

Perceptual coupling induces co-rotation and speeds up alternations in adjacent bi-stable structure-from-motion objects

Alexander Pastukhov

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



Christina Rita Zaus

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

Stepan Aleshin

Institute of Biology, Otto-von-Guericke University,
Magdeburg, Sachsen-Anhalt, Germany
Center for Behavioral Brain Sciences, Magdeburg,
Sachsen-Anhalt, Germany

Jochen Braun

Institute of Biology, Otto-von-Guericke University,
Magdeburg, Sachsen-Anhalt, Germany
Center for Behavioral Brain Sciences, Magdeburg,
Sachsen-Anhalt, Germany

Claus-Christian Carbon

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

When two bi-stable structure-from-motion (SFM) spheres are presented simultaneously, they tend to rotate in the same direction. This effect reflects a common state bias that is present for various multistable displays. However, it was also reported that when two spheres are positioned so that they touch each other, they tend to counterrotate instead. The latter effect is interpreted as a frictional interaction, indicating the influence of the embedded physics on our visual perception. Here, we examined the interplay between these two biases in two experiments using a wide range of conditions. Those included two SFM shapes, two types of disambiguation cues, the presence or absence of the disambiguation cues, different layout options, and two samples of observers from two different universities (in sum 26 participants). Contrary to the prior report, we observed a robust common state bias for all conditions, including those that were optimized for frictional and “gear meshing” interactions. We found that stronger

coupling of perceptual states is accompanied by more frequent synchronous perceptual reversals of the two objects. However, we found that the simultaneity of the individual switches does not predict the duration of the following dominance phase. Finally, we report that stronger perceptual coupling speeds up perceptual alternations.

Introduction

When our visual system is confronted with a multistable display—a visual display that is compatible with several comparably likely interpretations—our perception becomes unstable, oscillating between the alternatives. When several copies of such displays are viewed simultaneously, their perceptual states tend to couple, so that a single percept dominates all displays

Pastukhov, A., Zaus, C. R., Aleshin, S., Braun, J., & Carbon, C.-C. (2018). Perceptual coupling induces co-rotation and speeds up alternations in adjacent bi-stable structure-from-motion objects. *Journal of Vision*, 18(4):21, 1–14, <https://doi.org/10.1167/18.4.21>.



and they tend to switch in accord. This common state bias appears to reflect the influence of a general mechanism, because it was observed for very different multistable displays such as Necker cubes (Adams & Haire, 1958), Attneave's triangles (Attneave, 1968), motion quartets (Ramachandran & Anstis, 1983; Ramachandran & Anstis, 1985), or ambiguously rotating structure-from-motion (SFM) objects (Eby, Loomis, & Solomon, 1989; Gillam, 1972; Gillam, 1976; Grossmann & Dobbins, 2003; Landy, 1987).

Gilroy and Blake (2004) reported a curious exception to this rule. They observed that *counterrotation* dominated the perception when two SFM spheres were presented side by side so that they were touching each other. This perceptual coupling, which forced *opposite* dominance states onto the two spheres, was specific to the “touching” layout, as an introduction of a small gap eliminated this effect. Based on the nature and the specificity of the bias, authors of the original study interpreted it as an influence of physics of a frictional interaction embedded in visual perception. A similar interaction was also reported for a bi-stable point-light walker positioned on top of an unambiguously rotating sphere (Jackson & Blake, 2010).

The embedded physics interpretation bodes well with other known physics-based or, to put it more accurately, statistically based ecologically valid perceptual biases (Kersten & Yuille, 2003). For example, a hollow mask tends to be perceived as convex, as concave faces are highly unlikely (Gregory, 1997). For the shape from shading, the default assumed position of the light source is “from above,” presumably matching that of the sun (Gerardin, Kourtzi, & Mamassian, 2010). For SFM displays, their perceptual stability is modulated by their shape and their orientation, reflecting an expected stability of their bi-stable depth and motion (Pastukhov, Vonau, & Braun, 2012).

The frictional interaction is unique in that it affects two multistable objects in an *opposing* manner while working against a very potent common state bias described above (Attneave, 1968; Eby et al., 1989; Gillam, 1972; Gillam, 1976; Grossmann & Dobbins, 2003; Landy, 1987; Ramachandran & Anstis, 1983; Ramachandran & Anstis, 1985). Here, we sought to examine the interplay between these two perceptual biases in a systematic way. In addition to the frictional interaction, we introduced a display configuration that could be interpreted as meshing gears. The latter interaction has two potential advantages over the frictional one with two spheres. First, it is mechanically stronger and, therefore, has a higher potential for perceptual coupling via embedded physics. Second, meshing gears may require a less precise alignment of the objects, making “meshing” perception potentially more robust. Below, we present experimental results for

two samples of participants from two different universities.

Methods

Participants

All procedures were in accordance with the national ethical standards for human experimentation and with the Declaration of Helsinki of 1975, as revised in 2008, and were approved by the University of Bamberg and by the medical ethics board of the University of Magdeburg “Ethik-Kommission der Otto-von-Guericke-Universität an der Medizinischen Fakultät.”

Experiment 1 was performed at the University of Bamberg. Informed consent was obtained from all participants prior to the experimental session. All observers had a normal or corrected-to-normal vision and were naive as to the purpose of the experiments, apart from the second author (observer AZM90F). Seventeen observers, 12 females, aged 19–36 years, including the second author, and five males, aged 21–29 years, participated in Experiment 1. A single experimental session for the observer SJW1988M was excluded from the analysis because due to a programming error it lacked the *overlap* condition. To ensure full transparency, this dataset is also included in the online repository in a “missing overlap condition” subfolder.

Experiment 2 was carried out at the University of Magdeburg as part of the practicum for master's students of the Integrative Neuroscience program. Nine observers, three females, aged 22–27 years, had normal or corrected-to-normal visual acuity. Because the decision-making about the experimental design was part of the practicum, all participants were aware of the purpose of the study and of the experimental hypotheses (e.g., that a touching layout should induce the perception of the counterrotation, but a gap in between two objects should eliminate this effect).

Apparatus

In Experiment 1, displays were presented on a Samsung SyncMaster 2233RZ monitor (Samsung Electronics GmbH, Schwalbach am Taunus, Germany). The size of the visible area was 47.5×29.5 cm with resolution of $1,680 \times 1,050$ pixels and a refresh rate of 120 Hz. A single pixel subtended approximately 0.032° of the visual angle at a viewing distance of 50 cm. The observer's head was stabilized with a chin rest.

For Experiment 2, displays were presented on an Iiyama Vision Master Pro514 CRT monitor (iiyama

Corporation, Hoofddorp, The Netherlands). The size of the visible area was 40.8×30.6 cm with resolution of $1,600 \times 1,200$ pixels and a refresh rate of 85 Hz. A single pixel subtended approximately 0.02° of the visual angle at a viewing distance of 70 cm. The observer's head was stabilized with a chin rest.

Displays and procedure

In all experiments, observers viewed two rotating SFM objects. Spherical and “gear” shapes were used in Experiment 1. Spheres only were used in Experiment 2. Movies for all experimental conditions are available in the online repository. A smaller subset of videos is included as Supplementary Movies S1 through S7.

