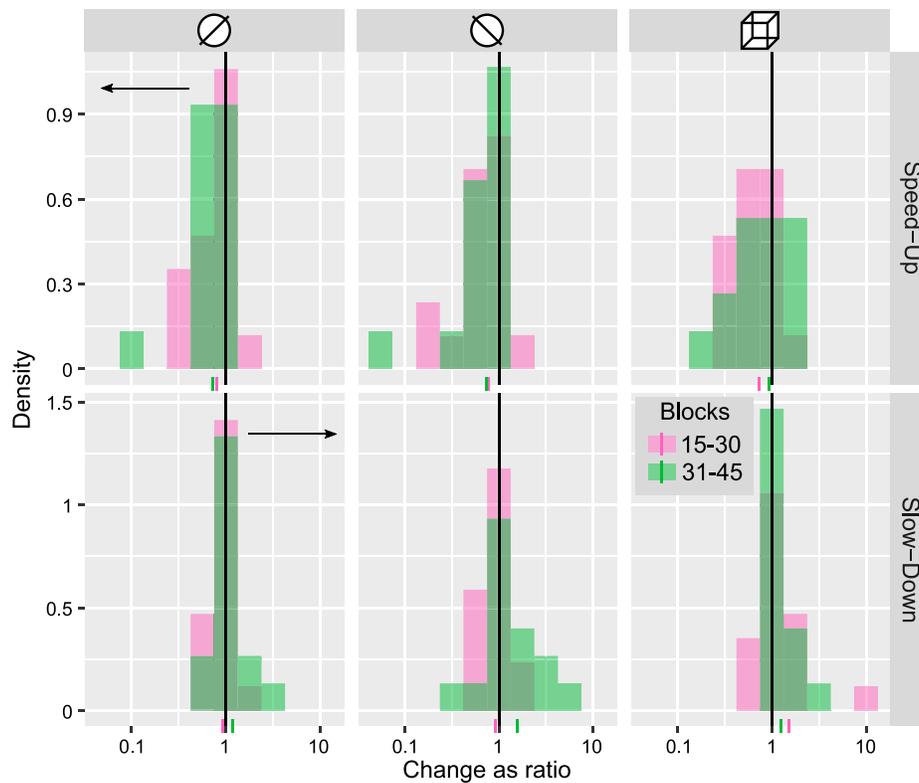


Figure 3. Volitional control measured as a ratio of geometric means of the selected condition and the passive view condition. Columns show different displays, whereas rows correspond to the two volitional control conditions. Colored ticks on the x axis depict distributions’ means for the *early* (15–30) and *late* (31–45) blocks. Arrows show the expected shift relative to a 1:1 ratio.

different neural representations. Specifically, we used two KDE displays that rotated around the orthogonally oriented axes, which ensured their independence (Maier et al., 2003; Nawrot & Blake, 1993). At the same time and consistent with the prior work (Brascamp et al., 2018), we observed no correlation between KDE and NC displays.

Although our results demonstrate that such correlation is, in principle, observable, it does not allow us to pinpoint the location of the pacesetter circuit for multistable perception. It merely shows that the lack of correlation cannot be used as evidence against the central mechanisms, as it may reflect differences in other relevant factors, such as the inputs to the neural

representations of the competing percepts or heterogeneous properties of neurons in these populations in different cortical areas. However, our results are equally compatible with the idea of local circuits. The two KDE displays we used were identical in every respect but for their axis of rotation. Even though their neural representations, as well as most of the upstream input neurons, are different, they map the same type of perception at the same spatial location. It is very likely that their output and internal architecture are highly similar to ensure that their output is easily comparable and that it can be treated as orientation-independent. The latter idea is supported by the fact that mean dominance phase durations were statistically similar for

