A. Eye movements and perceptual bistability

Involuntary eye movements could modulate neural activity and/or perceptual dynamics. If residual eye-movements around the fixation point were not evenly distributed during bound and unbound percepts, they may have altered neural activities differentially during the two percepts. Involuntary smooth pursuit of the diamond during bound percepts would overall have reduced the retinal slip and consequently the response of motion selective cells in V1 and hMT+ (Ilg and Thier, 1996). It is noteworthy that such a decreased activity could account for some of the middle temporal deactivation seen during bound percepts, but is at odds with the increased activity observed in V1 during bound percepts.

An alternative explanation would, on the contrary, consider that accurate fixation during a bound percept may be more demanding than fixation during an unbound percept, because observers would resist an automatic trend for ocular tracking. Middle temporal deactivation during the bound percept could thus reflect the active suppression of involuntary pursuit initiation.

Finally, blinks, saccades, and variation of pupil size may both induce perceptual transitions and elicit cortical activity in visual areas that would thus appear correlated to a switch toward a new percept.

We suspect that if pursuit, active suppression of pursuit, or any type of eye movement was more prominent during one percept compared to the other, a number of cerebral structures like Frontal Eye Field (FEF) and Supplementary Eye Field (SEF) should have also been significantly active (Berman et al., 1999), which was not the case. However, to ensure that our experimental conditions did not entail significant differences of ocular behavior between the bound and unbound percepts, we characterized eye movements and perceptual dynamics during bistable vision of the different motion displays used in the scanner, in a separate off-scan experiment, under fixation requirements, as during the fMRI experiment.

In addition, to control for possible perceptual effects of unnoticeable eye movements, we tested whether perceptual bistability is influenced by actively pursuing a small central disk undergoing a circular trajectory of the same amplitude and phase as the diamond displays.

Participants and procedure

Ten subjects (aged 21-35, 6 women) participated in the off-scan experiment. We recorded eye movements while subjects observed the same six displays as in the scanner and continuously reported their percepts. Eight of these subjects also participated in the fMRI experiment.

Depending on the experimental runs, subjects were either instructed to fixate on a centrally presented dot (Fixation, as in the fMRI experiment), or to pursue with the eyes a dot moving as if attached to the center of the rotating square (Pursuit). This led to a total of 12 blocks per subject (6 displays × 2 eye-movements). As in the scanner, viewing was binocular. Button presses (85-Hz sampling rate) and eye position (ISCAN infra-red eye-tracker, 240-Hz sampling rate, monocular recording) were continuously recorded.
Behavior

Perceptive reports were analyzed as for the fMRI experiment, with the addition of the Eye-movement within-subject factor (2 levels: Fixation and Pursuit).

The number of transitions was analyzed with a Stimulus × Induction × Eye-movement ANOVA. There was a significant effect of Induction (F(1,9)=5.173; p=0.05) and of Eye-movement (F(1,9)=30.582; p=0.0003), along with an Stimulus × Induction interaction (F(2,18)=5.383; ε=0.742; p=0.03), and marginally significant interactions: Induction × Eye-movement (F(1,9)=4.542; p=0.06) and Stimulus × Induction × Eye-movement (F(2,18)=3.305; ε=0.941; p=0.06). There were always fewer perceptual switches during pursuit eye-movements than during fixation. The difference was on average more pronounced for the spontaneous transitions than for the evoked transitions, which corresponds to what is observed for the Contrast and Motion stimuli. For the Shape stimuli however, the difference between fixation and pursuit was more pronounced in the evoked condition. Post-hoc Tukey HSD tests revealed a significant difference between Fixation and Pursuit for Contrast Spontaneous (p=0.002) and Motion Spontaneous (p=0.01); but not for Contrast Evoked and Motion Evoked (both n.s.); conversely for the Shape stimuli, there was a significant difference for Evoked (p=0.03), but not Spontaneous (n.s.), which explained the three-way interaction. These variations between the stimulation conditions might highlight perceptual differences associated to intrinsic vs. extrinsic line-ends, as only the Shape stimuli display extrinsic line-ends (with occluders materialized by light grey bars).

For the duration ratios \([\text{total bound percept duration} / (\text{total bound duration} + \text{total unbound duration})]\), there was a significant effect of Stimulus (F(2,18)=5.418; ε=0.790; p=0.02) and a significant interaction between Stimulus and Eye-movement (F(2,18)=5.244; ε=0.929; p=0.02), due to a tendency for overall less bound percept during pursuit (with respect to fixation), for the Motion and Contrast stimuli, compared to more bound percept (with pursuit) for the Shape stimulus.