Experiment 1

Two types of shapes—a sphere and a gear—were used in Experiment 1 (see videos in the online repository and Supplementary Movies S1 through S3). Individual shapes subtended approximately 6.5° of the visual angle vertically and horizontally and consisted of 500 dots distributed randomly over their surface. For the ambiguously rotating shapes, the dot diameter was equal to 0.23° (see Figure 1A through C). When we biased the direction of rotation via perspective cues, the dot diameter was systematically varied between 0.10° and 0.40° (see Figure 1B). When the direction of rotation was biased via the stereoscopic depth, participants wore red-green anaglyph glasses throughout the entire experiment, the dot diameter was 0.23° , and the object's presentation for two eyes differed by 2° of rotation around the vertical axis (see Figure 1A).

SFM objects rotated around the vertical axis with an angular speed of $72^\circ/\text{s}$ (0.2 Hz). Objects were placed on either side of the fixation so that they overlapped (width of the overlap region was 2°), touched (0° distance between the two objects), or had a 2° gap in between (see Figure 1A through C). In relative terms, the width of the overlap, as well as that of the gap, was 30% of the shapes' width. To facilitate perceptual grouping within a single object, dots that belonged to one shape were colored white, whereas dots of the other shape were colored yellow. Shapes alternated their color on every block.

Experiment 1 contained 18 conditions: two shapes (the sphere and the gear), three layout conditions (the overlap, touching, and the gap), and three ambiguity conditions (both ambiguous, the direction of rotation for the left object was biased, or the direction of rotation for the right object was biased). Presentation order was randomized. Each condition was presented twice in an ABBA order (36 blocks). For the disambiguated objects, the direction of rotation was

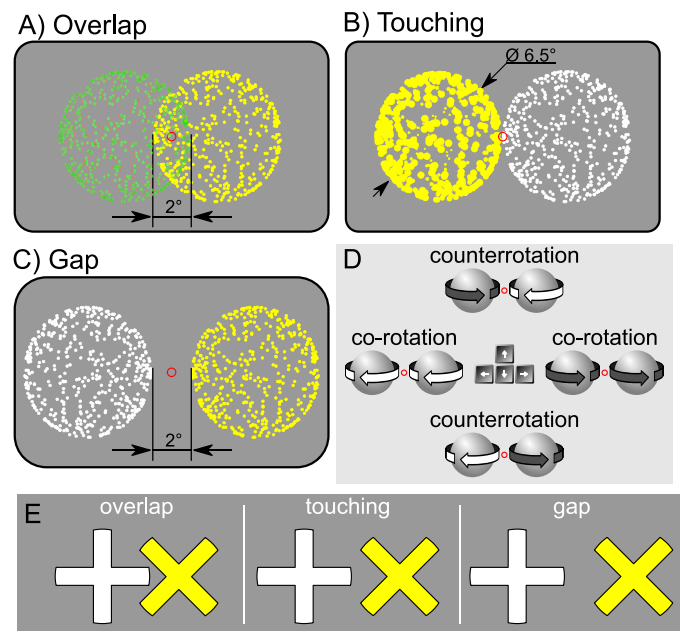


Figure 1. Experiment 1: Schematic display and response mapping. Two SFM objects, either two spheres or two gears (see E) were placed so that they overlapped (A), touched (B), or had a gap between them (C). Spheres were both ambiguous with respect to the kinetic depth (C), or one sphere (left or right) was biased towards a predefined direction of rotation using either the stereoscopic depth (A) or perspective cues (B). See also Supplementary Movies S1 through S3. (D) The perception-response mapping. Participants were instructed to press the left arrow key, if both objects rotated to the left; the right key, if both objects rotated to the right; the up key, if the left object rotated to the right and the right object rotated to the left (described as “into the screen” to participants); the down key, if the left object rotated to the left and the right object rotated to the right (described as “out of the screen” to participants). (E) The three layout conditions for the gear-shaped objects, as viewed from above (polar projection, schematic presentation).

chosen randomly for each block. Eleven participants viewed shapes that were disambiguated via perspective cues, whereas for the other six participants the direction of rotation was disambiguated via the stereoscopic depth. In the latter case, all participants were informally tested prior to the experimental session to ensure that they perceive the three-dimensional (3-D) rotation and can correctly identify the biased direction of rotation. For further details on the effectiveness of the disambiguation cues, please refer to the Results section.

Individual blocks lasted 1 min. During each block, observers viewed the continuous presentation of the two rotating SFM objects, while fixating at the central point. They were instructed to report on the perceived direction of objects' rotation using arrow keys as follows (see Figure 1D). Left, if for both objects the direction of rotation was to the left (i.e., object's front

surface moved to the left). Right, if both objects rotated to the right. Down, if the left object rotated to the left, whereas the right object rotated to the right. Up, if the left object rotated to the right and the right object rotated to the left. The first two cases correspond to the co-rotation (both shapes rotate in the same direction), whereas the latter two correspond to the counter-rotation (shapes rotate in opposite directions). All participants performed several practice blocks to accommodate themselves with the experimental procedure (data was not recorded). Participants reported no difficulty in carrying out the task.

Experiment 2

In Experiment 2, spheres subtended approximately 6° of the visual angle vertically and horizontally and consisted of 500 dots distributed randomly over their surface. The diameter of the individual dots was 0.06° . Spheres rotated around either vertical or horizontal axis with an angular speed of $72^\circ/\text{s}$ (0.2 Hz). Objects were placed to the left and to the right of the fixation (horizontal arrangement) or above and below the fixation (vertical arrangement). To facilitate perceptual grouping, one object was colored red, whereas the other one was colored green. The gap between the objects was 0, 0.1, 0.25, 0.5, and 0.8 sphere widths or, respectively, 0.0° (the touching layout configuration in Experiment 1), 0.6° , 1.5° , 2.0° (the gap layout in Experiment 1), and 4.8° of visual angle. See videos in the online repository and Supplementary Movies S4 through S7.

Displays were presented either on a uniform gray background (uniform background condition) or on the textured background (textured background condition). The textured background consisted of a randomly placed grayscale overlapping rectangles and was generated anew for each block.

Experiment 2 contained 40 conditions: five gap sizes (0, 0.1, 0.25, 0.5, and 0.8 sphere widths) \times two layouts (horizontally or vertically arranged objects) \times two directions of rotation (around the vertical or horizontal axis) \times presence/absence of the background. Presentation order was randomized. The 40 blocks were split into four experimental sessions.

Individual blocks lasted 5 min. During each block, observers viewed the continuous presentation of the two rotating SFM objects, while fixating at the central point. They were instructed to report on the direction of their rotation using arrow keys as follows. They used the left arrow when shapes rotated in the opposite directions (counter-rotation). The right key indicated when both shapes rotated in the same direction (co-rotation). The down key was for when the perception was unclear ($1.27\% \pm 0.18\%$ of the total time).

Statistical analysis

All statistical comparisons were performed in R (R Core Team, 2016), using *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016) for linear mixed-model analysis.

Data availability

All data files and the code, which performs the statistical analyses and produces the figures, are available under Creative Commons Attribution 4.0 International Public License at <https://osf.io/euzjb>.

