**Supplementary Methods and Results**

Quantifying univariate BOLD response changes in the attention task

To measure the change in the average magnitude of the BOLD response across all voxels in each ROI as a function of stimulus contrast and cue condition, we averaged BOLD responses from attended quadrants only. First, we selected voxels within each ROI that responded significantly to the stimulus locations in each quadrant. Using the mapping task data, we labeled the trials based on the quadrant that the wedge was presented in, and then using a similar procedure for identifying ROIs from the localizer task, performed a GLM with FSL FEAT. We contrasted beta estimates for one quadrant against estimates for all other quadrants and repeated this across all quadrants to select significant voxels for each quadrant ($p < 0.05$; FDR corrected). Next, we shifted all timepoints in the attention task by 4 seconds to account for the hemodynamic lag. Then, we averaged voxel responses during the cue-to-target delay period (6-8 s window after cue offset) for each trial in the attention task from the voxels that are selective for the cued quadrant, within each retinotopically defined ROI (Supplementary Figure 2).

To evaluate the impact of the attention cue and the contrast of the noise stimulus on the mean amplitude, we performed a randomization test for a three-way ANOVA on the mean BOLD responses with cue type, contrast level, and ROI as factors. Cue and ROI were considered as categorical factors, and contrast was considered as a continuous factor. First, we tested the main effect of the three factors. To test the effect of cue, cue labels were shuffled across trials, restricted within each contrast condition, ROI, and participant, and then averaged for each condition. Then, we performed ANOVA on this data, only testing for the three main effects without the interaction terms, to obtain the F-value for the main effect of cue. This procedure was repeated across 1000 iterations, yielding a distribution of 1000 F-values under the null hypothesis that cue conditions did not affect the mean amplitude. A p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels. The same procedure was repeated to test the main effects of contrast and ROI.

Next, we tested the interaction effects between the three factors. We first performed an ANOVA with only the main effects and obtained the residuals. This was done to exclude the variance explained by the main effects. To test the interaction between cue and contrast, we shuffled the labels of the residual data, restricted within each ROI and participant. Then, we performed ANOVA on the residuals, testing for the three main effects and the three two-way interaction effects, to obtain the F-value for the interaction effect between cue and contrast. This procedure was repeated across 1000 iterations, yielding a distribution of 1000 F-values, and a p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels. The same procedure was repeated to test interaction effects between cue and ROI, and contrast and ROI. Finally, to test the three-way interaction between cue type, contrast level, and ROI, we first performed an ANOVA with the main effects and the three two-way interactions and obtained the residuals. Then, we shuffled the labels of the residual data without restriction and subjected the data to a full three-way

ANOVA with the three main effects, three two-way interactions, and the three-way interaction. This procedure was repeated across 1000 iterations, yielding a distribution of 1000 F-values, and a p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels.

To further examine the interaction effect between cue type and contrast level, we performed a one-way ANOVA on the mean response amplitude with contrast level as a continuous factor, separately for each cue condition and within each ROI. We evaluated the statistical significance of this effect with a randomization test by shuffling the contrast label within each participant and each ROI and repeating the one-way ANOVA to obtain an F-value. Repeating this procedure 1000 times yielded a distribution of F-values under the null hypothesis that contrast had no impact on the responses. P-values were calculated by taking the proportion of F-values from the shuffled data that exceeded the F-value obtained using the unshuffled data. We also tested the effect of cue within each ROI by performing paired t-tests after collapsing across all contrast levels. We then repeated the t-test after shuffling the cue labels across 1000 iterations to obtain 1000 t-values under the null hypothesis that the attention-cue label had no effect on the mean BOLD responses. We then obtained p-values by calculating the proportion of shuffling iterations on which the shuffled t-value exceeded the real t-value, and the proportion on which the real t-value exceeded the shuffled t-value. We then took the minimum value between the two and multiplied it by two for a two-tailed test.

Assessing the effect of stimulus contrast on the top-down attentional representation

While not our primary manipulation of interest, we included the noise stimulus during the cue-to-target delay period in the attention task to determine if the precision of top-down attentional modulations interacted with the strength of concurrent bottom-up sensory drive. For example, a focused attention cue is expected to produce a more precise spatial representation than a diffuse cue. However, this relative increase in precision with focal attention might be further amplified (or dampened), if the top-down attention signals interact with the strength of bottom-up stimulus drive.

To evaluate the impact of the attention cue and the contrast of the noise stimulus on the top-down precision, we performed a randomization test for three-way ANOVA on the decoding accuracy with cue, contrast, and ROI as factors. Cue and ROI were considered as categorical factors, and contrast was considered as a continuous factor. First, we tested the main effect of the three factors. To test the effect of cue type, cue labels were shuffled across trials, restricted within each contrast condition, ROI, and participant, and then the classifier correctness for each trial was averaged to obtain decoding accuracy for each condition. Then, we performed an ANOVA on this data, only testing for the three main effects without the interaction terms, to obtain the F-value for the main effect of cue. This procedure was repeated across 1000 iterations, which yields a distribution of 1000 F-values under the null hypothesis that cue conditions did not affect the decoding accuracy. A p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels. The same procedure was repeated to test the main effects of contrast and ROI.

