A generic mechanism for perceptual organization in the parietal cortex

Pablo R Grassi¹,²,³, Natalia Zaretskaya¹,²,³,⁴ and Andreas Bartels¹,²,³

¹Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany
²Department of Psychology, University of Tübingen, Tübingen, Germany
³Max Planck Institute for Biological Cybernetics, Tübingen, Germany
⁴Institute of Psychology, University of Graz, Graz, Austria

DOI: 10.1523/JNEUROSCI.0436-18.2018

Received: 16 February 2018
Revised: 29 May 2018
Accepted: 21 June 2018
Published: 13 July 2018

Author contributions: P.R.G., N.Z., and A.B. designed research; P.R.G. and N.Z. performed research; P.R.G. and N.Z. analyzed data; P.R.G. wrote the first draft of the paper; P.R.G., N.Z., and A.B. edited the paper; P.R.G. and A.B. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

We would like to thank Therese Andrae for her assistance with the collection of the data. This work was supported by the German Excellence Initiative of the German Research Foundation (DFG) grant number EXC307, by DFG grant BA4914/1-1 to A.B., and by the Max Planck Society, Germany. The authors declare no competing financial interests.

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.0436-18.2018

Alerts: Sign up at www.jneurosci.org/cgi/alerts to receive customized email alerts when the fully formatted version of this article is published.
Title: A generic mechanism for perceptual organization in the parietal cortex

Abbreviated title: Confluence of scene segmentation in parietal areas

Pablo R Grassi1,2,3, Natalia Zaretskaya1,2,3,4 and Andreas Bartels1,2,3

1) Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany
2) Department of Psychology, University of Tübingen, Tübingen, Germany
3) Max Planck Institute for Biological Cybernetics, Tübingen, Germany
4) Institute of Psychology, University of Graz, Graz, Austria

Corresponding authors:

Pablo R. Grassi
Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen
Otfried-Müller-Str. 25, 72076 Tübingen
E-mail: pablo.grassi@cin.uni-tuebingen.de
Phone: +4970712989103

Andreas Bartels
Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen
Otfried-Müller-Str. 25, 72076 Tübingen
E-mail: andreas.bartels@tuebingen.mpg.de
Phone: +4970712989168

Acknowledgements: We would like to thank Therese Andrae for her assistance with the collection of the data. This work was supported by the German Excellence Initiative of the German Research Foundation (DFG) grant number EXC307, by DFG grant BA4914/1-1 to A.B., and by the Max Planck Society, Germany. The authors declare no competing financial interests.
ABSTRACT (220)

Our visual system’s ability to group visual elements into meaningful entities and to separate them from others is referred to as scene segmentation. Visual motion often provides a powerful cue for this process as parallax or coherence can inform the visual system about scene or object structure. Here we tested the hypothesis that scene segmentation by motion cues relies on a common neural substrate in the parietal cortex. We used fMRI and a set of three entirely distinct motion stimuli to examine scene segmentation in the human brain. The stimuli covered a wide range of high-level processes, including perceptual grouping, transparent motion and depth perception. All stimuli were perceptually bistable such that percepts alternated every few seconds while the physical stimulation remained constant. The perceptual states were asymmetric, in that one reflected the default (non-segmented) interpretation, and the other the non-default (segmented) interpretation. We confirmed behaviorally that upon stimulus presentation, the default percept was always perceived first, before perceptual alternations ensued. Imaging results showed that across all stimulus classes perceptual scene-segmentation was associated with an increase of activity in the posterior parietal cortex together with a decrease of neural signal in the early visual cortex. This pattern of activation is compatible with predictive coding models of visual perception, and suggests that parietal cortex hosts a generic mechanism for scene segmentation.
Significance statement (116)

Making sense of cluttered visual scenes is crucial for every-day perception. An important cue to scene segmentation is visual motion: slight movements of scene elements give away which elements belong to the fore- or background or to the same object. We used three distinct stimuli that engage visual scene segmentation mechanisms based on motion. They involved perceptual grouping, transparent motion and depth perception. Brain activity associated with all three mechanisms converged in the same parietal region with concurrent deactivation of early visual areas. The results suggest that posterior parietal cortex is a hub involved in structuring visual scenes based on different motion cues, and that feedback modulates early cortical processing in accord with predictive coding theory.
A fundamental task of our visual system is to make sense of complex visual scenes. This feat involves grouping visual elements into objects, fore- and background, or other coherent units. Visual motion provides a prominent segmentation cue due to the three-dimensional nature of our environment: small head- or object-movements give away which visual elements belong to the same object, occluder, or depth-level, and hence aide scene segmentation.

Prior knowledge of likely scene configurations clearly helps in this process. Scene segmentation can hence be thought of as an inference process in which bottom-up sensory information is combined with top-down prior knowledge to predict the most probable cause of the sensory inputs – a definition equivalent to the theory of predictive coding (Rao and Ballard, 1999; Friston, 2005). Within this predictive coding framework, during scene segmentation higher-level visual areas send top-down signal to the early visual cortex via feedback projections. However, the higher-level neural sources of the feedback signal mediating segmentation are largely unknown.

Prior studies focused primarily on the role of early or feature-selective visual regions in perceptual organization, such as the motion responsive V5+/MT+ complex (Murray et al., 2003; Muckli et al., 2005) or the shape responsive lateral occipital complex (LOC) (Murray et al., 2002b). The role of parietal cortex in scene segmentation has been by far less studied, even though clinical and experimental studies have shown that its disruption impairs grouping of visual items into a whole (Himmelbach et al., 2009; Karnath et al., 2000; Luria, 1959; Romei et al., 2011; Wolpert, 1924), and recent MEG evidence points to parietal cortex as initiating global shape detection, causally modulating early visual cortex subsequently (Liu et al., 2017).

Of the few previous studies examining a role of parietal cortex in scene segmentation, almost all either used visual stimuli that differed physically between conditions (Murray et al., 2003; Paradis et al., 2000; Yokoi and Komatsu, 2009; Zeki and Stutters, 2013), compared different attentional states and tasks (Fink et al., 1996; Mevorach et al., 2006; Romei et al., 2011), or examined a
singular specialized stimulus that did not allow for generalization (Grassi et al., 2016; Liu et al., 2017; Zaretskaya et al., 2013).

In this study we tested the hypothesis that parietal cortex is invariably involved in visual motion-based perceptual scene segmentation. We used a battery of stimuli that involved entirely distinct aspects of scene segmentation with or without the emergence of Gestalt, including perceptual grouping, occlusion and transparent motion. Importantly, all stimuli were *asymmetrically* bistable, such that observers either perceived a default (2D) or an alternative (3D segmented) interpretation. Such asymmetric bistable stimuli provide an ideal means to study scene segmentation: processes related to subjectively experienced scene segmentation can be tracked while physical stimulus properties stay constant and are thus fully controlled for. We examined behavioral responses and associated brain activity during viewing of three bistable displays, involving a wide range of scene interpretations, ranging from emerging shapes (moving occluded diamond versus moving elements), depth-structure (plaid versus component motion) and perceptual grouping (illusory Gestalt motion versus element motion). Most previous studies using the first two of these stimuli constrained their analyses to feature-selective regions LOC or V5/MT due to the expectation that these mediate the percepts (Castelo-Branco et al., 2002; Murray et al., 2002b; Villeneuve et al., 2005; Fang et al., 2008) and parietal involvement remained largely unexplored. The illusory Gestalt display had been examined previously (Grassi et al., 2016; Zaretskaya et al., 2013) but it was unclear whether results were specific to Gestalt perception or would generalize to other types of perceptual organization not involving Gestalt.