Results

Experiment 1

In our first experiment, we sought to replicate the frictional interaction reported by Gilroy and Blake (2004) and to extend their findings by using a wider range of conditions. Specifically, we added a gear shape (see Figure 1E and Supplementary Movies S1 through S3) and an overlap condition. We hoped that the combination of the two would produce a “gear meshing” interaction, which is potentially visually stronger than the friction between two spheres. We also systematically varied ambiguity of the SFM displays to examine whether the counterrotation is more prevalent when one of the shapes is disambiguated, as the co-rotation bias tends to be stronger when both shapes are rotating ambiguously (Grossmann & Dobbins, 2003). In addition to the between-group design for the disambiguation cues (see below), we used 12 conditions, as compared to two in the first experiment of the original study. Two different shapes (a sphere and a gear), two different ambiguity conditions (both shapes rotating ambiguously versus one ambiguous and one disambiguated shape), and three different spatial layouts (an overlap, touching, and a gap; see Figure 1A through C).

We used two types of the disambiguation cues for two groups of participants. Eleven participants saw the perspective cues, where dots “closer” to the viewer were rendered bigger. The other six participants viewed the rotation disambiguated via the stereoscopic depth and were wearing anaglyph glasses for the entire duration of the experiment. Stereoscopic depth appeared to be producing a stronger bias, although the difference between the two methods was not statistically significant (see Figure 2A and Table 1).

To characterize the perception of rotation, we computed the proportion of time that counterrotation was reported throughout a single block ($P_{counterrotation}$). We used the multilevel linear mixed-effects models (Bates et al., 2015; Kuznetsova et al., 2016) with the spatial layout, the presence of a disambiguated rotating object, the disambiguation method, the object’s shape, and the interaction between the shape and the layout as independent factors.

Results of Experiment 1 are summarized in Figure 2B through E and Table 2. In contrast to the prior study (Gilroy & Blake, 2004) and consistent with the common state bias effect (Grossmann & Dobbins, 2003), observers predominantly reported co-rotation in

Model	<i>df</i>	AIC	Log likelihood	χ^2	<i>p</i> value
Observer	3	-168.1	87.0		
+ layout	4	-168.3	88.2	2.3	0.13
+ shape	5	-168.4	88.7	1.0	0.31
+ disambiguation method	6	-168.7	90.3	3.3	0.07

Table 1. The effectiveness of the disambiguation cues in Experiment 1. *Notes:* The results of the statistical analysis using the linear mixed-effects models with the proportion of time a disambiguated object was rotating in the direction of the bias (P_{bias}) as a dependent variable. The spatial layout, the object’s shape, and the disambiguation method were independent factors. The observer identity was a nested random effect. *df* = degrees of freedom; AIC = Akaike’s Information Criterion.

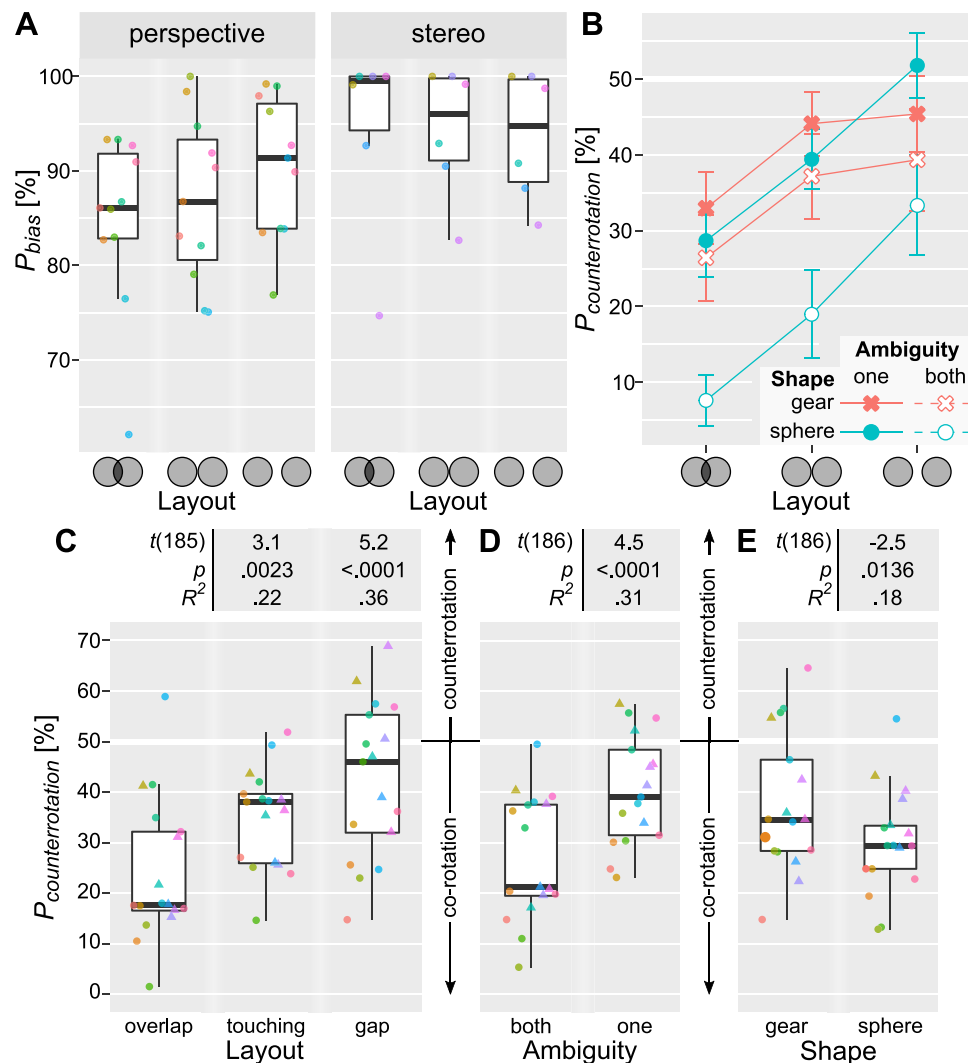


Figure 2. Experiment 1. (A) The effectiveness of perspective and stereoscopic depth disambiguation cues. P_{bias} is a proportion of time a disambiguated object was rotating in the direction of the bias. (B) The proportion of time that the objects were perceived as counterrotating ($P_{counterrotation}$). Error bars depict ± 1 SEM. (C–E) The main effect of the spatial layout (C), the ambiguity (D), and the shape (E). Circles and triangles depict individual observers (color) and the disambiguation method, respectively, via circle perspective cues and via triangle stereoscopic depth. Tables above the plot show the *t* statistics (Satterthwaite approximations to degrees of freedom), the corresponding *p* values, and effect sizes when comparing $P_{counterrotation}$ for the corresponding condition and that for the baseline (leftmost) condition. We performed the statistical comparison using a linear mixed-effects model with the corresponding factor as a single independent factor and an observer identity as a nested random effect.