Next, we tested the interaction effects between the three factors. We first performed an ANOVA on the decoding accuracy with only the main effects and obtained the residuals. This was done to exclude the variance explained by the main effects. To test the interaction between cue type and contrast level, we shuffled the labels of the residual data, restricted within each ROI and participant. Then, we performed an ANOVA on the residuals, testing for the three main effects and the three two-way interaction effects, to obtain the F-value for the interaction effect between cue type and contrast level. This procedure was repeated 1000 times, yielding a distribution of 1000 F-values, and a p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels. The same procedure was repeated to test the other interaction effects between cue type and ROI, and contrast level and ROI. To test the three-way interaction between cue, contrast, and ROI, we first performed ANOVA on the decoding accuracy with the main effects and the three two-way interactions and obtained the residuals. Then, we shuffled the labels of the residual data without restriction and subjected the data to a full three-way ANOVA with the three main effects, three two-way interactions, and the three-way interaction. This procedure was repeated across 1000 iterations, yielding a distribution of 1000 F-values, and a p-value was obtained by taking the proportion of F-values from this null distribution that exceeded the F-value obtained using the data with unshuffled labels.

To further examine the significant interaction effects (between cue type and ROI, contrast level and ROI) from the above analysis, we tested the effect of cue type and contrast level within each ROI. To test the effect of cue in individual ROIs, we first shuffled the cue labels for across trials within each ROI and participant, and averaged classifier accuracy across trials to obtain decoding accuracy for each cue condition. Next, we performed a paired t-test between the decoding accuracies from each cue condition to obtain the t-value for each ROI. This procedure was repeated across 1000 iterations, yielding a distribution of 1000 F-values for each ROI, under the null hypothesis that cue conditions did not affect decoding accuracy. Then, a p-value for each ROI was calculated by taking the proportion of iterations where shuffled t-values exceed the real t-value, the proportion of iterations where the real t-value exceeded the shuffled t-values, and then taking the minimum value and multiplying it by two for a two-tailed test. A similar procedure was repeated to test the effect of contrast level within each ROI, but we shuffled the contrast labels and used a one-way ANOVA with contrast as a continuous variable. In this case, the p-value for each ROI was calculated as the proportion of shuffled F-values that exceeded the real F-value.

Eye-tracking session

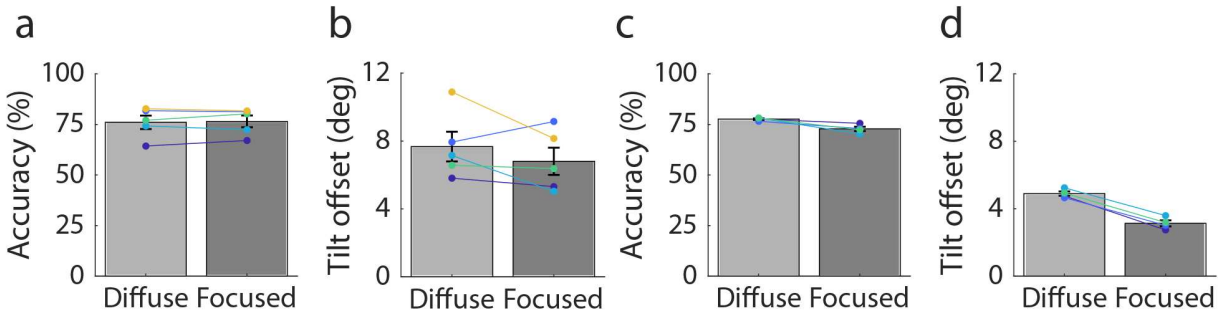
Eye-tracking data were collected in an entirely separate session to the MRI scanning session from four out of five fMRI participants who were available to come into the lab. An SR Eyelink 1000+ eye-tracker (SR Research) was used to collect continuous gaze-position data at a sampling rate of 1000 Hz. Participants performed only the top-down attention task and completed the same number of trials as in the scanner. All aspects of the task were the same, except that it was self-paced, which means that each trial terminated when the response was made, and participants initiated the next trial. Tilt offset for each condition was manually adjusted after every 27 trials to keep accuracies around 75%.