Regarding the fine dynamics of the perceptual alternations, we analyzed the distributions of normalized episode durations (after removal of the first and last episodes of each block), separately for Fixation-Spontaneous, Fixation-Evoked, Pursuit-Spontaneous, Pursuit-Evoked sessions. Kolmogorov-Smirnov tests revealed that episode durations followed a log-normal distribution for the Spontaneous sessions (p=0.20, for Fixation; p=0.10, for Pursuit), which was not the case for the case for the Evoked sessions (p<0.01, for Fixation; p<0.01, for Pursuit).

Altogether, pursuit eye-movements modified the number of transitions and the duration ratios between the two percepts in a stimulus-dependent way, suggesting that pursuit eye-movements may affect differentially form/motion binding (Hafed & Krauzlis, 2006) depending on whether extrinsic or intrinsic line-ends are present in the bistable stimulus. We however note that pursuit was sustained and ample in the pursuit condition.

Overall, the results related to perceptual alternations during the fixation runs are similar to those found in the scanner, suggesting that observers did accurately fixated in the two situations. If observers did involuntary track the stimuli during the fixation sessions in a way similar to the pursuit performed in the tracking sessions, no statistical differences between both conditions should be observed.
Eye movements

Overall, eye-movement amplitude was small in the fixation condition and the relative phase of horizontal and vertical eye-movements, expected to be close to ±90° when pursuing a circular translation, was highly variable, not phase-locked to the stimulus and similar for both percepts. In contrast, eye-movement were large in the pursuit condition and the relative phase between horizontal and vertical was close to 90°, as expected if observers accurately pursued the central target (Supplementary Figure). These results indicate that subjects maintained a fairly accurate fixation when instructed to do so (as in the scanner), and that residual eye-movements are similar for both percepts.

Saccades and blinks were slightly more numerous before and after a transition but these represented less than 0.04 saccades and 0.05 blinks per transition, in agreement with previous studies (van Dam and van Ee, 2006; Hupé et al., 2009). In addition, in an experiment using stimuli similar to the moving displays used here (Hupé et al., 2009, Experiment 2, Conditions 1 & 2), a pupil dilation was found, starting 300 to 400 ms before the report of a transition (Figure 2 in Hupé et al.). As pupil always dilates, independently from the particular percept following a switch (Figure 3 in Hupé et al.), it is unlikely that it accounts for the observed percept-dependent BOLD activities.

Interpretation

Overall, neither could we identify a systematic effect of active pursuit on perceptual bistability (see the behavioral data analysis above), nor could we correlate perceptual alternation dynamics with any oculomotor parameter (saccades, blinks, or fixation accuracy). It thus seems unlikely that eye-movements can account for the observed perceptual effects or can entirely account for BOLD signal changes.
**Supplementary Figure.** Eye-movements were recorded offline the MRI acquisition during perceptual reports of bound and unbound percepts, in Fixation and Pursuit conditions. Traces were averaged across the six displays used in the fMRI experiment (N=10, 8 subjects also took part in the fMRI experiment). a, b Horizontal and vertical eye movement gains around button presses signaling a transition from bound to unbound (Red) and from unbound to bound perceptual states (Green) during fixation (dark traces, small gain) and during pursuit of a central target moving as the diamond shape (light traces, high gain). c Relative phase between horizontal and vertical eye movements in the fixation (dark traces) and pursuit conditions (light traces).

**References**


**B. Comparison of hMT+ and Middle temporal clusters in the motion-area localizer and main contrasts**

In this analysis, we carefully selected for each subject (and each hemisphere) the voxel with maximal activation in a given contrast in the vicinity of the temporo-occipital junction, where activities in the hMT+ complex were expected to be found.