Model	df	AIC	Log likelihood	χ^2	p value
Observer	3	−18.1	12.0		
+ layout	5	−40.1	25.0	26.0	<0.0001
+ shape	6	−45.1	28.6	7.1	0.0078
+ ambiguity	7	−66.4	40.2	23.3	<0.0001
+ disambiguation method	8	−64.9	40.5	0.5	0.4806
+ shape:layout	10	−65.3	42.6	4.3	0.1140

Table 2. The proportion of time observers reported the counter-rotation in Experiment 1. *Notes:* The results of the statistical analysis using the linear mixed effects models with the proportion of time that the objects were perceived as counter-rotating ($P_{counterrotation}$) as a dependent variable. The spatial layout, the object's shape, the ambiguity of the displays, the disambiguation method, as well as the interaction between the shape and the layout were independent factors. The observer identity was a nested random effect. *df* = degrees of freedom; AIC = Akaike's Information Criterion.

all conditions that we used. Moreover, the preference for co-rotation was maximal for the overlap layout and minimal for the gap condition (Figure 2B and C). In agreement with the previous results (Grossmann & Dobbins, 2003; Ramachandran & Anstis, 1983, 1985), we found a stronger tendency towards co-rotation when both objects were ambiguous (Figure 2D). Finally, the co-rotation was significantly more dominant for the sphere than for the gear shape (Figure 2E). All in all, we failed to replicate the frictional interaction effect observed by Gilroy and Blake (2004).

Given that our results reflect predominantly the common state bias, we examined them further by looking at the frequency and the timing of perceptual reversals. We quantified the strength of the perceptual coupling as $|P_{counterrotation} - 50|/50$, with values close to zero indicating independence and values close to 1 showing a strong perceptual coupling, due to either co- or counterrotation. The analysis of the blocks when both objects were ambiguously rotating showed that the stronger perceptual coupling tended to destabilize perception of rotation for the individual objects ($t[188.52] = 2.7$, $p = 0.0078$, $R^2 = 0.19$, linear mixed-model analysis with the mean number of switches as a dependent variable, strength of perceptual coupling as an independent variable, and observer identity as a random factor). Consistently, perceptual switches were more numerous for the spatial layout conditions that produced stronger perceptual coupling (see Figure 2C and 3D).

Next, we examined whether perceptual reversals were synchronized so that both objects changed their state simultaneously. In this case, observers would report a switch from one co-rotation state to another, rather than to a counterrotation state (or vice versa). This perceptual “trapping” within a pair of states

would be similar to that observed for a quadro-stable binocular rivalry (Suzuki & Grabowecky, 2002). Alternatively, the perceptual transition could happen predominantly from the co-rotation to the counter-rotation or vice versa. This perceptual “switching” would indicate that the two objects switch independently from each other.

To quantify the perceptual trapping versus switching, we selected blocks that had at least five perceptual switches reported for each object. This yielded 123 blocks from 15 observers. For each block, we computed the proportion of time that counterrotation was reported throughout a block ($P_{counterrotation}$) and the proportion of perceptual switches that were reported to occur simultaneously for both objects ($P_{simultaneous}$; Figure 3A; note, however, that due to the time required for a participant to make a perceptual decision, at least some simultaneous switch reports are likely to reflect consecutive switches that occurred within a short time span). We found a strong and highly significant dependence between the perceptual coupling and the proportion of simultaneous switches ($t[121] = 12.3$, $p < 0.0001$, $R^2 = 0.55$, a linear mixed-model analysis with the proportion of simultaneous switches as a dependent variable, strength of perceptual coupling as an independent variable, and observer identity as a random factor). Thus, strong perceptual coupling synchronized both the perceptual states and the perceptual reversals of the two ambiguously rotating SFM objects.

To further examine perceptual trapping versus perceptual switching, we selected only the blocks within 25%–75% range for $P_{counterrotation}$ (46 blocks from 11 participants) and used randomization testing to quantify how the observed switching pattern was different from the one expected by chance. We computed the probability of the perceptual trapping for the perceptual sequence as

$$P_{trapping} = \frac{\sum_{i=1}^{N-1} S_i == S_{i+1}}{N-1}, \quad (1)$$

where S_i is the i^{th} perceptual state (either counter-rotation or co-rotation) and N is the total number of reported perceptual states in the block. First, we computed $P_{trapping}$ for the original perceptual sequence and then 1,000 times for a randomly shuffled sequence (sampling without replacement). The latter gave us the chance-level distribution of $P_{trapping}$. Then, we converted the $P_{trapping}$ of the original sequence to a z score ($Z_{trapped}$) using the distribution of chance-level persistence. This gave us a common statistic across all blocks with positive values indicating perceptual trapping and negative values indicating switching (e.g., observers were more likely to switch to counterrotation from co-rotation than to a different co-rotation state). The distribution of z scores is shown in Figure 3B, with different colors representing individual observers. The