Model	df	AIC	BIC	logLik	Deviance	$\chi^2$	df( $\chi^2$ )	$p(\chi^2)$	$p(\text{perm})$	$R_m^2$	$R_c^2$	BF
Random effects	3	443	453	-219	437							
+ condition	4	427	441	-210	419	17.7	1	<0.0001	0.0001	0.08	0.15	985
+ display	6	429	448	-208	417	2.8	2	0.2457	0.2562	0.10	0.17	0.2
+ part	7	430	452	-208	416	1.0	1	0.3159	0.3641	0.10	0.17	0.3

Table 1. Permutation ANOVA with Display, Experimental part (blocks 15–30 or 31–45), and Condition (Speed-Up or Slow-Down) as factors. Statistical analysis of relative change in dominance phase durations for volitional control conditions. Hierarchical linear-mixed models that were compared pairwise via  $\chi^2$  test, permutation test, and Bayes Factor. Fixed factors were volitional control condition (Speed-Up or Slow-Down), display, and experimental part (blocks 15–30 or 31–45, with participants’ identity as a random factor. *df* = degrees of freedom; AIC = Akaike information criterion; BIC = Bayesian information criterion; logLik = log likelihood;  $R_m^2$  and  $R_c^2$  are, correspondingly, marginal and conditional coefficients of determination; BF = Bayes Factor.