If parietal cortex is generally involved in perceptual organization, we expect its activation during the alternative segmented percepts invariantly across all stimuli. In contrast, if scene-segmentation is completed in feature selective regions we expect their activation without systematic involvement of the PPC.


**Materials and Methods**

**Subjects**

For the Diamond and Plaid experiments, a total of eighteen volunteers (mean age 26.2 ± 2.9 SD, 11 female, 7 male) participated in the study after signing an informed consent form. All had normal or corrected-to-normal vision and no history of neurological impairments. The study was conducted according to the Declaration of Helsinki and was approved by the ethics committee of the University Clinic Tübingen. One subject was excluded from the analysis based on aberrant behavioral responses and head movement. A further subject was excluded from the ROI analysis of motion selective regions because of technical problems during the acquisition of motion localizer data. Hence, 17 subjects entered the final whole-brain analysis, and 16 the motion-selective ROI-based analysis (see below). Data for the illusory Gestalt stimulus were acquired in eighteen separate participants in our previous study (see Zaretskaya et al., 2013).

**Asymmetric bistable displays**

The three distinct bistable stimuli are presented in Figure 1: transparent plaids (Stoner et al., 1990), the occluded Diamond (Lorenceau and Shiffrar, 1992; Murray et al., 2002b), and the illusory Gestalt display (Anstis and Kim, 2011). All data for the first and second stimuli are novel. For the third stimulus, behavioral analyses shown here are novel, and the previously published fMRI data (Zaretskaya et al., 2013) were used here for a new conjunction analysis for all three stimulus sets.

All bistable displays led to perceptual alternations between two different interpretations while the physical stimulation remained constant. Importantly, the present bistable displays were perceptually asymmetric: in the more commonly used symmetric bistable displays, both possible perceptual interpretations are more or less balanced in content and complexity, like two distinct orientations in space (e.g. Necker cube or gratings in binocular rivalry), motion directions (e.g. rotating random-dot sphere), or object-identities (face-vase or old-young woman illusion). In contrast, the perceptual interpretations of the present stimuli were *asymmetric* or imbalanced in their content. In the
following sections we refer to the two perceptual interpretations as to the “default” and “alternative” percepts, respectively, because these terms can easily be operationally defined as described below. Note though that the default percept reflected a 2D interpretation, whereas the alternative percept implied a 3D spatial arrangement.

When stimuli were adjusted such that both perceptual interpretations were balanced in their perceptual times during prolonged viewing (described in detail below), upon stimulus presentation, consistently one of the two interpretations was perceived with much higher probability. This initial percept can be defined as the “default” percept, as it is based on the initial perceptual processes, whereas the “non-default” or “alternative” interpretation is based on scene segmentation processes that follow after some time. These observations have been previously described for the plaid display (Hupé and Rubin, 2003).

Only indirect evidence consistent with this has been reported for the Diamond display in that during short stimulus presentations shape perception was difficult and was facilitated only under specific parameters (e.g. peripheral vision) (Lorenceau and Shiffrar, 1992). However, in contrast to this observation, a previous fMRI study using different variations of the Diamond stimulus did not find any systematic preference for a particular percept (Caclin et al., 2012). For the illusory Gestalt display the local percept was previously reported to occur with higher-than-chance probability at stimulus onset (Anstis and Kim 2011).

Our first analysis was hence a behavioral one examining relative frequencies of the first percept type at the onset of stimulus presentation for each of the stimuli, which allowed us to define “default” and “alternative” percepts for each stimulus type. Please note that the labeling of the percepts in Figure 1 is based on these behavioral results. Based on previous literature, we optimized each display so that both interpretations achieved roughly equal dominance times during prolonged viewing as described below.

**Plaid**

The first bistable display ("Plaid") consisted of two rectangular-wave gratings of different orientations moving in opposite directions (see Figure 1A). This bistable stimulus can be seen either as a single pattern moving in one direction
or as two segregated gratings sliding over each other, perceived as depth-structured transparent motion. The rectangular gratings were presented through a circular aperture with a diameter of 9.15° of visual angle. Individual gratings were composed of half-transparent white stripes (0.46° width, 553 cd/m²) on a gray background moving with a speed of 0.45 cycles/sec. The duty cycle, defined as the width of the stripes divided by the width of one spatial period, was 30%, i.e. white stripes were thinner than the gaps between them. The intersection regions of the rectangular gratings were brighter (646 cd/m²). The angle between individual grating motion vectors was 60°. The global pattern was moving in an oblique direction (20° from vertical), which has been shown to help achieve equiprobability between the two percepts (Hupé and Rubin, 2004).

Diamond

The second display (“Diamond”) consisted of a black and white contour drawing of a Diamond whose four corners were occluded by three vertical bars of the same gray color as the background (369 cd/m², see Figure 1B). This stimulus can be perceived as four lines moving independently up and down, but the lines can also be perceived as a bound diamond shape translating horizontally behind three illusory occluders (Murray et al., 2002b). We achieved roughly matched dominance duration times with the following parameters: the Diamond moved horizontally at a speed of 1.3° (13 subjects) or 2° (4 subjects) per second and changed direction every 1.2 s or 0.8 s, respectively. The sides of the Diamond were 5.94° long and 1° in width. In the starting position, the center of each corner of the Diamond was located at 4.2° eccentricity from fixation.

Illusory Gestalt

The illusory Gestalt stimulus consisted of four pair of dots moving in-phase on circular paths and has been studied extensively before (Anstis and Kim, 2011; Grassi et al., 2016; Grassi et al., 2017; Zaretskaya et al., 2013). The stimulus can be perceived as local motion of individual dots or as two illusory squares sliding over each other in transparent motion (i.e. global Gestalt motion). Individual dots had a size of 0.5°, the radius between dots was 2° and the distance between each dot par and center was 5°. All dots had the same contrast polarity (either black or white). Between each run, dot-rotation direction and dot-contrast
polarity was randomly varied. Mean speed of dot rotation was 2.49 ± 0.2 rotation per second. For further details see Zaretskaya et al. (2013).

**Display methods**

Displays were generated using MATLAB 2013a (MathWorks, Natick, MA) with Psycho toolbox 3 extensions (Brainard 1997, Pelli 1997) and presented using a linearized projector with a resolution of 1920 x 1080 pixels at 60 Hz. The semitransparent screen covered 29 x 16.5° visual degrees, and was viewed at 80 cm via a mirror attached to the head coil.