Analysis of eye-tracking data

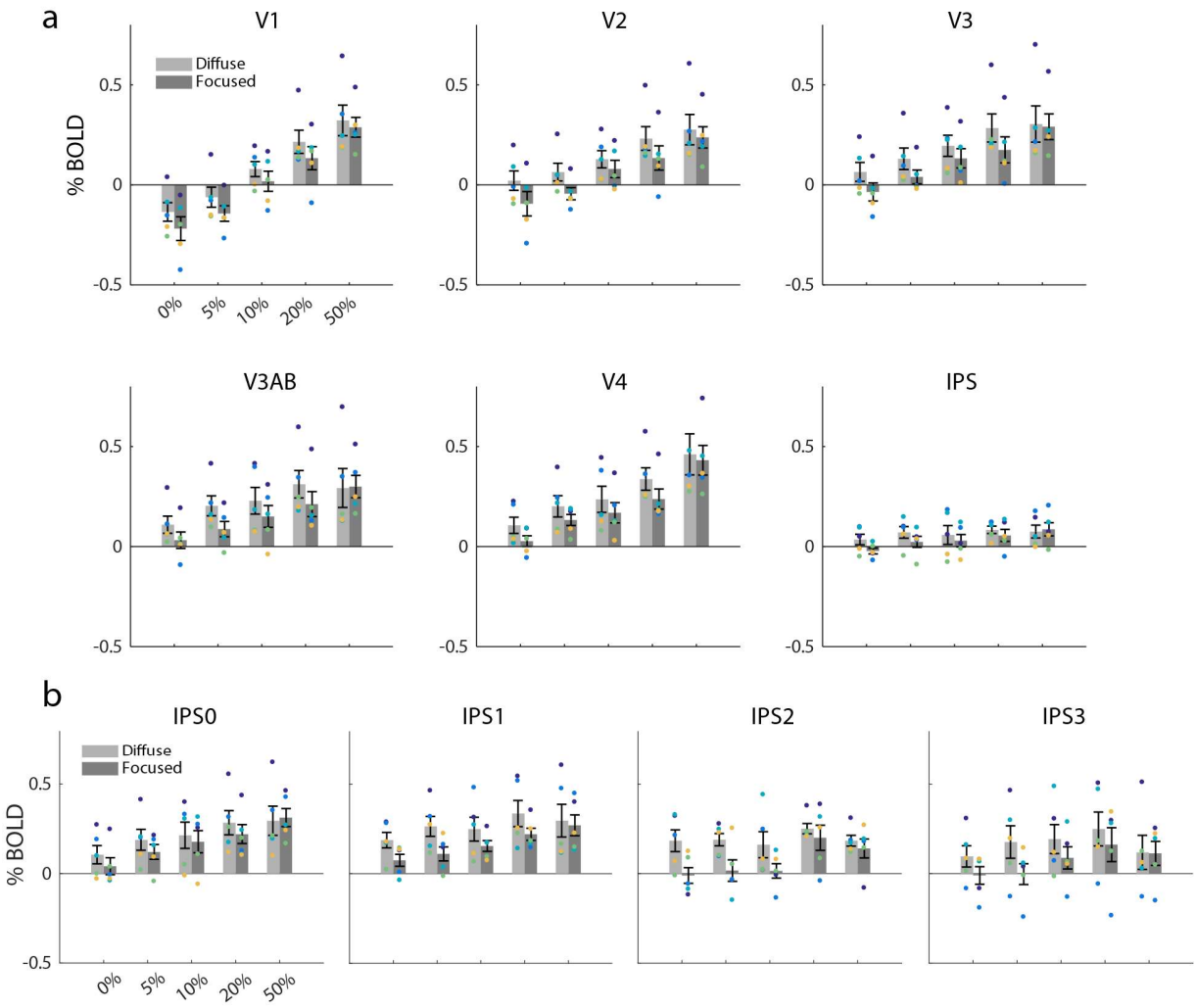
Gaze positions during each run were preprocessed by first removing data points during blinks and filling in those data points by interpolating between eye positions before and after the blink. Then, separately for each run, gaze positions were detrended, high-pass filtered at 0.1Hz, and low-pass filtered at 10Hz. Next, gaze positions for each trial were corrected by subtracting the mean position during the last 200-ms in the pre-cue period, to account for task-unrelated drifts between trials and runs.
To examine eye movement away from fixation, average gaze positions during pre-cue, cue, delay period, and target presentation was calculated for each trial. To test if the eye movement away from fixation was in the direction of the cued location in each trial, the gaze position vector was projected onto the vector of the cued location to attain the magnitude of the projected vector (Supplementary Figure 7a).

Control analysis for an using data from the top-down attention task for voxel selection

Our main analysis used a separate localizer task to select visually responsive voxels for the decoding analyses. Because the localizer task used flickering checkerboard stimuli, there is a chance that this process might have biased our pool of selected voxels in favor of bottom-up processing/representations that could have inflated bottom-up decoding accuracy in early visual areas. To address this possibility, we conducted control analyses with two alternative voxel selection methods using only the top-down attention task to select the voxels. First, we re-labeled the trials of the top-down task based on the quadrant that was cued and used all runs to run one-way ANOVA for individual voxels within each retinotopically defined ROI and found voxels that were significantly modulated across the cued quadrants during the delay period between the cue and the target ($p < 0.05$; FDR corrected; see Supplementary Table 2 for the number of selected voxels). After a single set of voxels has been selected for each ROI, the same procedure for the decoding analysis described in the main text was carried out. Note that this method is circular because data from all runs were used in the ANOVA to select the most selective voxels. This was done intentionally as we wanted to bias this analysis in favor of finding high-precision representations in the top-down attention task (see Supplementary Figure 8 for decoding results). In addition to this circular analysis, we also used a second method of voxel selection that avoided this circularity by using a leave-one-run-out procedure: All but one run of the top-down task was used in the voxel selection process and the decoder was tested only on the left out run, and this was repeated leaving out a different run on each iteration (20 iterations total). The bottom-up decoding, in this case, was done on every set of selected voxels and the resulting decoding accuracy was averaged across iterations (see Supplementary Figure 9 for decoding results).

