### Reporting Checklist for Nature Neuroscience

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Please note that in the event of publication, it is mandatory that authors include all relevant methodological and statistical information in the manuscript.

#### ∟ Statistics reporting, by figure

- Please specify the following information for each panel reporting quantitative data, and where each item is reported (section, e.g. Results, & paragraph number).
- Each figure legend should ideally contain an exact sample size (n) for each experimental group/condition, where n is an exact number and not a range, a clear definition of how n is defined (for example x cells from x slices from x animals from x litters, collected over x days), a description of the statistical test used, the results of the tests, any descriptive statistics and clearly defined error bars if applicable.
- For any experiments using custom statistics, please indicate the test used and stats obtained for each experiment.
- Each figure legend should include a statement of how many times the experiment shown was replicated in the lab; the details of sample collection should be sufficiently clear so that the replicability of the experiment is obvious to the reader.
- For experiments reported in the text but not in the figures, please use the paragraph number instead of the figure number.

Note: Mean and standard deviation are not appropriate on small samples, and plotting independent data points is usually more informative. When technical replicates are reported, error and significance measures reflect the experimental variability and not the variability of the biological process; it is misleading not to state this clearly.

<table>
<thead>
<tr>
<th>FIGURE NUMBER</th>
<th>WHICH TEST?</th>
<th>SECTION &amp; PARAGRAPH #</th>
<th>n</th>
<th>DESCRIPTIVE STATS (AVERAGE, VARIANCE)</th>
<th>P VALUE</th>
<th>DEGREES OF FREEDOM &amp; F/T/Z/R/ETC VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>one-way ANOVA</td>
<td>Fig. legend</td>
<td>9, 9, 10, 15 mice from at least 3 litters/group</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>p = 0.044</td>
</tr>
<tr>
<td>Results para 6</td>
<td>unpaired t-test</td>
<td>Results para 6</td>
<td>15 slices from 10 mice</td>
<td>error bars are mean +/- SEM</td>
<td>Results para 6</td>
<td>p = 0.0006</td>
</tr>
</tbody>
</table>