**Procedure and experimental design**

**Behavioral acquaintance and fMRI task**

All participants underwent a behavioral test prior to fMRI scanning to ensure that they could perceive and report both perceptual interpretations in all displays. We described both possible percepts and let the participants observe each display for at least 4 minutes.

During the fMRI experiments, subjects were asked to constantly fixate the central red fixation dot and to report their current spontaneously occurring percept by pressing and holding down one of the two buttons (one for the default percept, one for the alternative percept) with their right hand. Participants were instructed not to press any button if unsure of their percept and not to enforce any particular percept.

**fMRI paradigm**

Each fMRI run was dedicated to one of the display types and consisted of two stimulus presentations of 120 s each, each followed by a fixation-only block of 20 s. We alternated the stimulus presentation between each fMRI run. Each subject underwent 4 experimental runs per display type (i.e. 8 runs per participant). In each fMRI run we acquired 243 volumes, including four extra volumes at the beginning of the session to control for T1 equilibration effects. Data for the third display type had been collected in a previous study with the same instructions and similar design (90 s stimulus presentation followed by 15 s fixation, repeated four times in each of five to six runs) (Zaretskaya et al., 2013).
Post-scan questionnaire

Finally, after the fMRI scan, we asked our subjects to fill in a questionnaire to assess subjective experiences beyond the percept dominances. Among other questions not relevant to this study, the questionnaire contained the following questions: Q1: “How confident are you that your button presses reflected your actual percepts?”, Q2: “Did you try to steer/guide your perception to enforce any particular percept?”, Q3: “Was one of the possible interpretations of the stimulus more difficult to see than the other one?”. Each question was rated in a seven-point Likert scale (first question: 1 = not confident, 7 = very confident; second and third questions: 1 = Alternative percept, 4 = neutral, 7 = Default percept). We tested the responses against expected median values by means of Wilcoxon signed rank tests.

MRI data acquisition

We acquired data for the first two displays on a 3 Tesla Siemens Prisma system with a 64-channel head coil (Siemens, Erlangen, Germany). All Data were acquired in the same session. Functional data were acquired using a gradient-echo echo planar imaging (EPI) sequence using T2* weighted blood oxygenation level dependent (BOLD) contrast with the following parameters: repetition time (TR) = 2.2 s, echo time (TE) = 30 ms, flip angle = 79°, isotropic voxel size of 3x3x3 mm, 36 slices and a field of view of 192 x 224 mm for whole-brain coverage. We also acquired a high-resolution T1-weighted anatomical scan of each participant (ADNI, 192 slices, voxel size 1 mm³, TR = 2 s, TE = 3.06 ms, TI = 1.1 s, flip angle = 9°, FOV = 232 x 256 mm). For details of the data acquisition parameters of the third display see Zaretskaya et al. (2013).

Data analysis

Behavioral data analysis

Determination of default and alternative percepts. To differentiate between the two percept types when viewing the asymmetric bistable stimuli, we operationally defined the initial percept upon stimulus presentation as the “default” interpretation, and the perceptual interpretation that followed as the “alternative” interpretation.
In our first analysis we hence examined relative frequencies of the first percept type at the onset of stimulus presentation to define the “default” and “non-default” (“alternative”) percept for all three stimuli. We tested the first percept bias (probability of first percept) against chance (p = 0.5) by means of a one-sample t-test. The definition of the “default” and “alternative” percept was used for all subsequent analyses. Importantly, all stimuli had been optimized to yield more or less balanced durations of both percept types during prolonged viewing. Behavioral data obtained on the illusory Gestalt stimulus were taken from our previously collected data set (Zaretskaya et al., 2013) and were re-analyzed for the current study to obtain first-percept biases.

Test for percept duration changes over time. We performed two control analyses to test for potential changes in percept durations across runs and within a run that could have biased the fMRI results. In the first control analysis, we performed a 2x4 repeated measures ANOVA with the factors “percept” and “run” to test for differences in median percept durations across runs. In the second control analysis, we split each run into two halves and tested for differences between percept durations in the first half compared to the second half by means of a 2x2 repeated measures ANOVA (factors “percept” and “half”). We treated the total duration of a run as the start of the first percept until the end of the last complete percept. The percept belonging to both halves was excluded from the analysis. The reported p-values from the ANOVA analyses were adjusted using a Greenhouse-Geisser correction.

Definition of regions of interest (ROIs)
According to the primary hypothesis of this study we paid particular attention to parietal ROIs, but also to other regions that could potentially exhibit differential modulation for default versus alternative percepts based on the previous literature.

Parietal cortex. First, we specifically wanted to test the involvement of previously identified posterior parietal areas in the organization of ambiguous displays. For the group ROI definition we used fMRI data from our previous study (Zaretskaya et al., 2013). We first identified peaks of activity from the
contrast “alternative > default” percept at a threshold of $p < 0.001$ (uncorrected),
which were located in the left and right anterior Intraparietal Sulcus. We then
defined our ROI as two spheres with 20 mm radius positioned to include the
most activated voxels (center: $x = \pm 26, y = -54, z = 56$) (Zaretskaya et al., 2013).

**Early visual cortex (V1-V3).** A large number of studies revealed early
visual cortex modulations as a function of perceptual interpretations (Murray et
al., 2002b; Fang et al., 2008; Grassi et al., 2017). These differential modulations
have frequently been interpreted as a signature of feedback signals carrying
information about a high-level interpretation. In this context, we decided to
examine whether activity related to the alternative vs. default interpretation
elicited similar responses in early visual cortex for our stimuli. We defined the
occipital visual fields V1, V2 and V3 in each subject individually using surface-
based probabilistic templates (Wang et al., 2015). For this, we generated cortical
surface models of each subject using the FreeSurfer software (Fischl et al., 1999)
and projected the maximum probabilistic maps of the topographic regions V1, V2
and V3 from the atlas space onto the individual surfaces. This procedure was
done using a virtual container written by Noah C. Benson
(https://hub.docker.com/r/nben/occipital_atlas/). Thereafter, we transformed
the surface-based ROIs into individual volume space and normalized them into
MNI space for beta estimate extraction. Examples of early visual cortex ROIs can
be seen in Figure 3C.

**Shape and motion regions (LOC, V5/MT, MST, V3A, V6, CSv, Pc).** In
addition to the posterior parietal and early visual cortex ROIs, we also
investigated the involvement of ventral shape-selective lateral occipital area
(LOC) and of dorsal motion selective areas V5/MT, MST, V3A, V6, and two
recently described motion responsive hubs in the cingulate sulcus (CSv) and
precuneus (Pc). Most previous fMRI studies using the Diamond display focused
their analysis on LOC and excluded dorsal motion-selective and parietal ROIs
from their analyses (De-Wit et al., 2012; Fang et al., 2008; Murray et al., 2002b),
with the exception of Caclin et al. (2012) who also included the V5+/MT+
complex. On the other hand, fMRI studies using bistable plaid stimuli conducted
ROI analyses only on the early visual cortex, dorsal V5+/MT+ and V3A (Castelo-
Branco et al., 2002; Villeneuve et al., 2012), thus not including ventral shape-
responsive areas or other dorsal motion selective areas.