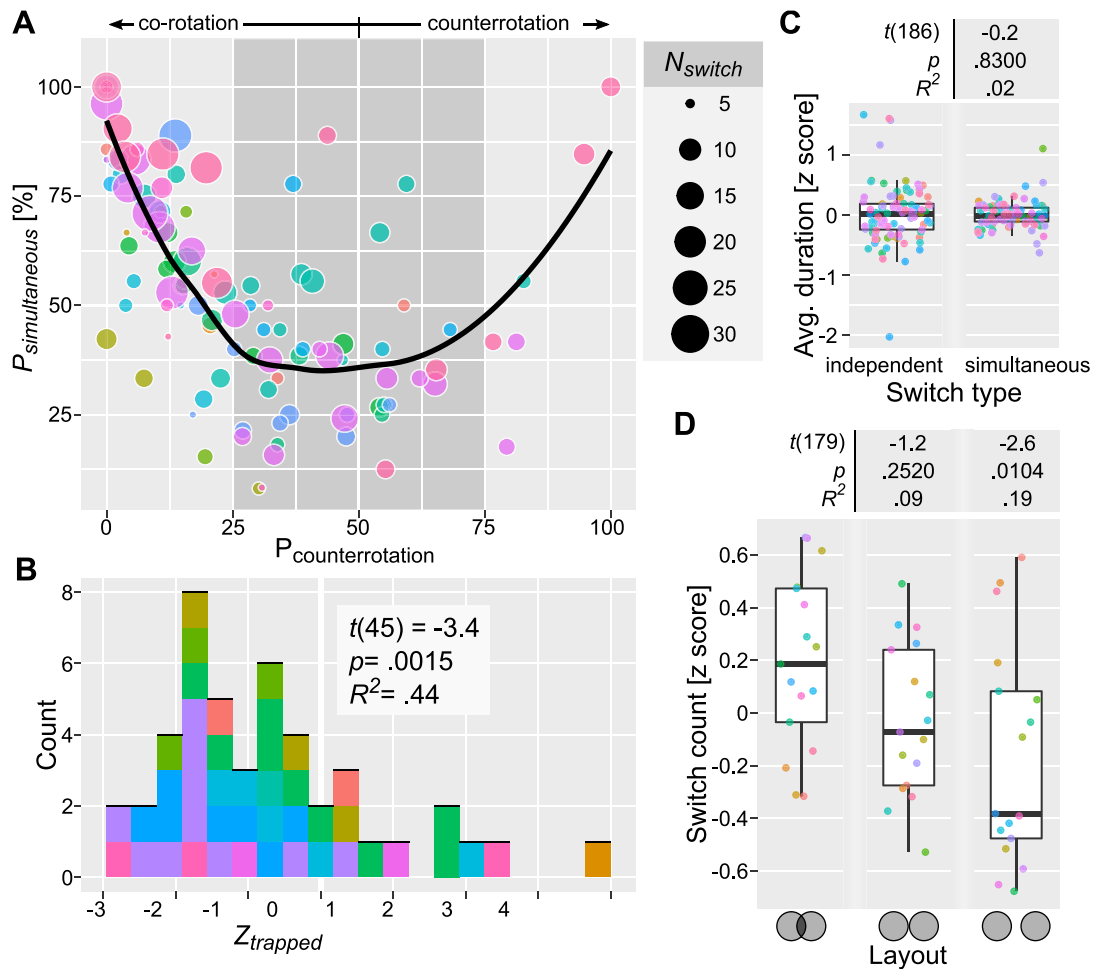


Figure 3. Simultaneous switching and perceptual stability. (A) The proportion of simultaneous switches in both objects ($P_{\text{simultaneous}}$) versus the proportion of the time participants reported counter-rotation ($P_{\text{counterrotation}}$). The size of a circle depicts the mean number of dominance switches reported for both objects (N_{switch}). Colors label individual observers. The solid line depicts a sliding average computed via the loess method. (B) The distribution of z scores that characterizes whether perception on the individual trials tended to be trapped within counterrotation or co-rotation percept pairs ($Z_{\text{trapped}} > 0$) or tended to switch primarily from counterrotation to co-rotation states or vice versa ($Z_{\text{trapped}} < 0$). Colors depict individual observers. Z_{trapped} was computed only for the blocks with $P_{\text{counterrotation}}$ between 25% and 75% (dark gray stripe in A). The presented statistics is for the one-sample t test against the mean of zero. (C) A normalized average dominance phase duration following an independent perceptual reversal (the other object remained stable) or when both objects switched simultaneously. Colors label individual observers. The table above the plot shows the t statistics (Satterthwaite approximations to degrees of freedom), the corresponding p values, and the effect size when comparing average dominance phase duration between conditions. We performed the statistical comparison using a linear mixed-effects model with the reversal type as a single independent factor and an observer identity as a nested random effect. (D) Dependence of the mean number of switches (z score) on the spatial layout. The table above the plot shows the t statistics (Satterthwaite approximations to degrees of freedom), the corresponding p values, and the effect size when comparing an average number of switches to the overlap condition. We performed the statistical comparison using a linear mixed-effects model with the spatial layout as a single independent factor and an observer identity as a nested random effect.

analysis shows that switching dominated the perception for the selected blocks.

As shown above, perceptual coupling increased the proportion of simultaneous switches but, at the same time, destabilized the individual objects. Given this negative dependence between the perceptual coupling and perceptual stability, we wondered whether the same is true for the individual dominance phases.

Specifically, we compared durations of the dominance phases that were preceded by a simultaneous switch in both objects and when objects switched independently. In the former case, the two objects retained the original perceptual coupling, whereas in the latter one this original coupling breaks down. To test this hypothesis, we computed an average dominance phase duration for individual *fully ambiguous* objects either following an

independent perceptual reversal ($D_{independent}$) or following a simultaneous switch in both objects ($D_{simultaneous}$). The duration of the dominance phase was normalized by computing a z score for all dominance durations of the corresponding perceptual state within a block. Please note that the first and the last dominance phases were excluded from the analysis because the former is not a perceptual switch but a perceptual choice (Noest, van Ee, Nijs, & van Wezel, 2007) and the latter dominance phase is curtailed by the end of the block. As can be seen in Figure 3C, the duration of the dominance phase was independent of the type of the preceding perceptual switch. Moreover, simultaneous switches did not lead to longer dominance phases even for the blocks with strong perceptual coupling: $t(91.88) = 0.04$, $p = 0.9696$, $R^2 = 0.004$, linear mixed-model analysis with the ratio of the $D_{simultaneous}$ to $D_{independent}$ as a dependent variable, balance ($|P_{counter-rotation} - 50|/50$) as an independent variable, and the observer identity as a random factor. Thus, the simultaneity of perceptual switches does not lead to either perceptual stability or destabilization.

To summarize, we observed that the perception of two rotating SFM objects was dominated by co-rotation, rather than by counterrotation, even when two objects were touching each other. Consistent with prior work (Grossmann & Dobbins, 2003), the strength of perceptual coupling was reduced when two objects were separated by the gap and when the rotation for one of the objects was disambiguated. In addition, we found that stronger perceptual coupling decreased the overall perceptual stability and increased the number of simultaneous perceptual reversals. However, the reversals themselves (simultaneous or independent) did not predict the stability of the following perceptual state.

Experiment 2

In our first experiment, we failed to replicate the frictional interaction observed by Gilroy and Blake (2004). Although our number of participants was more than three times larger than in the original study ($N = 17$ vs. $N = 5$), we could not rule out the possibility that it was not representative. To be more certain, a similar experiment was carried out at the University of Magdeburg as part of the practicum for master's students of the Integrative Neuroscience program. Participants were not informed about the results of Experiment 1. However, because it was part of the practicum, all participants read the original study and, therefore, had a full knowledge of the experimental hypotheses and of expected perceptual outcomes (e.g., that counterrotation was expected to dominate the perception when objects touched each other). This

Model	df	AIC	Log likelihood	χ^2	p value
Observer	3	-255	130		
+ distance	4	-355	182	102.4	<0.0001
+ layout	5	-357	184	3.9	0.047
+ rotation axis	6	-361	186	5.7	0.017
+ axes configuration	7	-457	236	98.4	<0.0001
+ background	8	-461	239	5.9	0.016
+ axes configuration: background	9	-463	241	4.2	0.040

Table 3. The proportion of time observers reported the counter-rotation in Experiment 2. *Notes:* Results of the linear mixed-effects model analysis with the distance between the objects, the spatial arrangement (horizontally or vertically arranged objects), the axis of rotation (vertical or horizontal), the axes layout (collinear or parallel, effectively an interaction term for the spatial arrangement and the axis of rotation), the background type (uniform or textured), and an interaction between the axis layout and the background as fixed terms. The observer identity was a nested random effect. df = degrees of freedom; AIC = Akaike's Information Criterion.

preknowledge means that the results of the experiment cannot stand on their own. However, in conjunction with Experiment 1, they are indicative of whether the original results are representative.

The experimental conditions were modified to better suit the practicum. Specifically, the objects were always two ambiguously rotating spheres. There were five spatial layout conditions: 0, 0.1, 0.25, 0.5, and 0.8 sphere widths or, respectively, 0° (the touching layout in Experiment 1), 0.6°, 1.5°, 2° (the gap layout in Experiment 1), and 4.8°. The spheres were arranged horizontally (as in Experiment 1) or vertically (one above the other). The ambiguous spheres rotated around either the vertical or the horizontal axis. In combination with the spatial arrangement, this meant that they were either collinear (axes of rotation were aligned) or parallel to each other (as in Experiment 1). Finally, the background was either uniform gray (as in Experiment 1) or textured (see Supplementary Movies S4 through S7).

The statistical analysis was similar to the one performed in Experiment 1, with the results of the linear mixed-effects model analysis summarized in Table 3. Here, the fixed factors were the distance between the objects, the spatial arrangement (horizontally or vertically arranged objects), the axis of rotation (vertical or horizontal), the axes layout (collinear or parallel, effectively an interaction term for the spatial arrangement and the axis of rotation), the background type (uniform or textured), an interaction between the axis layout and the background. The observer identity was a nested random effect.