Nature Neuroscience: doi:10.1038/nn.3805
<table>
<thead>
<tr>
<th>FIGURE NUMBER</th>
<th>WHICH TEST?</th>
<th>SECTION &amp; PARAGRAPH #</th>
<th>EXACT VALUE</th>
<th>DEFINED?</th>
<th>REPORTED?</th>
<th>SECTION &amp; PARAGRAPH #</th>
<th>EXACT VALUE</th>
<th>P VALUE</th>
<th>DEGREES OF FREEDOM &amp; F/T/Z/R/ETC VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Gaussian fit to measure RF center and preferred orientation</td>
<td>Methods para 7</td>
<td>MG 37-43, MJ 39-43</td>
<td>number of effective recording sites</td>
<td>Methods para 7</td>
<td></td>
<td></td>
<td></td>
<td>R-squared &gt; 0.7 for goodness of fit</td>
</tr>
<tr>
<td>1e</td>
<td>linear regression of learning curve for each stimulus condition, t-test for slope significance</td>
<td>Results para 7-8</td>
<td>15, 17 (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td>error bars (shaded areas) are mean +/- SEM</td>
<td>Fig. legend</td>
<td>p-values are used for categorization (p&lt;0.05 defined as learned condition)</td>
<td>Results para 7-8</td>
</tr>
<tr>
<td>2c-h</td>
<td>Wilcoxon rank sum test for recording sites significantly modulated by visual contours</td>
<td>Fig. legend</td>
<td>40</td>
<td>number of trials in each condition</td>
<td>Results para 4</td>
<td>recording sites with p&lt;0.05 are marked in the plots</td>
<td>Fig. legend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>one-tailed unpaired t-test (pre-training vs 1st training day)</td>
<td>Results para 16</td>
<td>594/570 459/436 621/564 461/461</td>
<td>pooling recording sites in different conditions</td>
<td>Results para 16</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>p = 0.00003, p &lt; 1e-6, p = 0.001, p &lt; 1e-6 (Note: p-values smaller than 1e-6 are always indicated as p&lt;1e-6)</td>
<td>Results para 16</td>
</tr>
<tr>
<td>3b</td>
<td>linear regression of averaged d' over days, t-test for slope significance</td>
<td>Results para 16</td>
<td>15, 17 (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>p = 0.002, p = 0.0002, p = 0.0003, p &lt; 1e-6</td>
<td>Results para 16</td>
</tr>
<tr>
<td>Results, para 17-18</td>
<td>linear regression of averaged d' over days, t-test for slope significance</td>
<td>Results para 17-18</td>
<td>15, 17 (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td></td>
<td>saturated conditions: p = 0.12, p = 0.0003, p = 0.46, p = 0.33, non-learnable conditions: p = 0.01, p = 0.17, p = 0.04, p = 0.06</td>
<td>Results para 17-18</td>
<td>t (13) = -1.7, t(15) = -4.6, t(13) = -0.7, t(15) = 1.0, t(13) = 2.8, t(15) = -1.5, t(15) = 2.2, t(15) = -2.0</td>
</tr>
<tr>
<td>+</td>
<td>3c</td>
<td>linear regression of percent significant recording sites over days, t-test for slope significance</td>
<td>Results para 20</td>
<td>$15, 17$ (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>$p = 0.00004$, $p = 0.00006$, $p &lt; 1e-6$</td>
</tr>
<tr>
<td>+</td>
<td>Results para 22</td>
<td>Pearson correlation between animal’s behavior and neuronal responses across training days</td>
<td>Results para 22</td>
<td>$15, 17$ (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5 (p.6 para 2)</td>
<td>$p &lt; 1e-6$, $p &lt; 1e-6$</td>
<td>Results para 22</td>
<td>$r = 0.91$, $r = 0.76$</td>
</tr>
<tr>
<td>+</td>
<td>4d</td>
<td>permutation test by randomly shuffling the stimulus labels</td>
<td>Results para 23</td>
<td>1000</td>
<td>bootstrap resampling</td>
<td>Methods para 10</td>
<td>$p&lt;0.05$ as a criterion for defining significant contour-related responses in each time bin</td>
<td>Methods para 10</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>5c, 5d</td>
<td>10-fold cross-validation</td>
<td>Results para 28</td>
<td>$135,115$ (MG, MJ)</td>
<td>5 classifier training-testing repeats x number of learned conditions</td>
<td>Methods para 11</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>SVM vs LFD $p &lt; 1e-6$ (MG), $p = 0.00001$ (MG), $p = 0.00001$ (MJ), $p = 0.00001$ (MJ)</td>
</tr>
<tr>
<td>+</td>
<td>6c</td>
<td>linear regression, t-test for slope significance</td>
<td>Results para 31</td>
<td>$15, 17$ (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td>error bars (shaded areas) are mean +/- SEM</td>
<td>Fig. legend</td>
<td>$p = 0.000001$, $p = 0.00007$, $p = 0.28$, $p = 0.59$</td>
</tr>
<tr>
<td>+</td>
<td>6d</td>
<td>one-tailed paired t-test for difference between classifiers</td>
<td>Results para 33-34</td>
<td>$81,69$ (MG, MJ)</td>
<td>3 days x number of learned conditions</td>
<td>Methods para 11</td>
<td>error bars are mean +/- SEM</td>
<td>Fig. legend</td>
<td>SVM vs LFD $p &lt; 1e-6$ (MG), $p = 0.00001$ (MJ), LFD vs DoM $p = 0.006$ (MG), $p = 0.001$ (MJ)</td>
</tr>
<tr>
<td>+</td>
<td>Results para 31</td>
<td>linear regression of noise correlation over days</td>
<td>Methods para 12</td>
<td>$15, 17$ (MG, MJ)</td>
<td>number of days for perceptual training</td>
<td>Results para 5</td>
<td>$p = 0.00007$, $p = 0.00002$, $p = 0.009$, $p = 0.00004$, $p = 0.00005$ $p = 0.001$</td>
<td>Methods para 12</td>
<td>$t(13) = -5.8$, $t(13) = -5.1$, $t(13) = -3.1$, $t(15) = -5.8$, $t(15) = -5.6$, $t(15) = -3.9$</td>
</tr>
</tbody>
</table>
Representative figures

1. Are any representative images shown (including Western blots and immunohistochemistry/staining) in the paper?
   If so, what figure(s)?

   1. Fig. 2 c-h show examples of the learned (c, d), saturated (e, f), and non-learnable (g, h) group of conditions; Fig. 2a, b show examples of near and far recording sites marked in the lower-right panel of Fig 2e, f.