We defined dorsal motion selective ROIs using standard methods (Huk et
al., 2002; Smith et al., 2006; Fischer et al., 2012a) in 16 of our 17 subjects (in one
subject technical problems prevented data acquisition of the motion localizer).
Our motion localizer consisted of seven conditions presented in a randomized
block design, as described before (Fischer et al., 2012a). Attentional effects were
balanced by applying the same letter back-matching task in all conditions (Huk
et al., 2001). Using contrasts detailed below we defined a large set of specialized

The human V5+/MT+ complex was defined as the contralateral response
using the contrast of coherent hemifield motion versus static dots. MST+ was
defined as the respective ipsilateral response (Dumoulin, 2000; Huk et al., 2002).
Thereafter, we excluded the correspondent MST+ voxels from the V5+/MT+
complex to define V5/MT. We use the term MST+ as several additional motion
responsive satellite regions of monkey V5/MT have receptive fields extending
into the ipsilateral hemifield like those of MST that are most likely included in
our MST+ (Nelissen, 2006; Kolster et al., 2010).

Area V3A was defined as the region below the parietal-occipital sulcus
and extending into the transverse occipital sulcus that responded significantly
stronger to coherent planar motion with moving fixation disk compared to the
moving fixation disk on a static background. This contrast reliably leads to
selective activation of voxels overlapping with retinotopically defined V3A
(Fischer et al., 2012a).

Areas V6 (Pitzalis et al., 2006; Fischer et al., 2012a), CSV (Wall and Smith,
2008; Fischer et al., 2012b) and Pc (Cardin and Smith, 2010, 2011) were defined
using the contrast comparing coherent optic flow to random motion exploiting
their established preferences to coherent motion. Note that the region defined
here as V6 most likely also includes voxels of the neighboring V6Av as this region
also marginally prefers coherent compared to random dot motion (Pitzalis et al.,
2013, 2015).

To define ventral shape-selective areas LOC, we used the well-established
procedure of contrasting images of objects with their fourier-scrambles (Grill-
Spector et al., 2001) in 17 subjects. Both blocks of the localizer were repeated 15
times and consisted of 6 images, each presented for 2 s. Subjects were asked to
perform an image back-matching task to ensure constant attentional loads
across conditions.

Please note that we did not have independently drawn ROIs for the subset
of subjects observing the illusory Gestalt stimulus. However, responses of
independently defined motion selective areas (V5/MT, MST, V3A, V6) to the
illusory Gestalt stimulus have been analyzed elsewhere (Grassi et al. 2016)
following a similar procedure and revealed no percept-driven modulation
favoring the alternative percept, regardless of stimulus size. Also, whole-brain
analyses of responses while viewing the illusory Gestalt stimulus revealed no
involvement of ventral shape-selective areas.

**MRI data analysis**

**Preprocessing.** All data were processed using SPM8 (Wellcome
Department of Imaging Neuroscience, London, UK). Preprocessing included slice-
time correction (reference slice: TR/2), motion correction and spatial
normalization using segmentation to the MNI space. For the (single-subject) ROI
analyses we smoothed the EPI volumes with a Gaussian Kernel of 3 mm full-
width at half maximum. For group-level whole-brain analyses we spatially
smoothed the images with a Gaussian Kernel of 9 mm.

**First level GLM analysis.** Data of each subject were analyzed separately
for each stimulus using a standard GLM approach. Reports of default and
alternative percept onsets were used to build two regressors of interest using
stick functions. We also included two further regressors modeling the onsets of
the fixation-only periods and of the stimulus presentation, the latter to account
for possible effects related to the appearance of the stimuli on the screen. We
modeled the onsets using the standard “double-gamma” hemodynamic response
model (with T0 = 0). Moreover, to account for variability in the hemodynamic
response function (HRF) and in the timing of key presses we also included the
first temporal derivative of each regressor in the model. Finally, we included six
nuisance regressors of the realignment parameters to model variance related to
head motion and an orthogonalized global mean regressor. Low-frequency signal
drifts were removed using a high-pass filter with 128 s cutoff.

**Group level analysis and derivative boost to account for temporal**
response variability. We performed two types of analyses at the group level: a
region of interest analysis and a whole-brain analysis. In order to include the
variance accounted for by all basis functions (i.e. standard HRF and the first
temporal derivative) in these second-level analyses, we used a combination of
the magnitude calculated across both beta estimates called "derivative boost"
(Calhoun et al., 2004; Steffener et al., 2010; Pernet, 2014). The "derivative boost"
is calculated as:

$$H = \sqrt{\beta_1^2 \sum_{t=1}^{N} x_1^2 + \beta_2^2 \sum_{t=1}^{N} x_2^2} \times \frac{\beta_1}{|\beta_1|}$$

with the following parameters: H derivative boost, $\beta_1$ parameter estimate for the
canonical HRF, $x_1$ regressor convolved with the HRF, $\beta_2$ parameter estimate for
its first derivative, $x_2$ regressor convolved with the derivative. We constrained
the derivative boost by a time window of ±1 s relative to the canonical HRF as
suggested by Steffener et al., 2010. The code used for boosting the images is
available at the repository:

https://github.com/CPerNet/spmup/blob/master/spmup_hrf_boost.m.

For the ROI analyses, we merged the independently defined ROIs from
both hemispheres into one ROI whenever possible. We then extracted the
boosted beta estimates of each condition (i.e. perceptual onset) within each ROI
for all subjects. Estimates were averaged over runs and voxels for each ROI. The
mean beta estimates were then transformed to percent signal change using the
respective ROI mean signal as normalization reference (cf. Pernet, 2014).

Paired t-tests were conducted across subjects to determine differences
between alternative and default percepts for each ROI. For the hypothesis-driven
ROI analysis testing the hypothesis that the parietal cortex is activated by
motion-based segmentation and that the EVC is suppressed by modulatory
feedback signals during alternative interpretations we tested template ROIs
using one-sided paired t-tests. For the exploratory ROI analysis of individually
defined mid-level ventral and dorsal regions two-sided paired t-tests were used.
Statistics were Holm-Bonferroni corrected for multiple comparisons for the
number of tests conducted.
Whole-brain random effects analyses were used in order to test the possible involvement of additional regions in perceptual organization. To identify voxels favoring alternative perceptual states, we performed a voxel-wise comparison between “alternative > default” perceptual states in every subject. Based on the boosted contrasts we performed random effects t-tests across all participants for each display type separately. To formally test for activations common to all stimuli we performed a conjunction analysis by plotting overlapping voxels from the contrast “alternative > default” from all stimuli (thresholded at t = 3).
Results

Behavioral results

Relative percept frequencies at stimulus onset

Figure 2A shows relative frequencies of each percept for each of the three stimulus classes. For the Plaid stimulus, pattern motion (i.e. a single coherent pattern moving in a single direction) was perceived first upon stimulus presentation with a probability of 93.4% of the trials ($t_{16} = 13.4, p = 4.0772 \times 10^{-10}$, Cohen’s $d = 3.25$). This is consistent with previous observations (Hupé and Rubin, 2003; Hupé and Pressnitzer, 2012). For the Diamond stimulus, in accord with our expectations, subjects perceived the four unbound elements upon stimulus presentation in 83.8% of the trials ($t_{16} = 5.46, p = 5.2191 \times 10^{-5}$, Cohen’s $d = 1.32$), which later alternated with the grouped percept of a diamond shape moving behind three occluders. For the illusory Gestalt display, 82.8% of the trials began with the percept of unrelated dot-pairs ($t_{17} = 5.21, p = 7.2667 \times 10^{-5}$, Cohen’s $d = 1.23$), which then alternated with the grouped Gestalt percept. We define the respective first percepts as “default” percepts (Plaid: pattern motion; Diamond: unbound elements; Gestalt: local motion). After the initial default interpretation, perceptual alternations with the non-default, alternative interpretation (Plaid: transparent component motion; Diamond: bound shape; illusory Gestalt: global Gestalt motion) ensued.