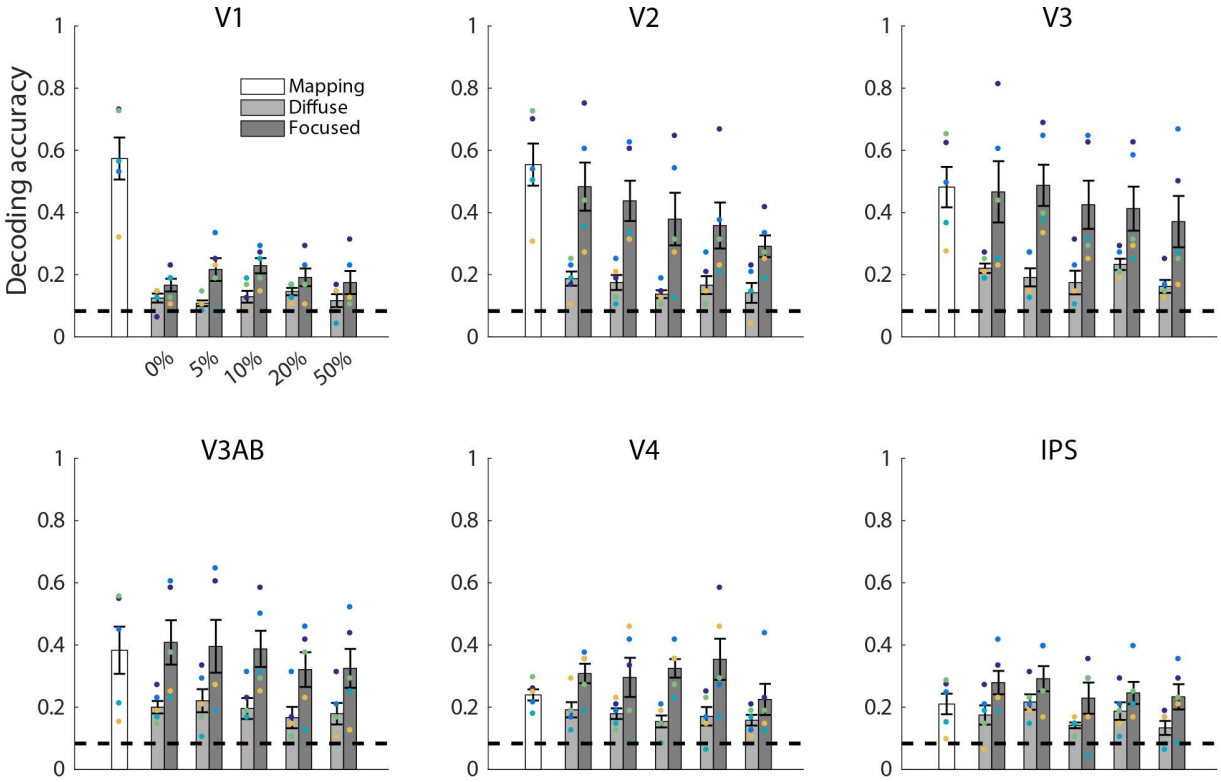



**Supplementary Figure 1**. Behavioral data for the top-down spatial attention task during the scanning and the eye-tracking session plotted separately. (a, b) Behavioral performance for the top-down spatial attention task during the scanning session. (a) Mean behavioral accuracy. Accuracy was comparable between the diffuse and the focused conditions. (b) Mean tilt offset. Tilt offsets were higher in the diffuse than in the focused condition. The effect of cue type on accuracy ($p = 0.33$) and on the tilt offset ($p = 0.12$) was not statistically significant. (c, d) Behavioral performance for the top-down spatial attention task during the eye-tracking session. (c) Mean behavioral accuracy. Accuracy was comparable between the diffuse and the focused conditions. (d) Mean tilt offset. Tilt offsets were higher in the diffuse than in the focused condition. The effect of cue type on accuracy and on the tilt offset were both significant ($p$'s $< 0.001$). Colored dots represent data from individual participants. Error bars represent $\pm 1$ standard error of the mean (SEM).