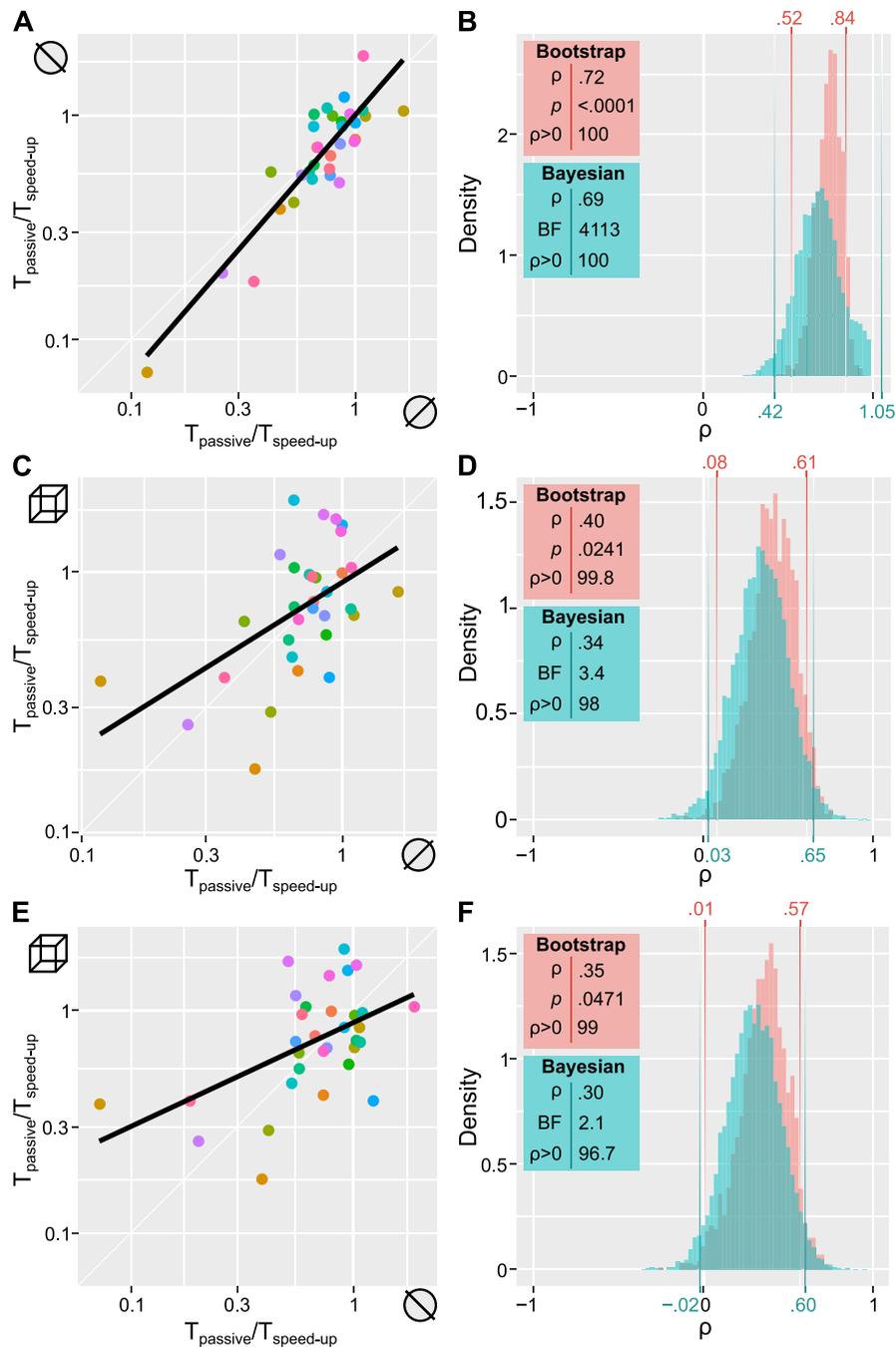


Figure 4. Change of dominance phase duration in the speed-up versus passive viewing condition. For labeling and details, see Figure 2.

the absolute majority of observers for the two KDE displays (see Figure 2A). Thus, both extreme models, as well as any intermediate ones, could explain this correlation.

Our results also extend prior work that showed correlated switching rates for KDE displays that differed in shape but all rotated around the same vertical axis of rotation (T. Cao et al., 2018). In this case, the observed correlation is expected as all these

rotating displays share the same competing neural subpopulations that code two directions of rotation (Maier et al., 2003; Nawrot & Blake, 1991; Pastukhov et al., 2013, 2014). However, the same study also showed weaker correlations between a subset of displays that included binocular rivalry, vase-face illusion, moving plaid, Necker cube, and rolling wheel. These subset specific correlations, coupled with our results, paint a more complex story than a mere lack of