These behavioral results suggest that across all stimuli, alternative percepts involved more complex scene segmentation processes compared to the initial default percepts: for the Plaid stimulus, perceiving two depth-segmented transparent objects sliding over each other in different directions involves higher-level processes (depth-organization and transparent motion) compared to perception of a simple pattern moving in one direction. Similarly, the alternative percept of a partially occluded diamond moving behind two invisible columns involves higher-level processes, as it implies understanding of visual 3D arrangements partly covering a moving object, in contrast to the perception of separately moving lines. The same argument holds for the Gestalt stimulus where the alternative Gestalt percept requires long-range spatial grouping.
mechanisms that are not required for perception of the local inducers in the
default percept.

Relative percept frequencies during prolonged viewing

Even though we had optimized all of the stimuli to yield balanced durations of
both percept types, the average median durations shown in Figure 2B reveal
some imbalances, some in favor of the default, others in favor of the non-default
(alternative) perceptual interpretation. The Diamond display yielded longer
percept durations for the default interpretation (7.62 ± 3.02 s mean ± S.D.) than
for the alternative one (5.47 ± 2.01 s, t16 = 5.28, p = 7.4994e-05, Cohen’s d = 1.28),
the Plaid display yielded no statistically significant difference between percept
types (default: 5.354 ± 2.26 s, alternative: 6.031 ± 1.71 s, t16 = 1.058, p = 0.306,
Cohen’s d = 0.26), and the illusory Gestalt display led to longer alternative
percept durations (7.02 ± 3.69 s) compared to default percept durations (5.68 ±
2.25 s, t17 = 2.46, p = 0.025, Cohen’s d = 0.58) (see Figure 2B). Importantly, the
balances during prolonged viewing were not related to the strong predominance
of the default percept upon stimulus presentation.

Note that the divergence of average dominance durations across stimuli
provided the advantage that any convergent fMRI results across stimuli would
hence not be attributable to systematic behavioral bias.

Finally, the distributions of dominance durations were well fitted with
gamma functions (maximum likelihood fit coefficient of determination $r^2$, all >
0.9, not shown).

Percept duration changes over time

We found a significant increase of the median percept durations across runs in
both displays (2x4 (percept x run) repeated measure ANOVA; main effect “run”,
Plaid: F(3,48) = 4.6958, p = 0.01755, Diamond: F(3,48) = 4.3013, p = 0.01289).
However, and relevant to our fMRI results, the relation between percepts
remained constant across time (Interaction: “percept” x “run”, Plaid: F(3,48) =
0.091, p = 0.961, Diamond: F(3,48) = 2.34, p = 0.102). A second analysis testing for
differences between percept durations in the first half of the runs compared to
those in the second half of the runs (2x2 (percept x half) repeated measures
ANOVA) revealed no significant results neither for the main effect nor for the
interaction in both displays (all $F < 0.2$ and all $p > 0.5$). Together, these results revealed no differences in relative percept durations over time that could have influenced the fMRI results presented below.

**Post-scan questionnaire**

The post-scan questionnaire revealed that subjects were confident in their responses (Q1; Likert range: 1 = not confident, 7 = very confident; Diamond: 5.94 ± 1.19 [mean ± std], median = 6; Plaid: 5.7 ± 1.1 [mean ± std], median = 6), with no difference against the expected value of 6 (Wilcoxon signed rank test, Diamond: $w = 38.5$, $p = 1$; Plaid: $w = 25$, $p = 0.2553$). For the Plaid display, most subjects reported they did not try to enforce any particular percept (Q2; Likert range: 1 = alternative percept, 7 = default percept; 4 ± 0.63 [mean ± std], median = 4, Wilcoxon signed rank test against median of 4: $w = 3$, $p = 1$) and no particular percept was more difficult to see than the other one (Q3; 3.65 ± 1.27 [mean ± std], median = 4, $w = 17.5$, $p = 0.3652$). For the Diamond display, six subjects reported a tendency to have enforced their percepts towards the alternative interpretation during viewing of the Diamond display (Q2; 3.56 ± 0.63 [mean ± std], median = 4, Wilcoxon signed rank test against median of 4: $w = 0$, $p = 0.03125$), in line with the fact that subjects reported the alternative percept of the Diamond display to be more difficult to see (Q3; 3.24 ± 1.25 [mean ± std], median = 3, $w = 12.5$, $p = 0.038$). All responses can be seen in Table 1.

To test whether mental effort as evidenced above could have affected any of our fMRI results we report additional whole-brain control analyses (shown below) showing that the above behavioral factors did not account for the main results (see the whole-brain responses section).

**Region of interest analysis**

In the following, we describe the results of our ROI analyses for the ROIs of posterior parietal cortex (PPC) and early visual cortex (EVC) and for the individually defined ROIs of dorsal motion selective and ventral shape selective areas. Statistical results of all ROI analyses are provided in Table 2.
Posterior parietal cortex and early visual cortex responses

One of the central questions of the current study was whether the posterior parietal cortex is generally involved in computing the alternative stimulus interpretations across distinct bistable motion stimulus classes. To test the generic role of the PPC in perceptual organization, we defined two parietal ROIs in the left and right parietal cortex using previously reported coordinates (Zaretskaya et al., 2013). Moreover, in view of previous studies showing EVC modulations as a function of perceptual states (Murray et al., 2002b; Fang et al., 2008; Zaretskaya et al., 2013; Grassi et al., 2017), we examined percept-specific modulations in three early visual ROIs (V1, V2 and V3) defined using individually inflated brain surfaces and a probabilistic atlas based on functional retinotopy of a large subject population (Wang et al., 2015).

This ROI analysis revealed a highly consistent response pattern for the tested bistable displays. The parietal ROI showed an enhancement of activity for the non-default interpretation compared to the default interpretation in all tested displays (Figure 3A). In contrast, early visual ROIs V1 and V2 were consistently suppressed in activity, irrespective of the stimulus used (Figure 3B). We found no percept-driven modulation in visual area V3.

Both results correspond to our previously published findings using the illusory Gestalt stimulus (Zaretskaya et al., 2013; Grassi et al., 2016) that demonstrated a specific involvement of parietal cortex during perception of the non-default Gestalt percept independent of stimulus dimensions, while EVC was suppressed.