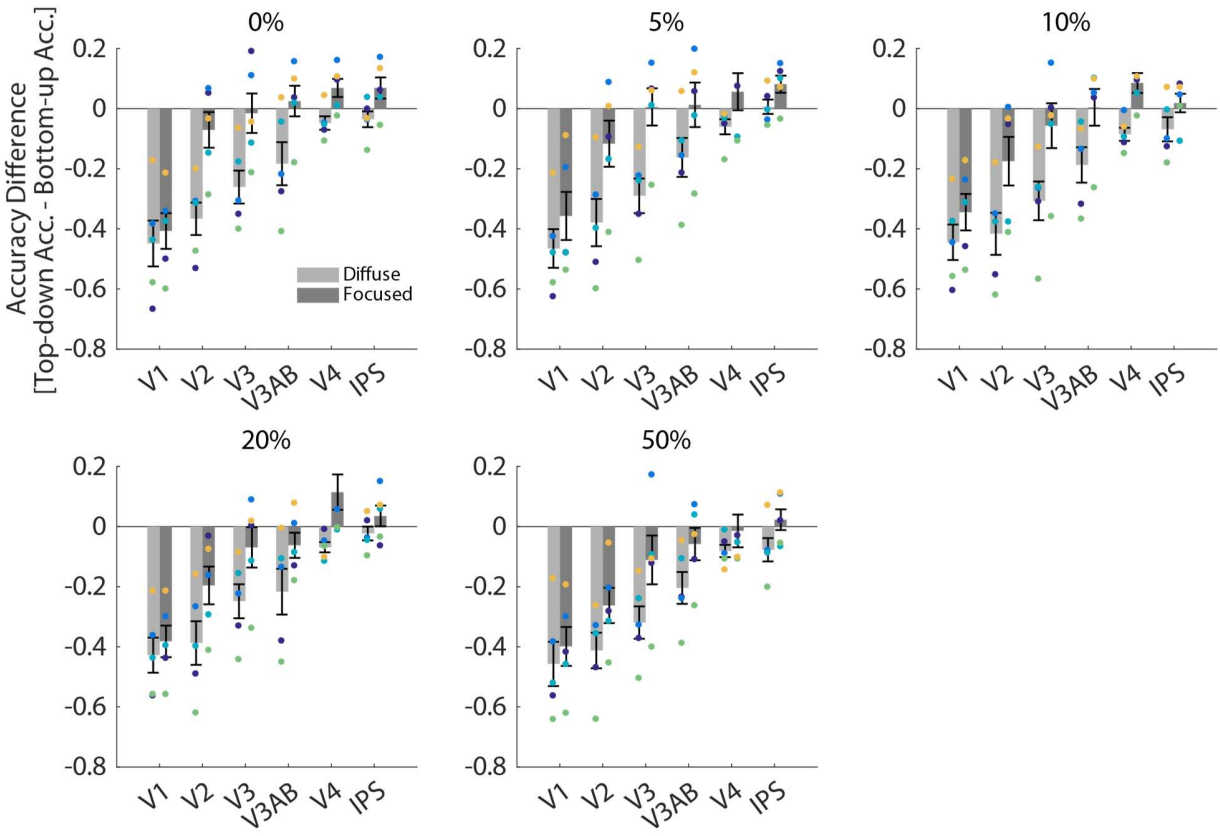



**Supplementary Figure 2.** (a) Univariate BOLD amplitude changes in the top-down spatial attention task. Mean BOLD amplitude was higher in the diffuse than in the focused condition, and it increased as a function of the contrast level of the noise stimulus that was presented during the cue-to-target interval. Mean response amplitude increased as a function of stimulus contrast. A three-way ANOVA on the mean BOLD amplitude with cue condition, contrast level, and ROI as factors revealed significant main effects of all three factors (all p's < 0.01, p-values computed via randomization test, see Supplementary Methods, section *Quantifying univariate BOLD response changes in the attention task*). There were also significant interaction effects between cue and contrast ($p < 0.001$) and contrast and ROI ($p < 0.001$), but other interaction effects did not reach significance (cue X ROI: $p = 0.13$; cue X contrast X ROI: $p = 0.81$). To determine the direction of the interaction between cue and contrast, we performed a one-way ANOVA with contrast level as a continuous factor, separately for each cue condition within each ROI. This revealed a significant effect of contrast in all ROIs for both diffuse and focused conditions, except for IPS in the diffuse condition (all ROIs except for IPS in both cue conditions: p's < 0.001; IPS in diffuse: $p = 0.16$; IPS in focused: $p < 0.01$). Colored dots represent data from individual participants. Error bars represent ±1SEM. (b) Univariate BOLD amplitude changes in the top-down spatial attention task for IPS

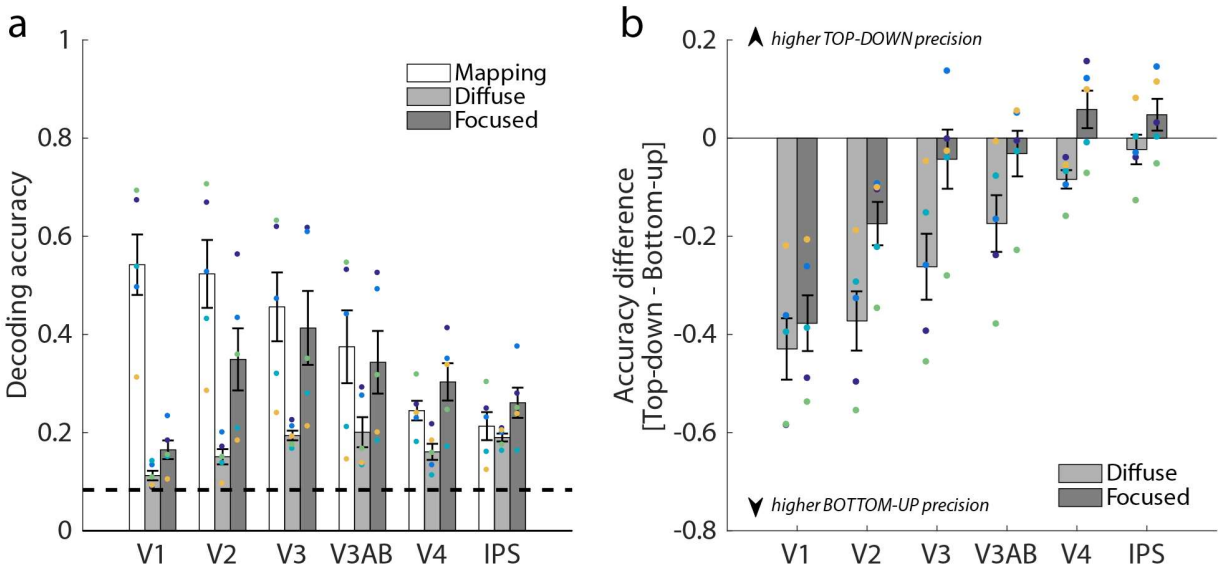
sub-regions. When collapsed across cue conditions, the effect of contrast was significant in IPS0 and IPS1 in both cue conditions (p's < 0.05) and IPS2 and IPS3 in the focused condition (p's < 0.05). However, stimulus contrast did not have a significant effect in IPS2 and IPS3 in the diffuse condition ($p = 0.67$ and $p = 0.51$, respectively). When collapsed across contrast levels, the effect of cue was significant in IPS1 ($p < 0.001$) but did not reach significance in the rest (IPS0: $p = 0.13$, IPS2: $p = 0.12$, IPS3: $p = 0.13$). Colored dots represent data from individual participants. Error bars represent $\pm$1SEM.