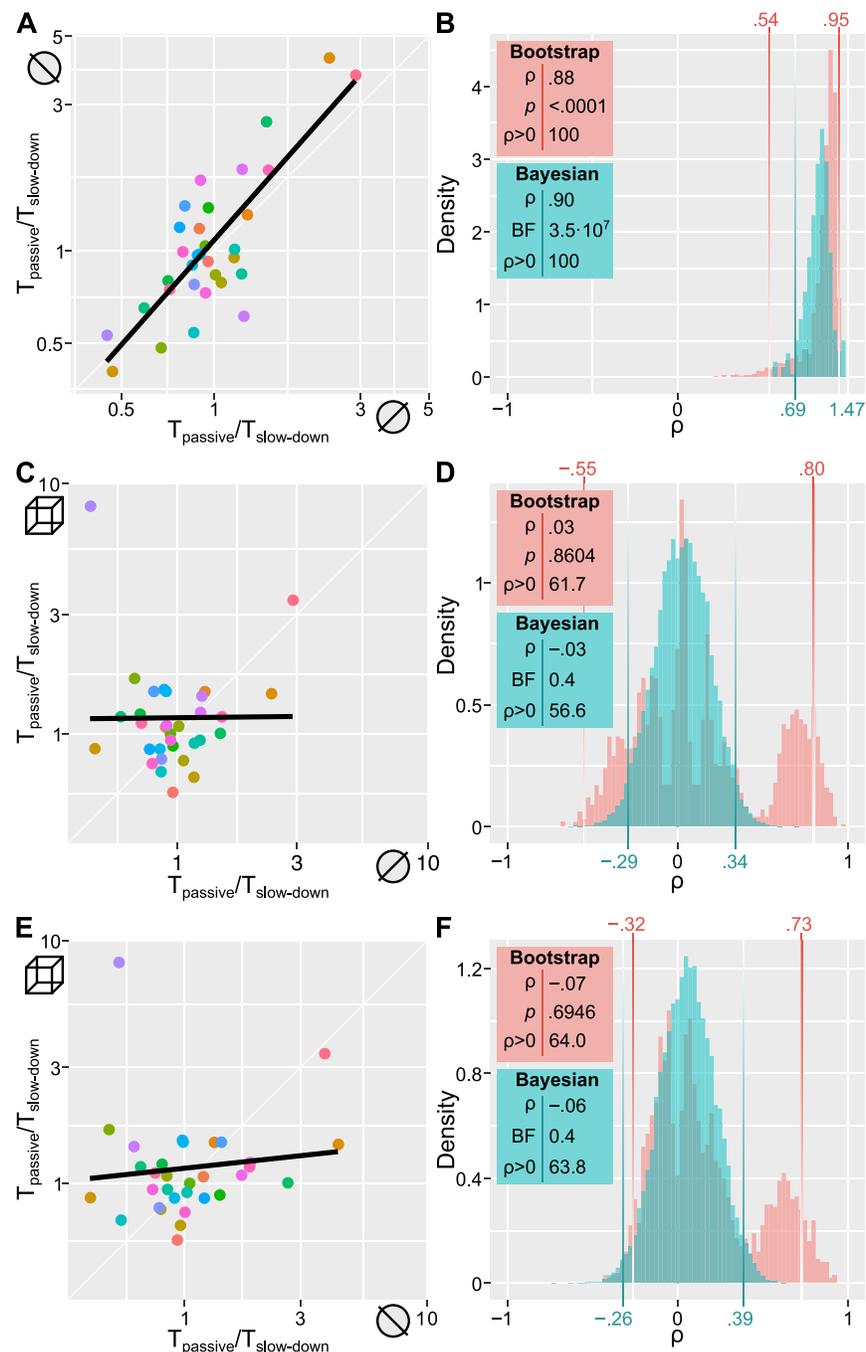


Figure 5. Change of dominance phase duration in the *slow-down* versus *passive viewing* condition. For labeling and details, see Figure 2.

relationship for switch rates across different multistable displays. Rather, they indicate that correlation strength depends on the similarity or overlap between underlying representations (but see also Wimmer et al., 2018). As noted above, that may not allow us to pinpoint the neural locus that determines the switch rate, but stimuli similarity measures coupled with potentially better characterization of dominance phase distribution (van Ee, Noest, Brascamp, & van den Berg, 2006) could be a useful guide for future studies.

### Volitional control

Our study also included two volitional control conditions where participants were asked to either speed-up or slow-down perceptual correlations. We found that there was a moderate correlation for changes in perceptual switching rates during the speed-up condition. However, there was no correlation for the participants' ability to slow-down alternation rates when we compared KDE and NC displays. Our results

complement a recent report that the ability to hold the dominance of a perceptual state was common across visual and auditory modality (Kondo et al., 2018).

It is possible that the dissociation between speed-up and slow-down conditions was a mere artifact of perceptual learning. The latter leads to an increased switch rate throughout and across experimental sessions (Pastukhov & Braun, 2013; Suzuki & Grabowecy, 2007). Accordingly, this increase would work against participants' attempts to slow-down perceptual switches and the lack of correlation may reflect an interaction between different individual learning rates and volitional control strength.

However, it is also possible that this dissociation reflects differences in the two modes of volitional control. For example, EEG evidence indicates that two conditions differ with slow-down condition eliciting a stronger response in delta- and gamma-bands (Mathes, Strüber, Stadler, & Basar-Eroglu, 2006). This was interpreted as the greater need for attentional resources when preventing perceptual reversals, as compared to initiating one. The two abilities are also dissociated by age or disease. For example, Parkinson's disease patients, as compared to healthy controls, are less capable of speeding switch rate up but show no difference for slow-down condition (Díaz-Santos et al., 2015). A further dissociation was observed when comparing young and older adults (Díaz-Santos et al., 2017). In this case, both groups were equally adept in speeding alternations up, and the older participants had difficulty in the slow-down condition. In our case, participants were significantly better in speeding perceptual switches up than slowing them down.