Hence, the current findings confirm our hypothesis by demonstrating that across three entirely distinct bistable stimulus classes, parietal cortex consistently responded more strongly to the alternative segmented percept interpretation, while early visual cortex was suppressed.

Responses of motion-selective dorsal regions and shape-selective ventral regions

Beyond examining our main hypothesis, we also wanted to quantify percept-driven responses of mid-level, functionally specialized dorsal motion- and ventral shape-responsive regions. We therefore defined individual ROIs for every subject using separate functional localizers. The results revealed
divergent, i.e. stimulus-specific, differential modulations across the stimulus
classes in both visual processing streams.
Shape-selective LOC in the ventral occipital cortex favored the alternative
interpretation during viewing of the Diamond (i.e. shape perception), but had no
significant preference for either percept of the Plaid stimulus (Figure 4), and
intriguingly, it was not involved in the perception of the (non-default) global
Gestalt illusion (see Zaretskaya et al., 2013; Grassi et al., 2016, 2017). Hence, for
shape-selective LOC there was no systematic modulation favoring either default
or alternative perceptual interpretations across the stimulus classes.
Also dorsal motion-selective ROIs showed no systematic modulation
favoring either default or alternative perceptual interpretations across all
stimulus classes. During viewing of the Plaid display, V5/MT, MST and V3A
showed an increase of activity with the alternative percept (i.e. transparent
motion)(Figure 4A). During viewing of the Diamond display, only area MST
showed a differential modulation, favoring the alternative interpretation (i.e.
Diamond percept). In contrast, a ROI analysis of areas MST, V3A and V6 showed
no differential modulation as a function of percept during viewing of the illusory
Gestalt display. Only area V5/MT showed a small selective deactivation during
the global Gestalt perception (Grassi et al., 2016).
The stimulus-specific activation patterns in shape-selective ventral and
motion-selective dorsal regions indicate that the involvement of these
extrastriate areas in resolving perceptual ambiguity depends on the exact
stimulus features, in contrast to the fully consistent modulation pattern found in
the posterior parietal and early visual cortex across all stimulus classes.

**Whole-brain responses**
In order to investigate whether regions other than those examined in the ROI
analyses were involved in perceptual inference and scene segmentation, we
performed a whole-brain random effects analysis. We tested the main contrast
“alternative percept > default percept” and its inverse for each stimulus type.
This analysis revealed similar activation maps for all displays (see Figure 5A). To
allow for an easy comparison with results for the corresponding contrast of the
Gestalt display, we additionally present the whole-brain responses to the illusory
Gestalt stimulus from Zaretskaya et al. (2013) (Figure 5A, lower panel), detailed results for which can be found in Zaretskaya et al. (2013) as well as in a replication study with different stimulus sizes in Grassi et al. (2016).

In line with the ROI analysis results we found a consistent activation of the posterior parietal cortex, together with a consistent deactivation of EVC during the perception of the alternative interpretation across all presented stimuli (see conjunction analysis in Figure 6). Peak MNI-values (xyz) of PPC activity for Plaid were (40 -36 52) and (32 -52 54) for the right hemisphere and (-22 -64 56) and (-28 -44 52) for the left hemisphere; for Diamond they were (28 -52 54) for the right and (-34 -40 48) for the left hemisphere. We also observed differential responses in areas along both, ventral and dorsal visual pathways, but these were not consistent across the different types of stimuli.

Control analyses

Although participants were instructed not to enforce any particular percept, the bistable paradigm does not fully exclude the possibility of attentional confounds due to differences in mental effort between percepts. We hence performed an additional whole-brain analysis using only subjects who reported after the fMRI session not to have enforced any particular percept (see Table 1; Diamond: n = 10, Plaid: n = 13). The results replicated those of the original analysis, for both displays (see Figure 5B).

Second, we examined whole-brain activity of the two participants who reported more difficulty in perceiving the default percept for the Diamond display, plus an additional two participants who reported the same for the Plaid display (see Table 1). Figure 5C shows that in all four participants PPC activity was significantly higher for the alternative versus the default percept, hence ruling out that difficulty drove the PPC results.

Finally, note that a strong argument against attention-related factors driving PPC during the alternative percept is the concurrent relative suppression of V1 and V2, which is not compatible with attentional enhancement.
**Discussion (1499)**

In this study, we used three asymmetric bistable stimuli to test the hypothesis that particular brain regions are consistently involved in motion-based scene segmentation, regardless of particularities of the percept or physical stimulus properties. While the stimuli differed vastly in their appearance, each of them had one default perceptual interpretation and one alternative interpretation that involved different scene segmentation processes. The stimuli remained physically constant while percepts alternated, and were hence well suited to identify neural substrates underlying perceptual organization. Our results showed that the PPC was consistently activated during perception of the alternative percept across distinct stimulus classes, no matter whether this percept included a Gestalt or whether it increased or decreased the number of perceived items. This PPC activation was systematically accompanied with a suppression of early visual areas V1 and V2 during non-default percepts across all stimuli used.

**PPC activation**

The enhancement of activity in the PPC during non-default interpretations is in line with an increasing number of studies that reported the involvement of the PPC in high-level visual tasks like perceptual grouping (Grassi et al., 2016; Liu et al., 2017; Reichert et al., 2014; Yokoi and Komatsu, 2009; Zaretskaya et al., 2013), motion segmentation (Duarte et al., 2017), object processing (Konen and Kastner, 2008) and 3D form extraction from motion (Orban, 2011). The common denominator in all these vision tasks appears to be scene segmentation: what is fore- and background, and which visual components belong to the same entity?

The role of the PPC in scene segmentation is likely not constrained to dynamic stimuli. Segmentation tasks based on static displays such as the alternative interpretation of the Schroeders Staircase (Karten et al., 2013), 3D perception of the Necker cube (Inui et al., 2000) and figure completion of Kanizsa displays (Murray et al., 2002a, 2004; Stanley and Rubin, 2003) have all been shown to involve the PPC. Given that in all of the stimuli used here (and in most used in prior literature) the alternative interpretation involved scene
segmentation in depth as well as grouping of elements, this could be common
function(s) of the PPC segmentation process.

Interestingly, beyond scene segmentation the PPC is also known to play a
crucial role in spatial attention and visual selection (Kastner and Ungerleider,
2000; Corbetta and Shulman, 2002). The anatomical convergence of these
processes in the PPC points to a functional relationship, including evidence that
objects are the units of attentional selection (Qiu and Sugihara, 2007; Bartels,
2009; Fang et al., 2009; Yokoi and Komatsu, 2009; McMains and Kastner, 2011;
Poort et al., 2012). However, the fact that perceptual integration involves
parietal cortex even when it is task-irrelevant suggests a degree of independence
between perceptual integration and attentional processing (Liu et al., 2017), and
the relative suppression of early visual cortex during PPC activation during scene
segmentation suggest distinct effects of segmentation and attention on visual
cortex. Similarly, our prior evidence that interference with parietal activity
selectively reduced durations of the non-default percept in the Gestalt display
suggests a causal role of parietal cortex in non-default percept formation rather
than its post-hoc activation through the percept change (Zaretskaya et al., 2013).
Future studies are needed to investigate whether the same applies for the
Diamond and Plaid displays, whether a single or multiple segmentation
processes reside in PPC, and how they interact with other processes such as
(voluntary) attention and perceptual selection.