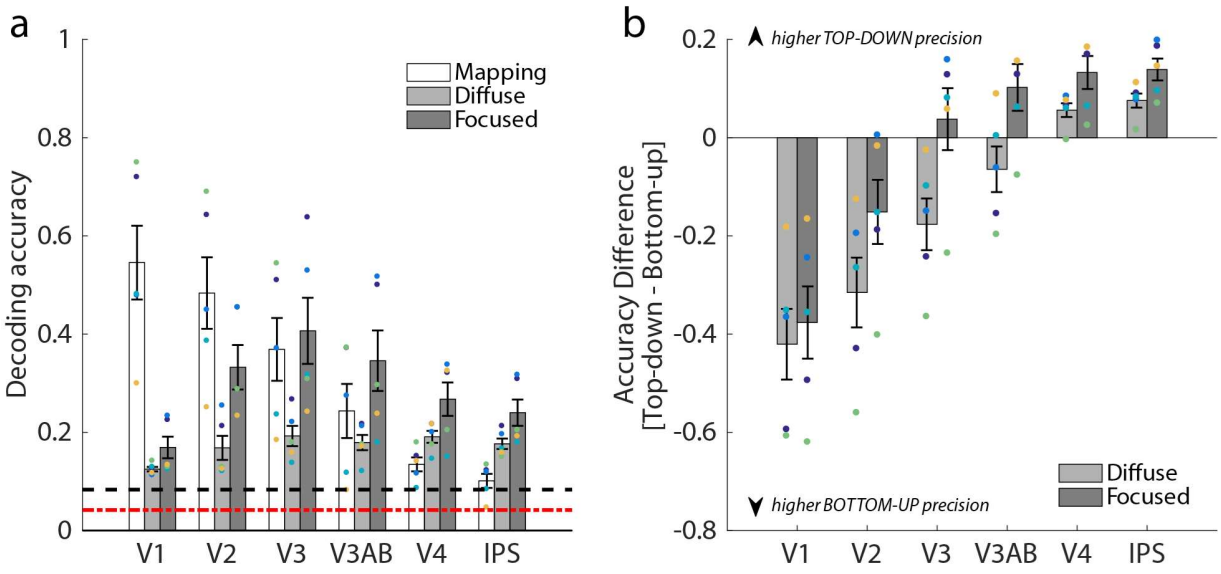

**Supplementary Figure 3.** Decoding accuracy in the top-down spatial attention task, plotted separately in each area for each contrast level and cue condition. For comparison, the same values for bottom-up decoding accuracies from Figure 2a are plotted together (white bars). Decoding accuracy in the attention task were generally higher in the focused than in the diffuse condition, and decreased as a function of contrast in V2, V3, V3AB, and IPS. While top-down decoding accuracy was much lower than bottom-up decoding accuracy in V1, accuracy in the mapping task and the focused condition were comparable in later areas (e.g. V3AB, V4, IPS). A three-way ANOVA on the mean decoding accuracies with cue condition, contrast level, and ROI as factors revealed significant main effects of all three factors (all p's < 0.01). There were also significant interactions between cue type and ROI ($p < 0.01$) and between contrast level and ROI ($p < 0.05$). Other interaction effects did not reach significance (cue X contrast: $p = 0.10$; cue X contrast X ROI: $p = 0.12$). To further explore the cause of the significant interactions, we tested the effect of cue and contrast within each ROI. When collapsed across contrast levels, decoding performance was significantly higher in the focused than in the diffuse condition in all ROIs except V1 (two-tailed t-test, V1: $p = 0.07$, other ROIs: p's < 0.05). Next, we evaluated the impact of the noise stimulus contrast presented during the cue-to-target delay period, by performing a one-way ANOVA with contrast level as a continuous variable after collapsing across cue conditions. This revealed a significant decrease in decoding accuracy as a function of contrast in V2, V3, V3AB, and IPS (p's < 0.05). This effect may have been due to higher contrast noise stimuli functioning as distractors that interfered with the maintenance of attention at the cued location(s). Colored dots represent data from individual participants. Error bars represent ±1SEM. The dotted line indicates chance performance (1/12, or ~.833).



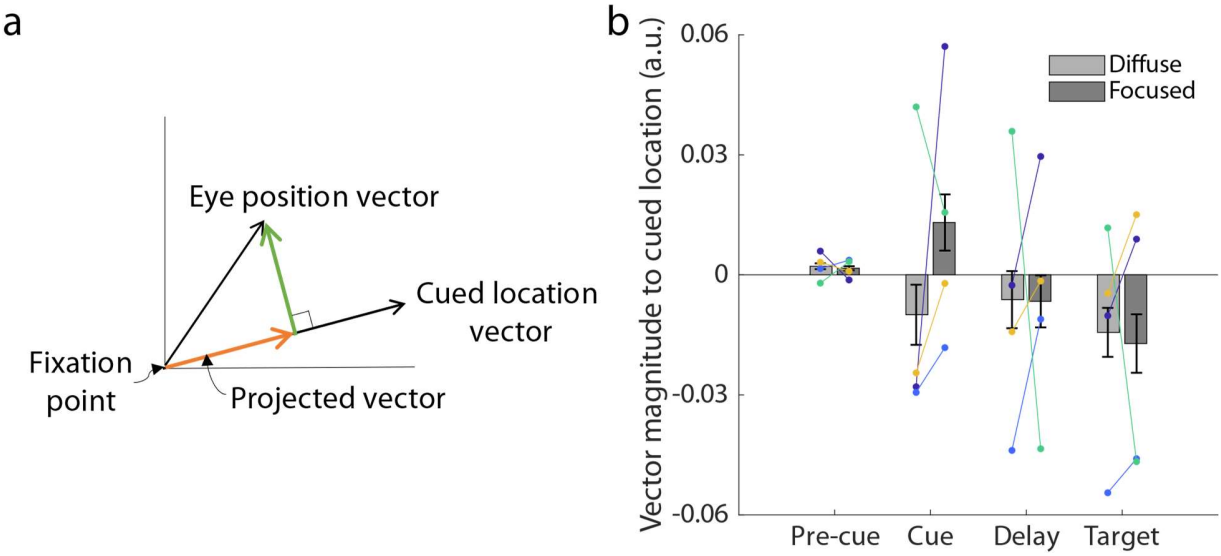

**Supplementary Figure 4.** Difference in decoding accuracies between top-down attention and bottom-up mapping task, plotted separately for each contrast level. Difference scores were obtained following the same procedure as described in Figure 2b but done separately for each contrast level instead of collapsing across the contrast levels. Negative values indicate that the ROI had higher decoding accuracy in the mapping task, consistent with higher bottom-up precision. Positive values indicate that the ROI had higher decoding accuracy in the top-down task, consistent with higher top-down precision. Difference scores for each contrast level showed a similar pattern to the collapsed result (shown in Figure 2). Colored dots represent data from individual participants. Error bars are ±1SEM.