To conclude, our results, in combination with prior work, indicate that different modes of the volitional control may reflect different attentional requirements or how attentional mechanisms interaction with the lower level sensory representations. They also complement prior work that shows that some multistable displays are easier to control than others (see Paffen & Alais, 2011 for a review). In particular, the binocular rivalry is thought to be much harder to control volitionally than other stimuli (Meng & Tong, 2004). With respect to our findings, it would be interesting to investigate whether similar dissociation would be observed for other pairs of multistable stimuli and whether this dissociation is modulated by a similarity in a degree of volitional control and by similar in stimuli representations.

## Conclusions

We report a strong correlation for switch rates between two kinetic-depth effect displays that were

identical but for the orthogonal axes of rotation and, therefore, relied on independent neural representations. Yet, consistent with prior work, we found no correlation between switch rates between kinetic-depth effect and Necker cube displays. This discrepancy may reflect not the absence of shared neural mechanisms for the initiation of perceptual switches but other differences between neural representations of the displays, such as differently modulated inputs.

In addition, we found that although the participants' ability to speed-up perceptual switching was similar across the kinetic-depth effect and Necker cube displays, their ability to slow them down was not. This indicates that these two types of volitional control differ in the way attention interacts with underlying sensory representations.

*Keywords:* multistable perception, interobserver variability, interstimulus variability, kinetic-depth effect, structure-from-motion, Necker cube, perceptual ambiguity, bistable perception, volitional control, visual illusion

## Acknowledgments

Commercial relationships: none.

Corresponding author: Alexander Pastukhov.

Email: Pastukhov.Alexander@gmail.com.

Address: Department of General Psychology and Methodology, University of Bamberg, Bamberg, Germany.

## References

- Brascamp, J. W., Becker, M. W., & Hambrick, D. Z. (2018). Revisiting individual differences in the time course of binocular rivalry. *Journal of Vision*, 18(7): 3, 1–20, <https://doi.org/10.1167/18.7.3>. [PubMed] [Article]
- Brascamp, J. W., & Blake, R. (2012). Inattention abolishes binocular rivalry: Perceptual evidence. *Psychological Science*, 23(10), 1159–1167, <https://doi.org/10.1177/0956797612440100>.
- Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R. H. A. H., & van den Berg, A. V. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *Journal of Vision*, 6(11):8, 1244–1256, <https://doi.org/10.1167/6.11.8>. [PubMed] [Article]
- Brouwer, G. J., & van Ee, R. (2006). Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Vision Re-*