The results are presented in Figure 4A and they closely mirror those of Experiment 1. Again, the co-

rotation rather than the counterrotation dominated the perception in this different group of informed participants in a different university. The effect of the interobject distance was also the same, with proximity leading to the stronger co-rotation bias.

Consistent with the prior report (Grossmann & Dobbins, 2003), the collinear axes configuration was very potent in inducing the co-rotation, was less sensitive to the interobject distance, and was not affected by the textured background ($\chi^2[1] = 0.1$, $p = 0.7522$, linear mixed-effects model with the distance and the background type as fixed factors, observer identity as a random effect). In contrast, the presence of the textured background significantly reduced the perceptual coupling for the parallel axes configuration ($\chi^2[1] = 10.1$, $p = 0.0015$). The nature of the latter effect is unclear, but it is possible that the textured background provides further clues that objects are separate and independent (Benmussa, Aissani, Paradis, & Lorenceau, 2011). In this case, the collinear axes might override this influence by providing stronger evidence for the single object or by benefiting from a more direct connectivity (Christiaan Klink, Noest, Holten, Van Den Berg, & Van Wezel, 2009). Conversely, the parallel axes configuration is harder to interpret as a single object and this implied independence might be amplified by the context provided by the textured background.

As in Experiment 1, we found that stronger perceptual coupling leads to more frequent simultaneous perceptual reversals (Figure 4B; $t[353.3] = 22.4$, $p < 0.0001$, $R^2 = 0.76$, a linear mixed-effects model analysis with the proportion of simultaneous switches as a dependent variable, strength of perceptual coupling as an independent variable, and observer identity as a random factor). Similarly, we found no evidence of perceptual trapping, but rather of perceptual switching (Figure 4C). Given the similarity of results of Experiments 1 and 2, we conclude that the results of Experiment 1 are unlikely to reflect a suboptimal participants sample and, therefore, are likely to be externally valid.

Discussion

The primary goal of the study was to replicate a “frictional interaction” effect observed for two touching rotating SFM spheres (Gilroy & Blake, 2004) and to examine its interaction with a more general common state bias (Attneave, 1968; Eby et al., 1989; Gillam, 1972, 1976; Grossmann & Dobbins, 2003; Landy, 1987; Ramachandran & Anstis, 1983, 1985). Moreover, we extended the original setup by creating a similar but potentially stronger “meshing gears” effect using gear-

shaped objects. Unfortunately, we were unable to replicate the original effect or to observe the gear meshing effect. We used a broader set of conditions, including two SFM shapes, two types of disambiguation cues, the presence or absence of the disambiguation cues, more layout options, and two samples of observers from two different universities totaling 26 participants. Despite all that, we only observed co-rotation consistent with the common state bias.

Physics embedded in visual perception

The lack of the frictional interaction in our setup might stem from the difference in the experimental displays. The SFM spheres in the original study were much smaller than in our study (1.06° vs. 6°) and, also, than a typical SFM display (e.g., Brouwer & van Ee, 2006; Klink, van Ee, & van Wezel, 2008; Maier, Wilke, Logothetis, & Leopold, 2003). They were also much denser (250 dots) and rotated somewhat faster (0.33 Hz vs. 0.2 Hz), although at a lower update rate of 20 Hz, as compared to 120 Hz in Experiment 1 and 85 Hz in Experiment 2. Thus, given a strictly monotonic relationship between the interobject distance and the dominance of co-rotation that we observed, it is possible that only very specific displays elicit the perception of the frictional interaction.

This high specificity might reflect the fact that the frictional interaction between the two spheres works only if they are precisely aligned. Even a minimal gap, misalignment in depth or along the vertical axis, or a low friction surface would prevent it in real life. In fact, all these considerations prompted us to add the meshing gears display. However, the same high specificity means that inferring a frictional interaction for the two spheres requires high certainty that the visual scene meets the requirements listed above. Even when such checks are possible, the physical configuration leading to it is uncommon. Or, at least, not more likely than encountering objects rolling on a surface in the same direction (e.g., apples rolling downhill). The latter situation is less reliant on the precise configuration of a visual scene. Thus, when the frictional interaction is compared to a more generally applicable and more robust common state bias, it raises questions both about a benefit that such narrowly aimed heuristic would provide for the visual system when employed at the perceptual level and about a feasible implementation of a model for this physical interaction.

A more general question would be whether we should expect embedded physics to extend beyond perceptual states of individual objects. As noted in the introduction, there are numerous examples of physics/world statistics dictating the perception of individual objects in the presence of noise or ambiguity (Gerardin