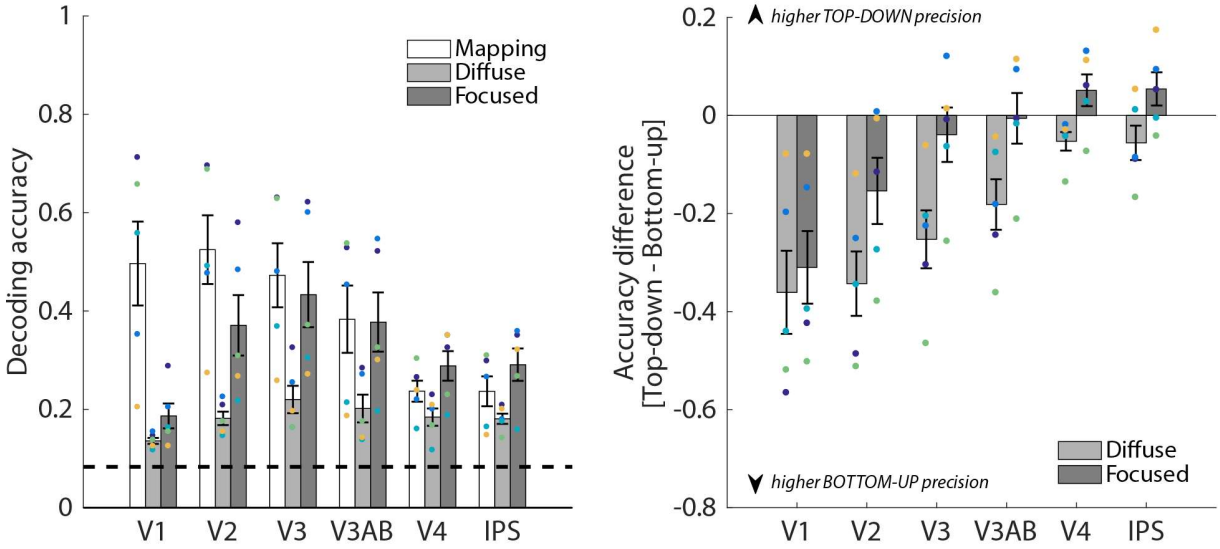

**Supplementary Figure 5.** (a) Decoding accuracy from top 300 voxels in each ROI in the bottom-up spatial mapping task (white bars) and the top-down spatial attention task (Diffuse and Focused conditions; light gray and dark gray bars). Decoding accuracies showed a similar pattern to when the number of voxels were not restricted (shown in Figure 2a). Filled, colored dots represent data from individual participants, and error bars represent ±1SEM. The dotted line indicates chance performance (1/12 or ~.833). (b) Difference in decoding accuracies from top 300 voxels in each ROI between the bottom-up mapping task and the top-down attention task in each ROI. To obtain this difference score, the bottom-up decoding accuracy was subtracted from the top-down decoding accuracy within each ROI, separately for the diffuse and the focused conditions. Negative difference score indicates that an ROI had higher decoding accuracy in the mapping task, consistent with higher bottom-up precision. Positive difference score indicates that an ROI had higher decoding accuracy in the top-down task, consistent with higher top-down precision. While V1 showed much higher bottom-up precision compared to top-down attentional precision, this discrepancy gradually decreased in the later areas, similar to Figure 2b. Colored dots represent data from individual participants. Error bars are ±1SEM.