- search, 46(20), 3393–3402, <https://doi.org/10.1016/j.visres.2006.03.016>.
- Brouwer, G. J., & van Ee, R. (2007). Visual cortex allows prediction of perceptual states during ambiguous structure-from-motion. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(5), 1015–1023, <https://doi.org/10.1523/JNEUROSCI.4593-06.2007>.
- Cao, R., Pastukhov, A., Maurizio, M., & Braun, J. (2016). Collective activity of many bistable assemblies reproduces characteristic dynamics of multistable perception. *Journal of Neuroscience*, 36(26), 6957–6972, <https://doi.org/10.1523/JNEUROSCI.4626-15.2016>.
- Cao, T., Wang, L., Sun, Z., Engel, S. A., & He, S. (2018). The independent and shared mechanisms of intrinsic brain dynamics: Insights from bistable perception. *Frontiers in Psychology*, 9(April), 589, <https://doi.org/10.3389/FPSYG.2018.00589>.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press, <https://doi.org/10.2307/1271471>.
- De Graaf, T. A., De Jong, M. C., Goebel, R., Van Ee, R., & Sack, A. T. (2011). On the functional relevance of frontal cortex for passive and voluntarily controlled bistable vision. *Cerebral Cortex*, 21(10), 2322–2331, <https://doi.org/10.1093/cercor/bhr015>.
- de Jong, M. C., Hendriks, R. J. M., Vansteensel, M. J., Raemaekers, M., Verstraten, F. A. J., Ramsey, N. F., . . . van Ee, R. (2016). Intracranial recordings of occipital cortex responses to illusory visual events. *The Journal of Neuroscience*, 36(23), 6297–6311, <https://doi.org/10.1523/JNEUROSCI.0242-15.2016>.
- Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology Paris*, 100(1–3), 154–170, <https://doi.org/10.1016/j.jphysparis.2006.09.012>.
- Díaz-Santos, M., Cao, B., Yazdanbakhsh, A., Norton, D. J., Nearing, S., & Cronin-Golomb, A. (2015). Perceptual, cognitive, and personality rigidity in Parkinson's disease. *Neuropsychologia*, 69, 183–193, <https://doi.org/10.1016/j.neuropsychologia.2015.01.044>.
- Díaz-Santos, M., Mauro, S., Cao, B., Yazdanbakhsh, A., Nearing, S., & Cronin-Golomb, A. (2017). Bistable perception in normal aging: Perceptual reversibility and its relation to cognition. *Aging, Neuropsychology, and Cognition*, 24(2), 115–134, <https://doi.org/10.1080/13825585.2016.1173646>.
- Eysenck, H. J., Granger, G. W., & Brengelmann, J. C. (1957). *Perceptual processes and mental illness*. London, UK: Oxford University Press.
- Fellows, I. (2012). Deducer: A data analysis GUI for R. *Journal of Statistical Software*, 49(8), 1–15, <https://doi.org/10.18637/jss.v049.i08>.
- Gelbard-Sagiv, H., Mudrik, L., Hill, M. R., Koch, C., & Fried, I. (2018). Human single neuron activity precedes emergence of conscious perception. *Nature Communications*, 9(1):2057, <https://doi.org/10.1038/s41467-018-03749-0>.
- George, R. W. (1936). The significance of the fluctuations experienced in observing ambiguous figures and in binocular rivalry. *The Journal of General Psychology*, 15(1), 39–61, <https://doi.org/10.1080/00221309.1936.9917904>.
- Hofer, H., Carroll, J., Neitz, J., Neitz, M., & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(42), 9669–7969, <https://doi.org/10.1523/JNEUROSCI.2414-05.2005>.
- Kanai, R., Bahrami, B., & Rees, G. (2010). Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current Biology: CB*, 20(18), 1626–1630, <https://doi.org/10.1016/j.cub.2010.07.027>.
- Knapen, T., Brascamp, J., Pearson, J., van Ee, R., & Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *Journal of Neuroscience*, 31(28), 10293–10301, <https://doi.org/10.1523/JNEUROSCI.1727-11.2011>.
- Kondo, H. M., Kitagawa, N., Kitamura, M. S., Koizumi, A., Nomura, M., & Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cerebral Cortex*, 22(8), 1915–1922, <https://doi.org/10.1093/cercor/bhr266>.
- Kondo, H. M., Pressnitzer, D., Shimada, Y., Kochiyama, T., & Kashino, M. (2018). Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. *Scientific Reports*, 8(1):14548, <https://doi.org/10.1038/s41598-018-32892-3>.
- Lack, L. C. (1969). The effect of practice on binocular rivalry control. *Perception & Psychophysics*, 6(6), 397–400, <https://doi.org/10.3758/BF03212798>.
- Laing, C. R., & Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *Journal of Computational Neuroscience*, 12(1), 39–53, <https://doi.org/10.1023/A:1014942129705>.
- Leopold, D. A., & Logothetis, N. K. N. (1999). Multistable phenomena: Changing views in per-