**Location of the voxels with maximal activation**

MT location was searched for in the localizer block (which used static and moving RDKs), in the contrast Motion > Static. Individual maps were inspected with the following thresholds: p<0.01 for the height threshold, and 30 voxels for the spatial threshold. For each hemisphere of each subject, we manually identified the voxel with maximal activation (highest t-value) in the vicinity of the temporo-occipital junction, where hMT+ is expected to lie. The position of the maximum activity was unambiguous in all cases but one (with several local maxima), where we retained the maximum closest to that of the group analysis. As expected from earlier work, the voxels identified lied closed to the posterior parts of the ITS (Inferior Temporal Sulcus) and MTG (Middle Temporal Gyrus) (see Supplemental Table).
The voxel with maximal activity at the temporo-occipital junction area was also searched for in the main contrasts Unbound > Bound (for both transient and sustained activity). Individual maps were inspected with the following thresholds: \( p < 0.01 \) for the height threshold, and 30 voxels for the spatial threshold. One (or more) cluster(s) in the hMT+ region was (were) evident in all subjects in both hemispheres for the transient response, and in 21 hemispheres out of 24 for the sustained response (see Supplemental Table). For the transient response, in all cases but four, the voxel selected was the voxel with the highest t-value of a cluster centered around the occipito-temporal junction (in the posterior part of the MTG/STG). In the other cases, the cluster of interest had its maximal activity anteriorly in the STG/STS (1 case), superiorly at the parieto-occipital junction (2 cases), or in the motor area (1 case). In these cases, we retained the local maximum that fell close to the TO junction. For the sustained activity, in all cases but one, the voxel selected was the voxel with the highest t-value of a cluster centered around the occipito-temporal junction (in the posterior part of the MTG/STG). In one case, the cluster of interest had its maximal activity anteriorly in the ITS, so we retained the local maximum closest to the TO junction.

**Supplemental Table.** MNI coordinates of the voxels with maximal activation in the vicinity of the TO junction in several contrasts.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Localizer (Motion &gt; Static)</th>
<th>Unbound &gt; Bound, transient activity</th>
<th>Unbound &gt; Bound, sustained activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x y z</td>
<td>x y z</td>
<td>x y z</td>
</tr>
<tr>
<td></td>
<td>Left Right</td>
<td>Left Right</td>
<td>Left Right</td>
</tr>
<tr>
<td>1</td>
<td>-54 -72 0</td>
<td>-58 -74 2</td>
<td>-54 -76 0</td>
</tr>
<tr>
<td>2</td>
<td>-50 -68 6</td>
<td>-50 -72 -4</td>
<td>-54 -72 -4</td>
</tr>
<tr>
<td>3</td>
<td>-46 -78 0</td>
<td>-50 -76 6</td>
<td>-52 -74 12</td>
</tr>
<tr>
<td>4</td>
<td>-46 -72 -4</td>
<td>-48 -70 8</td>
<td>-48 -72 4</td>
</tr>
<tr>
<td>5</td>
<td>-46 -70 0</td>
<td>-60 -72 12</td>
<td>-48 -70 10</td>
</tr>
<tr>
<td>6</td>
<td>-54 -70 2</td>
<td>-54 -72 8</td>
<td>-42 -72 4</td>
</tr>
<tr>
<td>7</td>
<td>-36 -68 6</td>
<td>-54 -68 12</td>
<td>-42 -72 14</td>
</tr>
<tr>
<td>8</td>
<td>-50 -66 4</td>
<td>-60 -70 2</td>
<td>-54 -64 10</td>
</tr>
<tr>
<td>9</td>
<td>-40 -68 6</td>
<td>-36 -66 8</td>
<td>-38 -74 2</td>
</tr>
<tr>
<td>10</td>
<td>-52 -68 6</td>
<td>-48 -60 10</td>
<td>-44 -72 10</td>
</tr>
<tr>
<td>11</td>
<td>-50 -64 2</td>
<td>-50 -72 14</td>
<td>-48 -68 18</td>
</tr>
<tr>
<td>12</td>
<td>-46 -74 10</td>
<td>-58 -76 24</td>
<td>-48 -74 20</td>
</tr>
</tbody>
</table>

**Comparison of the maximal activation coordinates in the different contrasts**

Comparing the coordinates of the maxima of activity of these clusters (Unbound > Bound, transient or sustained activity vs. localizer) did not reveal significant differences except on the z-axis.
where the maxima of activity were higher for the Unbound > Bound contrasts (transient or sustained activity) than for the localizer in both hemispheres (Wilcoxon paired-value tests: \( p<0.05 \), except for the comparison between the localizer and Unbound > Bound contrast for sustained activity in the right hemisphere: \( p=0.09 \)). The maximal activity for the localizer sequence was always below \( z=10 \) mm, whereas it extended up to \( z=24 \) mm in the main contrasts, with nevertheless subjects having peaks of activity below \( z=10 \) mm. This suggests that the clusters observed when comparing unbound and bound percepts encompass not only the core of the MT region, but also more dorsal areas (putatively MST). This interpretation is corroborated by the fact that a number of the subjects exhibiting maximal activities below \( z=10 \) mm in the main contrasts had additional local maxima around \( z=20 \) mm in the same contrasts (data not shown).

Interpretation

The analysis reported above along with the time-courses of activations reported in the main text point to a dissociation between the core of MT, with moderate differences between the two percepts and the upper part of the hMT+ complex (putatively MST), where the difference between the two percepts is more pronounced and a decrease of activity accompanies the emergence of the Unbound percept.