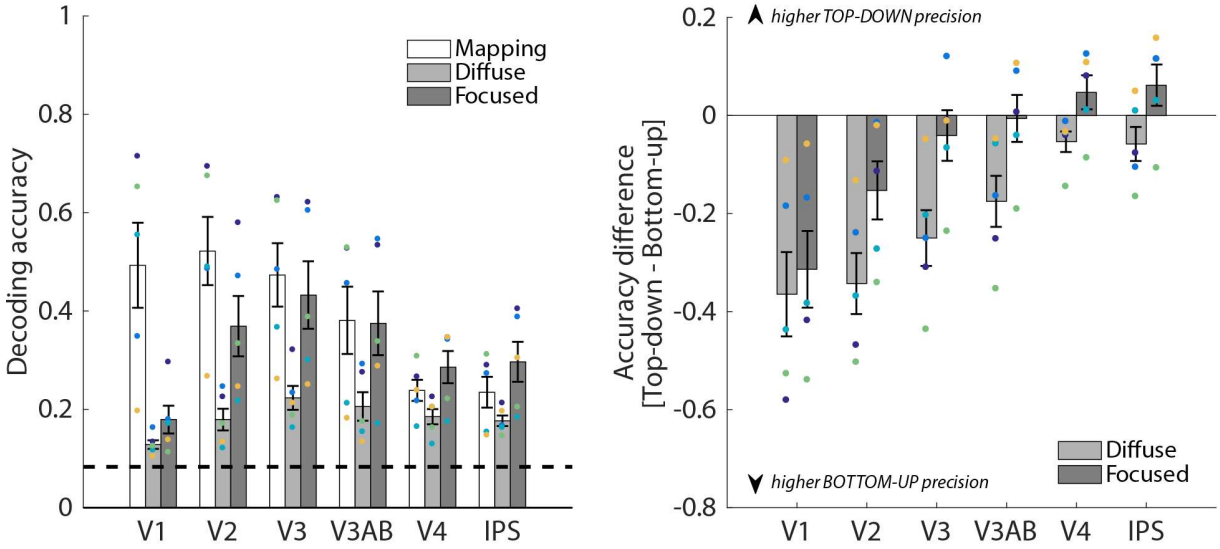

**Supplementary Figure 6.** Similar results to Figure 2 when using 24-way decoding of the mapping task. Classifiers were trained on the mapping task data without combining the wedges. Decoding accuracy for the mapping task was calculated by comparing the predicted location and the presented location. Decoding accuracy for the attention task was calculated by comparing the predicted location and the attended location, and a given trial was marked as correct if the predicted location from the 24-way decoder was included in the attended location. The general pattern of results was similar to Figure 2. (a) Decoding accuracy based on fMRI activation patterns in the bottom-up spatial mapping task (white bars) and the top-down spatial attention task (Diffuse and Focused; gray and dark gray bars). Note that the chance level of decoding accuracy is different across tasks: 1/24 for mapping (red, dash-dot line), 1/12 for attention task (black, dashed line). While top-down decoding accuracy was much lower than bottom-up decoding accuracy in V1, accuracy in the mapping task and the focused condition was comparable in later areas (e.g. V3AB, V4, IPS), leading to an interaction between task type (bottom-up vs top-down) and visual area. Filled, colored dots represent data from individual participants, and error bars represent ±1SEM. (b) Difference in decoding accuracies between top-down attention and bottom-up mapping task. While V1 showed much higher bottom-up precision compared to top-down attentional precision, this discrepancy gradually decreased in the later areas. Colored dots represent data from individual participants. Error bars are ±1SEM.




**Supplementary Figure 7.** Eye-tracking results from an independent eye-tracking session conducted outside of the scanner. Eye-tracking data were collected form four out of five fMRI participants. (a) An example of the vector projection calculation. A vector for average eye position was projected onto the vector for the cued location for that trial. The magnitude of the projected vector (orange) was calculated for each period (pre-cue, cue, delay, and target) within each trial. (b) Average magnitude of the projected vector towards the cued location. A larger value means that the gaze moved away from fixation in the direction of the cued location. Although overall eye movement away from fixation was larger for focused condition compared to diffuse condition during delay and target presentation, these eye movements were not systematically related to the cued location (all p's > 0.1).




**Supplementary Figure 8.** Similar results to Figure 2 when using another way to select voxels based on the top-down task. To bias the selection in favor of the top-down decoding as much as possible, all runs of the top-down task were used to select voxels that changed their response significantly across the cued quadrants during the delay period. The general pattern of results was similar regardless. (a) Decoding accuracy based on fMRI activation patterns in the bottom-up spatial mapping task (white bars) and the top-down spatial attention task (Diffuse and Focused; gray and dark gray bars). While top-down decoding accuracy was much lower than bottom-up decoding accuracy in V1, accuracy in the mapping task and the focused condition was comparable in later areas (e.g. V3AB, V4, IPS), leading to an interaction between task type (bottom-up vs top-down) and visual area. Filled, colored dots represent data from individual participants, and error bars represent ±1SEM. The dotted line indicates chance performance (1/12 or ~.833). (b) Difference in decoding accuracies between top-down attention and bottom-up mapping task. While V1 showed much higher bottom-up precision compared to top-down attentional precision, this discrepancy gradually decreased in the later areas. Colored dots represent data from individual participants. Error bars are ±1SEM.




**Supplementary Figure 9.** Similar results to Figure 2 when selecting voxels based on the top-down task. As a control analysis, instead of using the separate localizer task to define visually responsive voxels, voxels that changed their response significantly across the cued quadrants during the delay period in the top-down attention task were used in the decoding analysis. To avoid circularity in the top-down decoding accuracy, a leave-one-run-out method was used, in which the run that was not used in the voxel selection step was used as a test dataset in the decoding analysis (see section Control analysis for an alternative voxel selection method in the Supplementary Methods). (a) Decoding accuracy based on fMRI activation patterns in the bottom-up spatial mapping task (white bars) and the top-down spatial attention task (Diffuse and Focused; gray and dark gray bars). While top-down decoding accuracy was much lower than bottom-up decoding accuracy in V1, accuracy in the mapping task and the focused condition was comparable in later areas (e.g. V3AB, V4, IPS), leading to an interaction between task type (bottom-up vs top-down) and visual area. Filled, colored dots represent data from individual participants, and error bars represent ±1SEM. The dotted line indicates chance performance (1/12 or ~.833). (b) Difference in decoding accuracies between top-down attention and bottom-up mapping task. While V1 showed much higher bottom-up precision compared to top-down attentional precision, this discrepancy gradually decreased in the later areas. Colored dots represent data from individual participants. Error bars are ±1SEM.

**Supplementary Table 1.** Number of participants (out of 5 total) who showed significant effects ($p < 0.05$) from the main statistical tests reported when analyzed at a single-subject level with trial as the unit of variance.

*Two-way ANOVA on decoding accuracy*

| | |
|---|---|
| Cue | 5 |
| ROI | 5 |
| Cue X ROI | 3 |

*Effect of cue within each ROI*

| | V1 | V2 | V3 | V3AB | V4 | IPS |
|---|---|---|---|---|---|---|
| Cue | 2 | 5 | 5 | 5 | 3 | 3 |

*Two-way ANOVA on accuracy difference*

| | |
|---|---|
| Cue | 5 |
| ROI | 0 |
| Cue X ROI | 3 |

**Supplementary Table 2.** Average number of voxels selected for the main analysis based on the functional localizer task and the supplementary analyses based on the top-down attention task. Values are mean number of voxels selected for each retinotopic ROI across all subjects ±1SEM.

| *ROI* | *defined using localizer task (used in all main analyses)* | *defined using all runs of top-down task (Supp. Fig. 8)* | *defined using top-down task with the leave-one-run-out procedure (Supp. Fig. 9)* |
|---|---|---|---|
| V1 | 776.2±103.32 | 754±323.06 | 723.07±320.82 |
| V2 | 723.6±59.41 | 804.6±218.24 | 772.41±219.11 |
| V3 | 714±121.98 | 808.6±129.46 | 777.4±131.57 |
| V3AB | 608.6±250.11 | 629.2±175.08 | 609.32±173.6 |
| V4 | 342.2±47.38 | 321.4±72.35 | 309.99±73.24 |
| IPS | 513.4±162.44 | 407.6±117.03 | 385.75±114.93 |