- ception. *Trends in Cognitive Sciences*, 3(7), 254–264, [https://doi.org/10.1016/S1364-6613\(99\)01332-7](https://doi.org/10.1016/S1364-6613(99)01332-7).
- Liaci, E., Bach, M., Van Elst, L. T., Heinrich, S. P., & Kornmeier, J. (2016). Ambiguity in tactile apparent motion perception. *PLoS One*, 11(5), 1–22, <https://doi.org/10.1371/journal.pone.0152736>.
- Maier, A., Wilke, M., Logothetis, N. K., & Leopold, D. A. (2003). Perception of temporally interleaved ambiguous patterns. *Current Biology*, 13(13), 1076–1085, [https://doi.org/10.1016/S0960-9822\(03\)00414-7](https://doi.org/10.1016/S0960-9822(03)00414-7).
- Mathes, B., Strüber, D., Stadler, M. A., & Basar-Eroglu, C. (2006). Voluntary control of Necker cube reversals modulates the EEG delta- and gamma-band response. *Neuroscience Letters*, 402(1–2), 145–149, <https://doi.org/10.1016/j.neulet.2006.03.063>.
- Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10(1), 363–401, <https://doi.org/10.1146/annurev.ne.10.030187.002051>.
- Medith, G. M. (1967). Some attributive dimensions of reversibility phenomena and their relationship to rigidity and anxiety. *Perceptual and Motor Skills*, 24(3), 843–849, <https://doi.org/10.2466/pms.1967.24.3.843>.
- Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision*, 4(7):2, 539–551, <https://doi.org/10.1167/4.7.2>. [PubMed] [Article]
- Meredith, G. M., & Meredith, C. G. W. (1962). Effect of instructional conditions on rate of binocular rivalry. *Perceptual and Motor Skills*, 15(3), 655–664, <https://doi.org/10.2466/pms.1962.15.3.655>.
- Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004, May 27). Object-based attention determines dominance in binocular rivalry. *Nature*, 429(6990), 410–413, <https://doi.org/10.1038/nature02584>. R package version 0.9.12-4.2. <https://CRAN.R-project.org/package=BayesFactor>
- Morey, R. D., & Rouder, J. N. (2015). Bayesfactor: Computation of Bayes factors for common designs. R package version 0.9.12-2. *BayesFactor: Computation of Bayes factors for common designs*, <https://doi.org/10.1007/978-3-642-21037-2>.
- Mossbridge, J. A., Ortega, L., Grabowecky, M., & Suzuki, S. (2013). Rapid volitional control of apparent motion during percept generation. *Attention, Perception, & Psychophysics*, 75(7), 1486–1495, <https://doi.org/10.3758/s13414-013-0504-3>.
- Movshon, J. A., & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *The Journal of Neuroscience*, 16(23), 7733–7741, <https://doi.org/10.1523/JNEUROSCI.16-23-07733.1996>.
- Nawrot, M., & Blake, R. (1991). The interplay between stereopsis and structure from motion. *Perception & Psychophysics*, 49(3), 230–244, <https://doi.org/10.3758/BF03214308>.
- Nawrot, M., & Blake, R. (1993). On the perceptual identity of dynamic stereopsis and kinetic depth. *Vision Research*, 33(11), 1561–1571, [https://doi.org/10.1016/0042-6989\(93\)90149-Q](https://doi.org/10.1016/0042-6989(93)90149-Q).
- Necker, L. A. (1832). LXI. Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(5), 329–337, <https://doi.org/10.1080/14786443208647909>.
- Paffen, C. L. E., & Alais, D. (2011). Attentional modulation of binocular rivalry. *Frontiers in Human Neuroscience*, 5(September):105, <https://doi.org/10.3389/fnhum.2011.00105>.
- Pastukhov, A., & Braun, J. (2007). Perceptual reversals need no prompting by attention. *Journal of Vision*, 7(10):5, 1–17, <https://doi.org/10.1167/7.10.5>. [PubMed] [Article]
- Pastukhov, A., & Braun, J. (2013). Disparate time-courses of adaptation and facilitation in multistable perception. *Learning & Perception*, 5(Suppl., 2), 101–118, <https://doi.org/10.1556/LP.5.2013.Suppl2.7>.
- Pastukhov, A., Füllekrug, J., & Braun, J. (2013). Sensory memory of structure-from-motion is shape-specific. *Attention, Perception, & Psychophysics*, 75(6), 1215–1229, <https://doi.org/10.3758/s13414-013-0471-8>.
- Pastukhov, A., Lissner, A., & Braun, J. (2014). Perceptual adaptation to structure-from-motion depends on the size of adaptor and probe objects, but not on the similarity of their shapes. *Attention, Perception, & Psychophysics*, 76(2), 473–488, <https://doi.org/10.3758/s13414-013-0567-1>.
- Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2(January), 1–8, <https://doi.org/10.3389/neuro.11.010.2008>.
- Pressnitzer, D., & Hupé, J.-M. (2006). Temporal dynamics of auditory and bisual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13), 1351–1357, <https://doi.org/10.1016/j.cub.2006.05.054>.
- Raemaekers, M., van der Schaaf, M. E., van Ee, R., & van Wezel, R. J. A. (2009). Widespread fMRI