Feedback to EVC

In addition to the role of PPC in motion-based scene segmentation, our results
show a consistent down-modulation of V1 and V2 whenever the alternative
interpretation was perceived. This is consistent with prior imaging evidence of
EVC suppression during viewing of object or Gestalt content (Murray et al.,
2002b; Fang et al., 2008; Zaretskaya et al., 2013; Reichert et al., 2014; Grassi et
al., 2016). Our results extend these findings to scene segmentation also in the
absence of Gestalt.

The early modulation observed here is likely related to
electrophysiological evidence of neurons in early visual areas signaling border-
ownership (Zhou et al., 2000), illusory contours (Peterhans and von der Heydt,
and filling in of foreground surfaces (Roelfsema et al., 2007). These processes have also been shown to undergo top-down attentional modulation (Fang et al., 2009; Poort et al., 2012), with delayed signal modulation in upper and lower cortical layers known to receive feedback from higher-level areas (Self et al., 2013). In case of Gestalt displays, perceptual states differentially modulate V1 and V2 in ways that reflect the topographic layout of the alternative percept (Grassi et al., 2017; see also Kok and de Lange, 2014). Recent laminar signal evidence supports the view that feedback mediates this modulation (Kok et al., 2016). Since the strongest components of these percept-driven EVC modulations were consistently negative for the alternative percept, negative EVC signal would dominate smoothed multi-subject analyses.

The consistent involvement of the PPC during perceptual organization makes it a likely contributor to the EVC modulations. This does not rule out additional, stimulus specific contributors, such as V5+/MT+ and LOC, as has been suggested previously (Murray et al., 2002b; Guo et al., 2004; Schmidt et al., 2011).

The percept-driven EVC modulation is compatible with the theory of predictive coding (Rao and Ballard, 1999; Friston, 2005), and the present data extend experimental evidence to non-Gestalt scene segmentation processes.

**Numerosity**

Besides its involvement in attention and perceptual organization, the PPC has also previously been associated with the representation of numerosity (Eger et al., 2003; Harvey et al., 2013). However, the present data show that the consistent up-modulation of PPC with the non-default interpretation cannot be related to numerosity processing or perception: in two of our stimuli, the alternative percept corresponded to a reduction of perceived items (our grouping-related stimuli, i.e. illusory Gestalt display and Diamond display), but for the Plaid stimulus, the alternative percept corresponded to an increase in perceived items.

**Plaid**

One of the surprising findings in this study concerns the plaid stimuli. Traditionally, perception of pattern motion has been seen as the perceptual correlate of advanced neural integration: electrophysiological studies using
unambiguous sine-wave gratings revealed that neurons in V1 respond only to
cOMPONENT motion, whereas pattern motion is processed first in the V5/MT
complex (Movshon et al., 1985; Khawaja et al., 2013).

However, there are good reasons to view component motion as a result of
late visual segmentation. First, component motion is the more complex
perceptual interpretation: rather than a single surface moving in one direction,
component motion suggests transparency, three-dimensional arrangement, and
two objects gliding over each other independently in two distinct directions.

Second, behaviorally, pattern motion is perceived first, with component
motion perception developing later, as observed in the present data as well as
previously (Hupé and Rubin, 2003; Hupé and Pressnitzer, 2012).

Third, non-motion surface segmentation cues, such as depth and
occlusion, can modulate the perception of motion in the Plaid display, suggesting
that higher-level integration processes including multiple features are involved
(Kersten et al., 1992; Trueswell and Hayhoe, 1993; Stoner and Albright, 1996,
1998).

Finally, the V1 versus V5+/MT+ segregation does no longer apply to
rectangular-wave gratings similar to those used in our study, for which V5/MT
neurons can change their tuning properties based on transparency cues (Stoner
and Albright, 1992).

Our results are consistent with previous fMRI studies that reported a
relative deactivation of V5+/MT+ during pattern motion perception (Castelo-
Branco et al., 2002; Villeneuve et al., 2012) and a recent study in which pattern
motion direction was decoded in V1 (Van Kemenade et al., 2014). The increase of
activity in the PPC and motion-selective areas during the transparent motion
interpretation suggest their involvement in the computationally more
demanding segmentation of the two surfaces rather than in the integration of the
two components into a coherent pattern. This interpretation is consistent with
the fMRI results from a recent study that used a similar asymmetric bistable
stimulus revealing V5+/MT+ and parietal involvement during the alternative
perception of two segmented surfaces moving inwards compared to the default
perception of one surface moving downwards (Duarte et al., 2017).
Diamond

Most previous fMRI studies that used the Diamond display to study grouping of individual elements into coherent shapes focused their analysis on EVC and on LOC (Murray et al., 2002b; Fang et al., 2008; De-Wit et al., 2012). The majority of these studies consistently reported LOC activation together with V1 deactivation during shape compared to element perception and interpreted the differential modulation in context of predictive coding. One further fMRI study using different variations of the diamond display (Caclin et al., 2012) failed to report V1 deactivation, which may be due to differences in the stimulation parameters (see discussion in De-Wit et al., 2012 and Caclin et al., 2012). Our findings extend prior findings to show involvement of PPC during shape perception.

Conclusion

In sum, we found a consistent PPC activation and EVC deactivation specifically during perception of the alternative interpretations, which consistently involved 3D scene segmentation and grouping, across a range of asymmetric bistable stimuli. Our findings suggest a generic mechanism for late motion-based scene segmentation in the PPC.
References


Qiu FT, Sugihara T (2007) Figure-ground mechanisms provide structure for selective attention. Nat. Neurosci. 10, 1492–1499.


993  selective responses in MT, MST and the pulvinar of humans. Eur J Neurosci
994  36:2849–2858.
999  Wolpert I (1924) Die Simultanagnosie - Störung der Gesamtauffassung. Z
1000  Gesamte Neurol Psychiatr 93:397–415.
1001  Yokoi I, Komatsu H (2009) Relationship between neural responses and visual
1003  Zaretskaya N, Anstis S, Bartels A (2013) Parietal cortex mediates conscious
1006  grouping of stimuli based on colour and motion. Neuroimage 73:156–166.
1007  Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in
**Figure legends**

**Figure 1.** Illustration of the asymmetric bistable stimuli used. (A) Plaid stimulus. This stimulus could be perceived either as a single plaid moving upwards (default percept) or as two depth-separated transparent gratings sliding over each other in opposite directions (alternative percept). (B) "Occluded Diamond" stimulus and its two interpretations: local vertical motion of separate elements (default percept) and global horizontal motion of a bound shape (alternative percept). (C) Illusory Gestalt display used previously (Zaretskaya et al., 2013): four pairs of dots moving circularly in phase could be perceived either as local motion of separate pairs of dots (default percept) or as of two illusory squares sliding over each other in planar motion (alternative percept).