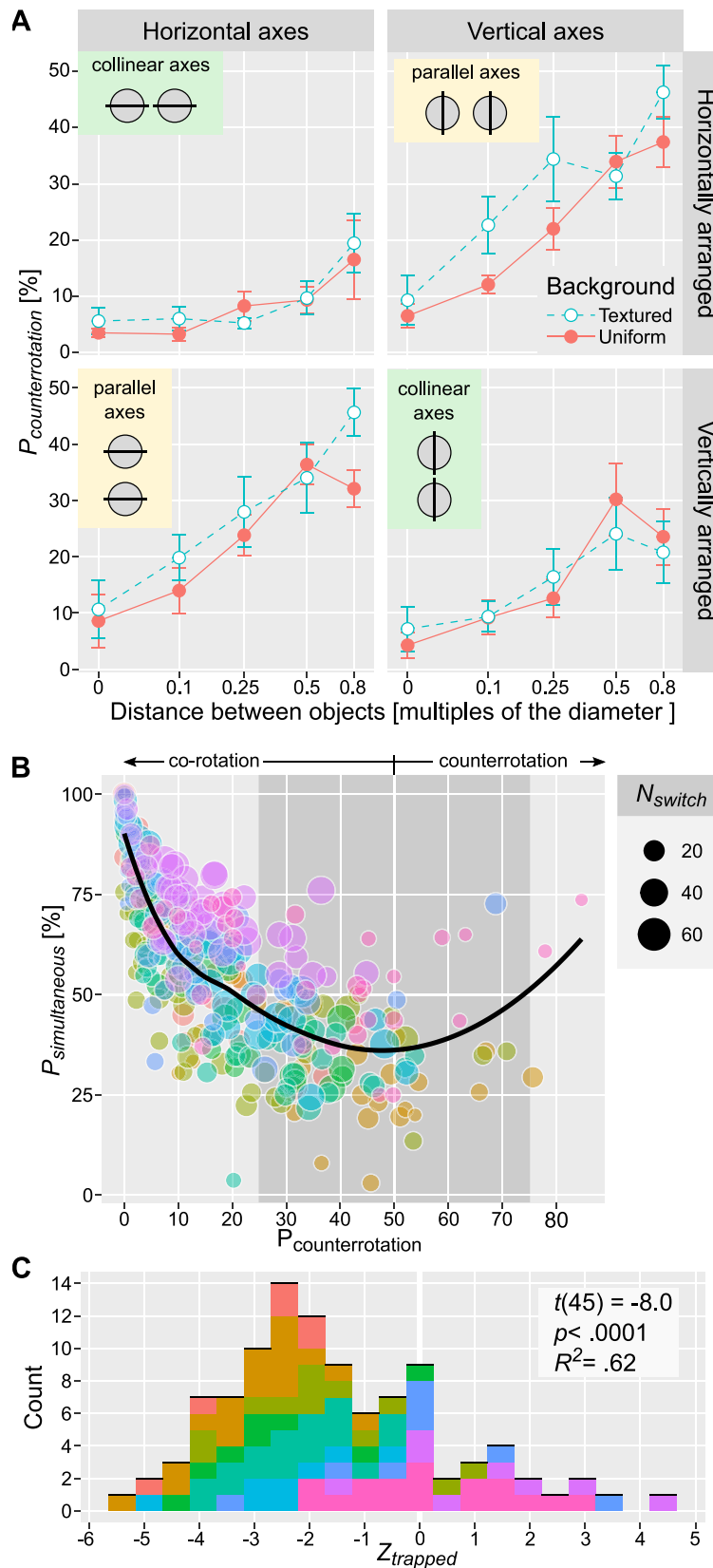


Figure 4. Experiment 2. The proportion of time when participants reported the perception of counterrotation ($P_{counterrotation}$) as a function of the interobject distance, the axes orientation, the axes layout, and the background type. Error bars depict ± 1 SEM. (B) The proportion of simultaneous switches in both objects ($P_{simultaneous}$) versus the proportion of the time participants reported

→

←

counterrotation ($P_{\text{counterrotation}}$). The size of a circle depicts the mean number of switches reported for both objects (N_{switch}). Colors label individual observers. The solid line depicts a sliding average computed via the loess method. (C) The distribution of z scores that characterizes whether perception on the individual trials tended to be trapped within counterrotation or co-rotation percept pairs ($Z_{\text{trapped}} > 0$) or tended to switch primarily from counterrotation to co-rotation states or vice versa ($Z_{\text{trapped}} < 0$). Colors depict individual observers. Z_{trapped} was computed only for the blocks with $P_{\text{counterrotation}}$ between 25% and 75% (dark gray stripe in B). Statistics for the one-sample t test against the mean of zero.

et al., 2010; Gregory, 1997; Pastukhov et al., 2012). However, examples of such multistable interactions are much harder to find with a streaming-bouncing display being the only obvious example (Metzger, 2009). Here, two objects either “stream” through each other or “bounce” off each other due to an elastic collision. However, even in this case, its model could be considered incomplete, as the auditory-visual integration relies on the timing of the perceptual events rather than of the physical ones (Arnold, Johnston, & Nishida, 2005). A possible reason for this scarcity is that maintaining a compendium of interactions could be very expensive due to a potential combinatorial explosion or due to a need for high precision models of such interactions (as is the case for the frictional interaction). We would be able to understand those limitations and mechanisms only once we have a list of reliable perceptual interaction effects at our disposal.

Influence of expectations on multistable perception

Another way to reconcile our results with those in Gilroy and Blake (2004) is by assuming a different set of expectations that observers had about the display. Participants can control the appearance of multi-stable displays to a large degree (Hol, Koene, & van Ee, 2003; Meng & Tong, 2004; Mitchell, Stoner, & Reynolds, 2004). Moreover, an expectation of a particular outcome may strongly bias the observer towards that perceptual state (Bugelski & Alampay, 1961; Sterzer, Frith, & Petrovic, 2008).

Accordingly, it is possible that the frictional interaction reflected the observers’ expectations about how the two spheres should interact. This would make it a cognitive rather than perceptual bias but would explain its high specificity. In that case, however, it is unclear why similar expectations failed to produce a comparable cognitive bias for participants of Experiment 2.

The common state bias

As noted above, the perception of the SFM objects was highly consistent with a common state bias reported for various multi-stable displays (Adams &

Haire, 1958; Attneave, 1968; Eby et al., 1989; Gillam, 1972; Gillam, 1976; Grossmann & Dobbins, 2003; Landy, 1987; Ramachandran & Anstis, 1983; Ramachandran & Anstis, 1985). Please note that we opted for the term “common state bias” instead of using a more general and commonly used term “perceptual coupling” because the two spheres can be considered perceptually coupled both when they are co-rotating and counterrotating.

Confirming the earlier results (Christiaan Klink et al., 2009; Grossmann & Dobbins, 2003), we observed weakening of this bias when one of the objects was disambiguated (Experiment 1). We also observed a stronger coupling between collinear than between parallel axes of rotation, as in the prior report (Eby et al., 1989). These results fit nicely with the idea that the common state bias in SFM is mediated by lateral connections between depth-tuned neural populations (Klink et al., 2008).

Interestingly, we found that the stronger perceptual coupling was associated with more frequent perceptual switches (see Figure 3A and D). In other words, the mutual influence of the two objects reduced rather than increased their perceptual stability. This could indicate that the perceptual coupling affects not only the dominant perceptual state, in which case we would have expected it to stabilize the perception (Chong, Tadin, & Blake, 2005), but that both currently dominant and currently suppressed neural representations reinforce their respective counterparts in the other object. This way, the perceptual coupling could increase the “effective contrast” of both perceptual alternatives equally, leading to a higher switching rate, just as an increase in the contrast does (Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006). This idea also fits well with the fact that the simultaneity of the switches in two objects does not predict the duration of the following dominance phase (Figure 3C). Again, if both the dominant and the suppressed states are perceptually coupled, a simultaneous switch would provide no additional benefits for a specific perceptual interpretation over the competitor.

In addition, we found that the perceptual coupling was stronger for spheres than for the gears (Figure 2E). This indicates that, similar to the perceptual adaptation (Pastukhov, Lissner, & Braun, 2014), the strength of the interaction might depend on the strength of the individual representations. These results appear to

contradict the lack of shape specificity observed by (Grossmann & Dobbins, 2003). However, this discrepancy most likely reflects the choice of individual objects within the study (Maier et al., 2003; Pastukhov, Füllekrug, & Braun, 2013). Accordingly, a study with a broader selection of SFM shapes would clarify the matter.

Finally, in Experiment 2 we found that the strength of perceptual coupling was affected by the presence of the textured background, although only for the parallel axes configuration. It is possible that the texture background provided evidence that two objects are separate, weakening perceptual coupling. Collinear axes overall produced stronger perceptual coupling and, therefore, may have been less affected. Alternatively, it is possible that two types of the background provided different cues about the relative motion. In case of the uniform background, it can be viewed as both static and moving, since there are no specific cues to assume its stationarity. In contrast, the textured background is clearly static and, therefore, may counteract the perception of objects rolling together. However, the prior work speaks against the mechanistic explanation of the effect (Sereno & Sereno, 1999), making the interference with lateral connections (Christiaan Klink et al., 2009) a more likely explanation of the phenomenon.

Conclusions

We report that perceptual coupling speeds up perceptual alternations and increases the proportion of simultaneous switches in two objects. However, we found that the simultaneity of the individual switches does not predict the duration of the following dominance phase. Finally, we found that the common state bias and not the frictional interaction determine the perception of two touching SFM spheres.

Keywords: embedded physics, multistable perception, structure-from-motion, kinetic-depth effect, prior knowledge

Acknowledgments

Commercial relationships: none.
Corresponding author: Alexander Pastukhov.
Email: pastukhov.alexander@gmail.com.
Address: Department of General Psychology and Methodology, University of Bamberg, Bamberg, Bavaria, Germany.