- activity differences between perceptual states in visual rivalry are correlated with differences in observer biases. *Brain Research*, 1252, 161–171, <https://doi.org/10.1016/j.brainres.2008.11.078>.
- Shannon, R. W., Patrick, C. J., Jiang, Y., Bernat, E., & He, S. (2011). Genes contribute to the switching dynamics of bistable perception. *Journal of Vision*, 11(3):8, 1–7, <https://doi.org/10.1167/11.3.8>. [PubMed] [Article]
- Suzuki, S., & Grabowecy, M. (2007). Long-term speeding in perceptual switches mediated by attention-dependent plasticity in cortical visual processing. *Neuron*, 56(4), 741–753, <https://doi.org/10.1016/j.neuron.2007.09.028>.
- Thurstone, L. L. (1944). *A factorial study of perception*. Chicago, IL: University of Chicago Press.
- van Ee, R. (2005). Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Research*, 45(1), 29–40, <https://doi.org/10.1016/j.visres.2004.07.039>.
- van Ee, R. (2009). Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: Evidence from serial correlations in perceptual bistability. *Journal of the Optical Society of America A*, 26(12), 2612–2622, <https://doi.org/10.1364/JOSAA.26.002612>.
- van Ee, R., Noest, A. J., Brascamp, J. W., & van den Berg, A. V. (2006). Attentional control over either of the two competing percepts of ambiguous stimuli revealed by a two-parameter analysis: Means do not make the difference. *Vision Research*, 46(19), 3129–3141, <https://doi.org/10.1016/j.visres.2006.03.017>.
- van Loon, A. M., Knapen, T. H. J., Scholte, H. S., St John-Saaltink, E., Donner, T. H., & Lamme, V. A. F. (2013). GABA shapes the dynamics of bistable perception. *Current Biology: CB*, 23(9), 823–827, <https://doi.org/10.1016/j.cub.2013.03.067>.
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45(4), 205–217, <https://doi.org/10.1037/h0056880>.
- Washburn, M. F., & Gillette, A. (1933). Motor factors in voluntary control of cube perspective fluctuations and retinal rivalry fluctuations. *The American Journal of Psychology*, 45(2), 315–319, <https://doi.org/10.2307/1414285>.
- Weilnhammer, V. A., Ludwig, K., Sterzer, P., & Hesselmann, G. (2014). Revisiting the Lissajous figure as a tool to study bistable perception. *Vision Research*, 98, 107–112, <https://doi.org/10.1016/j.visres.2014.03.013>.
- Wheatstone, C. (1838). XVIII. Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128, 371–394, <https://doi.org/10.1098/rstl.1838.0019>.
- Wimmer, M., Carmel, D., Kocsis, Z., Taranu, M., Farkas, D., van Ee, R., ... Denham, S. L. (2018). Similar but separate systems underlie perceptual bistability in vision and audition. *Scientific Reports*, 8(1), 1–10, <https://doi.org/10.1038/s41598-018-25587-2>.

## Supplementary materials

**Supplementary Movie S1.** Kinetic-depth effect display; the axis of rotation is 45° counterclockwise relative to the vertical.

**Supplementary Movie S2.** Kinetic-depth effect display; the axis of rotation is 45° clockwise relative to the vertical.