   These examples in Fig. 2 are simply used to illustrate how the raw data look like. They are followed by detailed and quantitative statistical analyses of similar conditions and recording sites in each group. The numbers of similar conditions and recording sites in each group are clearly indicated in the text (see the table above).

Statistics and general methods

1. Is there a justification of the sample size?
   If so, how was it justified?
   Where (section, paragraph #)?
   Even if no sample size calculation was performed, authors should report why the sample size is adequate to measure their effect size.

   In Results section, paragraphs 1-5, the numbers of recording sites, repetition trials, animals, training days, and stimulus conditions are explained.

   We used conventional tests (t-test, Wilcoxon rank sum test etc.) where appropriate, which are clearly described in the text.

   We used a permutation test for Fig 4d which is detailed in the Methods (“Calculating the onset time of contour modulation”, paragraph 10).

   The sample sizes can be deemed large enough for the t-test used for comparing two sample means. We confirmed this by running a non-parametric test (Wilcoxon rank sum test), which gave identical results.

   A detailed analysis of the variance of the contour versus non-contour stimulus condition is reported in Fig. 6a-c. The variance is almost constant across training days.

   Yes.

   N/A
3. Are criteria for excluding data points reported?
   Was this criterion established prior to data collection?
   Where is this described (section, paragraph #)?
   Yes. Electrodes were excluded that did not show a clear RF profile and orientation tuning. This method was established in previous experiments and thus known prior to data collection. (see Methods "Receptive fields mapping", paragraph 7).

4. Define the method of randomization used to assign subjects (or samples) to the experimental groups and to collect and process data.
   If no randomization was used, state so.
   Where does this appear (section, paragraph #)?
   We used two monkeys which performed identical tasks. All the stimulus conditions were randomized within a block of trials, as described in Results (paragraph 4)

5. Is a statement of the extent to which investigator knew the group allocation during the experiment and in assessing outcome included?
   If no blinding was done, state so.
   Where (section, paragraph #)?
   N/A

6. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included?
   Where (section, paragraph #)?
   Yes. See "Animal preparations" in Methods, paragraph 1

7. Is the species of the animals used reported?
   Where (section, paragraph #)?
   Yes. Macaca mulatta. Results (paragraph 1)

8. Is the strain of the animals (including background strains of KO/transgenic animals used) reported?
   Where (section, paragraph #)?
   N/A

9. Is the sex of the animals/subjects used reported?
   Where (section, paragraph #)?
   Yes. Male. Results (paragraph 1)

10. Is the age of the animals/subjects reported?
    Where (section, paragraph #)?
    It is reported in Results (paragraph 1) that adult monkeys were used. Their exact ages were 7 (MG) and 5 (MJ).

11. For animals housed in a vivarium, is the light/dark cycle reported?
    Where (section, paragraph #)?
    Not reported. But the animal room has automatic control of light cycle (6:00 am -8:00 pm)

12. For animals housed in a vivarium, is the housing group (i.e. number of animals per cage) reported?
    Where (section, paragraph #)?
    Not reported. One monkey was housed in one cage with enriched environment.

13. For behavioral experiments, is the time of day reported (e.g. light or dark cycle)?
    Where (section, paragraph #)?
    Not reported. All experiments were performed in the light cycle.
14. Is the previous history of the animals/subjects (e.g. prior drug administration, surgery, behavioral testing) reported?
   Where (section, paragraph #)?
   Yes. Naive animals were used. Results (paragraph 1). Animal preparations are reported in Methods paragraph 1 ("Animal preparations")

   a. If multiple behavioral tests were conducted in the same group of animals, is this reported?
      Where (section, paragraph #)?
      N/A

15. If any animals/subjects were excluded from analysis, is this reported?
   Where (section, paragraph #)?
   N/A

   a. How were the criteria for exclusion defined?
      Where is this described (section, paragraph #)?
      N/A

   b. Specify reasons for any discrepancy between the number of animals at the beginning and end of the study.
      Where is this described (section, paragraph #)?
      N/A

Reagents

1. Have antibodies been validated for use in the system under study (assay and species)?
   Where (section, paragraph #)?
   N/A

   a. Is antibody catalog number given?
      Where does this appear (section, paragraph #)?

   b. Where were the validation data reported (citation, supplementary information, Antibodypedia)?
      Where does this appear (section, paragraph #)?