References

- Adams, P. A., & Haire, M. (1958). Structural and conceptual factors in the perception of double-cube figures. *The American Journal of Psychology*, *71*(3), 548, <https://doi.org/10.2307/1420250>.
- Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*, *45*(10), 1275–84, <https://doi.org/10.1016/j.visres.2004.11.014>.
- Attneave, F. (1968). Triangles as ambiguous figures. *The American Journal of Psychology*, *81*(3), 447, <https://doi.org/10.2307/1420645>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48, <https://doi.org/10.18637/jss.v067.i01>.
- Benmussa, F., Aissani, C., Paradis, A.-L., & Lorenceau, J. (2011). Coupled dynamics of bistable distant motion displays. *Journal of Vision*, *11*(8):14, 1–19, <https://doi.org/10.1167/11.8.14>. [PubMed] [Article].
- 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 Research*, *46*(20), 3393–402, <https://doi.org/10.1016/j.visres.2006.03.016>.
- Bugelski, B. R., & Alampay, D. A. (1961). The role of frequency in developing perceptual sets. *Canadian Journal of Psychology*, *15*, 205–211. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/13874512>
- Chong, S. C., Tadin, D., & Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of Vision*, *5*(11):6, 1004–1012, <https://doi.org/10.1167/5.11.6>. [PubMed] [Article].
- Christiaan Klink, P., Noest, A. J., Holten, V., Van Den Berg, A. V., & Van Wezel, R. J. A. (2009). Occlusion-related lateral connections stabilize kinetic depth stimuli through perceptual coupling. *Journal of Vision*, *9*(10):20, 1–20, <https://doi.org/10.1167/9.10.20>. [PubMed] [Article].
- Eby, D. W., Loomis, J. M., & Solomon, E. M. (1989). Perceptual linkage of multiple objects rotating in depth. *Perception*, *18*(4), 427–444, <https://doi.org/10.1068/p180427>.
- Gerardin, P., Kourtzi, Z., & Mamassian, P. (2010).

- Prior knowledge of illumination for 3D perception in the human brain. *Proceedings of the National Academy of Sciences, USA*, 107(37), 16309–16314, <https://doi.org/10.1073/pnas.1006285107>.
- Gillam, B. (1972). Perceived common rotary motion of ambiguous stimuli as a criterion of perceptual grouping. *Perception & Psychophysics*, 11(1), 99–101, <https://doi.org/10.3758/BF03212694>.
- Gillam, B. (1976). Grouping of multiple ambiguous contours: Towards an understanding of surface perception. *Perception*, 5(2), 203–209, <https://doi.org/10.1068/p050203>.
- Gilroy, L. A., & Blake, R. (2004). Physics embedded in visual perception of three-dimensional shape from motion. *Nature Neuroscience*, 7(9), 921–922, <https://doi.org/10.1038/nn1297>.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 352(1358), 1121–1127, <https://doi.org/10.1098/rstb.1997.0095>.
- Grossmann, J. K., & Dobbins, A. C. (2003). Differential ambiguity reduces grouping of metastable objects. *Vision Research*, 43(4), 359–369, [https://doi.org/10.1016/S0042-6989\(02\)00480-7](https://doi.org/10.1016/S0042-6989(02)00480-7).
- Hol, K., Koene, A., & van Ee, R. (2003). Attention-biased multi-stable surface perception in three-dimensional structure-from-motion. *Journal of Vision*, 3(7):3, 486–498, <https://doi.org/10.1167/3.7.3>. [PubMed] [Article].
- Jackson, S., & Blake, R. (2010). Neural integration of information specifying human structure from form, motion, and depth. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(3), 838–848, <https://doi.org/10.1523/JNEUROSCI.3116-09.2010>.
- Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, 13(2), 150–158, [https://doi.org/10.1016/S0959-4388\(03\)00042-4](https://doi.org/10.1016/S0959-4388(03)00042-4).
- Klink, P. C., van Ee, R., & van Wezel, R. J. A. (2008). General validity of Levelt's propositions reveals common computational mechanisms for visual rivalry. *PLoS One*, 3(10), e3473, <https://doi.org/10.1371/journal.pone.0003473>.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmerTest: Tests in linear mixed effects models. Retrieved from <https://cran.r-project.org/package=lmerTest>
- Landy, M. S. (1987). Parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, 4(5), 864, <https://doi.org/10.1364/JOSAA.4.000864>.
- 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).
- 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].
- Metzger, W. (2009). *Laws of seeing* (L. Spillmann, S. Lehar, M. Stromeyer, & M. Wertheimer, Trans.). Cambridge, MA: MIT Press.
- Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature*, 429(6990), 410–413, <https://doi.org/10.1038/nature02584>.
- Noest, A. J., van Ee, R., Nijs, M. M., & van Wezel, R. J. A. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *Journal of Vision*, 7(8):10, 1–14, <https://doi.org/10.1167/7.8.10>. [PubMed] [Article].
- 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>.
- Pastukhov, A., Vonau, V., & Braun, J. (2012). Believable change: Bistable reversals are governed by physical plausibility. *Journal of Vision*, 12(1):17, 1–16, <https://doi.org/10.1167/12.1.17>. [PubMed] [Article].
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: Authors. Retrieved from <https://www.r-project.org/>
- Ramachandran, V. S., & Anstis, S. M. (1983, August 11–17). Perceptual organization in moving patterns. *Nature*, 304(5926), 529–531, <https://doi.org/10.1038/scientificamerican0686-102>.
- Ramachandran, V. S., & Anstis, S. M. (1985). Perceptual organization in multistable apparent motion. *Perception*, 14(2), 135–143, <https://doi.org/10.1068/p140135>.
- Sereno, M. E., & Sereno, M. I. (1999). 2-D center-surround effects on 3-D structure-from-motion. *Journal of Experimental Psychology: Human Per-*

ception and Performance, 25(6), 1834–1854, <https://doi.org/10.1037/0096-1523.25.6.1834>.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: Expectations alter visual awareness. *Current Biology*, 18(16), R697–R698, <https://doi.org/10.1016/j.cub.2008.06.021>.

Suzuki, S., & Grabowecky, M. (2002). Evidence for perceptual “trapping” and adaptation in multi-stable binocular rivalry. *Neuron*, 36(1), 143–157, [https://doi.org/10.1016/S0896-6273\(02\)00934-0](https://doi.org/10.1016/S0896-6273(02)00934-0).

Supplementary material

Supplementary Movie S1. Experiment 1. Shape – cross, layout – overlap, stereoscopic depth – right object.

Supplementary Movie S2. Experiment 1. Shape – sphere, layout – touching, perspective cues – left object.

Supplementary Movie S3. Experiment 1. Shape – cross, layout – gap, both ambiguous.

Supplementary Movie S4. Experiment 2. Axis – horizontal, layout – horizontal, distance – 0, background – texture.

Supplementary Movie S5. Experiment 2. Axis – horizontal, layout – vertical, distance – 0.1, background – uniform.

Supplementary Movie S6. Experiment 2. Axis – vertical, layout – horizontal, distance – 0.5, background – texture.

Supplementary Movie S7. Experiment 2. Axis – vertical, layout – vertical, distance – 0, background – uniform.