**Figure 2.** First percept bias and mean percept durations during continuous viewing. (A) Bar plots show the probability for a given percept to be perceived first upon stimulus presentation, for all three bistable stimuli (mean ± SEM). For all stimuli one of the two percepts had an overwhelming predominance of being initially perceived. We defined the first percept as the default percept. The right bar plot shows the first percept bias for the Gestalt illusion, which had not been analyzed in Zaretskaya et al., (2013). We statistically tested the first percept bias against chance (p = 0.5) by means of a one-sample t-test. (B) The bar graphs depict the average perceptual durations for all percepts during continuous viewing (mean ± SEM). Durations for the illusory Gestalt display have been published in Zaretskaya et al., (2013) and are included for completeness. ****: p<0.001; ***: p<0.01; *: p<0.05, uncorrected.

**Figure 3.** ROI analysis results. Responses of the posterior parietal cortex ROI and early visual cortex ROIs for the Plaid (A) and Diamond (B) displays. The PPC showed differential modulation favoring the alternative percept in both stimuli. Early visual areas V1 and V2 were consistently deactivated during perception of the alternative percept in both stimuli. Area V3 revealed no percept-driven modulation. The baseline corresponds to the mean fixation-only response. (C) Group parietal ROI together with three individual examples of the early visual ROIs in MNI space. ****: p<0.001; ***: p<0.01; *: p<0.05, Holms-Bonferroni corrected; +: p<0.05, uncorrected.

**Figure 4.** ROI analysis of the ventral shape selective and dorsal motion selective areas for the Plaid (A), Diamond (B) displays. Responses of dorsal and ventral areas to perceptual changes during viewing of the Gestalt illusion display can be found in Grassi et al. 2016 and Grassi et al. 2017. None of the regions was consistently up- or down-modulated for a given percept type across all three bistable stimulus classes. ****: p<0.001; **: p<0.01; *: p<0.05, Holms-Bonferroni corrected; +: p<0.05, uncorrected.

**Figure 5.** Whole-brain results showing the main contrast "alternative > default" for all stimuli displays. (A) Whole-brain results for the Plaid, Diamond and illusory Gestalt displays. Illusory Gestalt results are reproduced here for comparison from Zaretskaya et al. 2013. (B) First control analysis: whole-brain results for the Plaid and Diamond displays testing subjects who reported not to have guided their perception to any particular interpretation of the stimuli (Diamond: n = 10, Plaid: n = 13). (C) Second control analysis: axial section through PPC for the two participants who reported more difficulty in perceiving the default percept for the Diamond display, plus an additional two participants who reported the same for the Plaid display (see Table 1). PPC activity was hence driven by percept type, not by difficulty. Results of A and B are projected onto average inflated cortical hemispheres from the FreeSurfer software (Fischl et al., 1999). For better visualization we thresholded the results at the liberal value of t = 2. LH: left hemisphere; RH: right hemisphere.

**Figure 6.** Conjunction analysis. The figure shows overlapping voxels in a glass brain (A) and cross-sections (B and C) from the contrast "alternative > default" that reached a threshold of t > 3 and t < -3 (p < 0.005) in each of the three display types. Across all three stimuli (Plaid, Diamond and illusory Gestalt) we found consistent PPC activation (B) and EVC deactivation (C) during perception of the non-default interpretation. LH: left hemisphere; RH: right hemisphere.
Table 1. Post-scan questionnaire answers

<table>
<thead>
<tr>
<th>Question</th>
<th>Likert-Scale</th>
<th>Plaid</th>
<th>Diamond</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: Confidence</td>
<td>0 0 0 1 1</td>
<td>3 5 7</td>
<td>0</td>
</tr>
<tr>
<td>Q2: Enforce</td>
<td>0 0 2 0 1</td>
<td>10 0 0 0</td>
<td>1</td>
</tr>
<tr>
<td>Q3: Difficult</td>
<td>0 4 3 7 1</td>
<td>2 0 0 0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Results of the paired t-tests for all ROIs for both stimuli.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Plaid</th>
<th>Diamond</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V1</td>
<td>$t_{16} = -5.39, p = 0.0001, d = 1.31$, ***</td>
<td>$t_{16} = -5.02, p = 0.0001, d = 1.22$, ***</td>
</tr>
<tr>
<td>V2</td>
<td>$t_{16} = -4.65, p = 0.0003, d = 1.13$, ***</td>
<td>$t_{16} = -3.63, p = 0.0023, d = 0.88$, **</td>
</tr>
<tr>
<td>V3</td>
<td>$t_{16} = -0.7152, p = 0.4848, d = 0.17$</td>
<td>$t_{16} = 0.15, p = 0.8847, d = 0.04$,</td>
</tr>
<tr>
<td>V5/MT</td>
<td>$t_{15} = 6.34, p = 1.323e-05, d = 1.59$, ***</td>
<td>$t_{15} = 1.63, p = 0.1242, d = 0.41$,</td>
</tr>
<tr>
<td>MST</td>
<td>$t_{15} = 6.19, p = 1.703e-05, d = 1.55$, ***</td>
<td>$t_{15} = 3.94, p = 0.0013, d = 0.98$, **</td>
</tr>
<tr>
<td>V3A</td>
<td>$t_{14} = 3.37, p = 0.0046, d = 0.87, *$</td>
<td>$t_{14} = 2.25, p = 0.0413, d = 0.58, +$</td>
</tr>
<tr>
<td>V6</td>
<td>$t_{15} = 0.71, p = 0.4857, d = 0.18$</td>
<td>$t_{15} = 0.75, p = 0.4635, d = 0.19$</td>
</tr>
<tr>
<td>CSv</td>
<td>$t_{12} = 0.48, p = 0.6382, d = 0.13$</td>
<td>$t_{12} = 1.28, p = 0.2236, d = 0.36$</td>
</tr>
<tr>
<td>Pc</td>
<td>$t_{14} = 1.30, p = 0.2142, d = 0.34$</td>
<td>$t_{14} = 1.12, p = 0.2831, d = 0.29$</td>
</tr>
<tr>
<td>LOC</td>
<td>$t_{16} = 0.69, p = 0.4990, d = 0.17$</td>
<td>$t_{16} = 6.02, p = 1.768e-05, d = 1.46$, ***</td>
</tr>
</tbody>
</table>

$d$: effect size (Cohen’s $d$), ***: $p<0.001$; **: $p<0.01$; *: $p<0.05$, Holm-Bonferroni corrected; +: $p<0.05$, uncorrected.
A

**Plaid**

% signal change

---

**% signal change**

PPC  V1  V2  V3

---

B

**Diamond**

% signal change

---

% signal change

PPC  V1  V2  V3

---

C

**Parietal and early visual ROIs**

S5  S8  S10

---

z = 58  x = -10  x = -10  x = 16

---

Yellow PPC  Red V1  Green V2  Blue V3
Conjunction analysis: ◯ + ○ + □

A
PPC

B
PPC
LH
RH
z = 54
y = -46

C
EVC
x = 3
z = 10

Figure legend:
- Red: t > 3
- Blue: t < -3