2. If cell lines were used to reflect the properties of a particular tissue or disease state, is their source identified?
   Where (section, paragraph #)?
   N/A

   a. Were they recently authenticated?
      Where is this information reported (section, paragraph #)?
Data deposition

Data deposition in a public repository is mandatory for:
- a. Protein, DNA and RNA sequences
- b. Macromolecular structures
- c. Crystallographic data for small molecules
- d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available here. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad.

1. Are accession codes for deposit dates provided?
   Where (section, paragraph #)?
   N/A

Computer code/software

Any custom algorithm/software that is central to the methods must be supplied by the authors in a usable and readable form for readers at the time of publication. However, referees may ask for this information at any time during the review process.

1. Identify all custom software or scripts that were required to conduct the study and where in the procedures each was used.
   For all data analyses we used standard MatLab functions (e.g. svmtrain, classify, robustfit, corcoef, ranksum, ttest etc.), which are reported in the text where applicable.

2. Is computer source code/software provided with the paper or deposited in a public repository? Indicate in what form this is provided or how it can be obtained.
   No source code is provided.
   If necessary, MatLab source code for performing the analysis and generating the figures can be obtained from the authors.

Human subjects

1. Which IRB approved the protocol?
   Where is this stated (section, paragraph #)?
   N/A

2. Is demographic information on all subjects provided?
   Where (section, paragraph #)?
   N/A

3. Is the number of human subjects, their age and sex clearly defined?
   Where (section, paragraph #)?
   N/A

4. Are the inclusion and exclusion criteria (if any) clearly specified?
   Where (section, paragraph #)?
   N/A

5. How well were the groups matched?
   Where is this information described (section, paragraph #)?
   N/A
6. Is a statement included confirming that informed consent was obtained from all subjects?
   Where (section, paragraph #)?
   N/A

7. For publication of patient photos, is a statement included confirming that consent to publish was obtained?
   Where (section, paragraph #)?
   N/A

- fMRI studies

For papers reporting functional imaging (fMRI) results please ensure that these minimal reporting guidelines are met and that all this information is clearly provided in the methods:

1. Were any subjects scanned but then rejected for the analysis after the data was collected?
   N/A
   
   a. If yes, is the number rejected and reasons for rejection described?
      Where (section, paragraph #)?

2. Is the number of blocks, trials or experimental units per session and/or subjects specified?
   N/A
   Where (section, paragraph #)?

3. Is the length of each trial and interval between trials specified?
   N/A

4. Is a blocked, event-related, or mixed design being used? If applicable, please specify the block length or how the event-related or mixed design was optimized.
   N/A

5. Is the task design clearly described?
   Where (section, paragraph #)?
   N/A

6. How was behavioral performance measured?
   N/A

7. Is an ANOVA or factorial design being used?
   N/A

8. For data acquisition, is a whole brain scan used? If not, state area of acquisition.
   N/A
   
   a. How was this region determined?
9. Is the field strength (in Tesla) of the MRI system stated?
   - N/A

9a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?
   - N/A

9b. Are the field-of-view, matrix size, slice thickness, and TE/TR/flip angle clearly stated?
   - N/A

10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?
   - N/A

11. Is the coordinate space for the anatomical/functional imaging data clearly defined as subject/native space or standardized stereotaxic space, e.g., original Talairach, MNI305, ICBM152, etc? Where (section, paragraph #)?
   - N/A

12. If there was data normalization/standardization to a specific space template, are the type of transformation (linear vs. nonlinear) used and image types being transformed clearly described? Where (section, paragraph #)?
   - N/A

13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?
   - N/A

14. Were any additional regressors (behavioral covariates, motion etc) used?
   - N/A

15. Is the contrast construction clearly defined?
   - N/A

16. Is a mixed/random effects or fixed inference used?
   - N/A

16a. If fixed effects inference used, is this justified?
   - N/A

17. Were repeated measures used (multiple measurements per subject)?
   - N/A

17a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?
   - N/A

18. If the threshold used for inference and visualization in figures varies, is this clearly stated?
   - N/A

19. Are statistical inferences corrected for multiple comparisons?
   - N/A

19a. If not, is this labeled as uncorrected?
   - N/A
20. Are the results based on an ROI (region of interest) analysis?
   a. If so, is the rationale clearly described?
   b. How were the ROI’s defined (functional vs anatomical localization)?

21. Is there correction for multiple comparisons within each voxel?

22. For cluster-wise significance, is the cluster-defining threshold and the corrected significance level defined?

▶ Additional comments

Additional Comments