Robust selectivity to two-object images in human visual cortex

Yigal Agam[1], Hesheng Liu[4], Alexander Papanastassiou[6], Calin Buia[1], Alex Golby[5], Joseph R. Madsen[6] and Gabriel Kreiman[1,2,3,*]

[1] Department of Ophthalmology and Kirby Center for Neurobiology, Children's Hospital, Harvard Medical School

[2] Center for Brain Science Harvard University

[3] Swartz Center for Theoretical Neuroscience, Harvard University

[4] Massachusetts General Hospital

[5] Department of Neurosurgery, Brigham and Women's Hospital

[6] Department of Neurosurgery, Children's Hospital, Harvard Medical School

* To whom correspondence should be addressed

gabriel.kreiman@tch.harvard.edu

List of supplementary materials:

5 Supplementary Figures

1 Supplementary Table

See also additional material at:

http://klab.tch.harvard.edu/resources/objectpairs/index.htm

Figure S1: Responses of the electrode shown in Figure 1 to all category pairs and more examples of selective responses

A. For the same electrode illustrated in **Figure 1** in the main text (left fusiform gyrus), here we show the responses to each pair of object categories. Responses are aligned to stimulus onset (see scale bar in the top left corner). Entry (*i*,*j*) indicates the average responses obtained when the image contained category *i* and *j* (*i*,*j*=1,…,5). The matrix is symmetric. Along the diagonal (*i*=*j*), the entries indicate the average responses when the two objects belonged to the same category. For each category, we show one exemplar object (top and left) but the neural responses represent the average across all exemplar combinations and position combinations in the corresponding category pairs. The small number in each subplot indicates the total number of repetitions. The responses are colored according to the category in each row.

B-E. Responses of four more example electrodes using the same format used in **Figure 1** in the main text. These four electrodes come from three different subjects. Electrodes **B** (left inferior temporal cortex; Talairach coordinates = [-45.8 -3.2 -42.8]) and **C** (left fusiform gyrus; Talairach coordinates = [-49.6 -66.4 -12.6]) were in the same subject. Electrode **D** (left fusiform gyrus; Talairach coordinates = [-23.4 -36.5 -19.4]) was in the same subject as the one in **Figure 1** in the main text. Electrode **E** was in a different subject and was located in the left inferior temporal cortex (Talairach coordinates = [45.4 -36.5 -20.6]). On the right we depict the electrode locations (see Experimental Procedures and [1] for electrode localization). In the first 3 columns, the gray rectangle denotes the image presentation time (100 ms) and the dark gray horizontal line at the bottom shows the main analysis interval (50 to 300 ms after stimulus onset). Error bars denote one SEM. The number of repetitions is indicated in the upper right corner of each subplot. The color of each trace indicates the preferred category (see part **A**); the non-preferred category is shown in gray. The first column shows the responses to images containing single objects. The second column shows the responses to images containing two objects that belong to different categories. The third column compares the responses to the preferred category in isolation (thin line, copied from the first column) against the preferred category paired with another category (thick line, copied from the second

column). Note that there are always two lines in the third column but it is often hard to distinguish them in the initial ~200 ms (i.e., there is only a small amount of suppression in the initial response). The last column shows the distribution of the IFP response magnitudes (range from 50 to 300 ms) for the images containing two objects from different categories. The dashed curves show a Gaussian fit to the data (two parameters for each curve, mean and s.d.). We used the Akaike Information Criterion (AIC) to compare mixture-of-Gaussian fits to the response distributions. Only in one of the 35 fits was a 2-Gaussian fit better than the 1-Gaussian fit suggesting that the response distributions were unimodal.

Figure S2: Spatial summation models (expanding on Figure 2)

To quantify the degree of response suppression, let r_1 (R_1) indicate the response to exemplar 1 (category 1) when presented alone, r_2 (R_2) indicate the response to exemplar 2 (category 2) alone and r_{12} (R_{12}) indicate the response to an image containing both exemplar 1 and exemplar 2 (both category 1 and category 2).

A-B. These plots are derived from **Figures 1E** and **2D** in the main text except that here the responses were normalized by the maximum response (**A**: $\max(r_1,r_2)$; **B**: $\max(R_1,R_2)$) and the x-axis shows the sum of the responses normalized by the maximum response [2]. A single-electrode example is shown in **A** (same electrode as in **Figure 1E**) whereas all visually selective electrodes are shown in part **B**. The black dashed line shows the expected result according to a model predicting complete robustness to two-object images ($r_{12} = max(r_1,r_2)$ or $R_{12} = max(R_1,R_2)$). The blue line shows the expected results according to a model predicting that the response to a pair of objects is the average of the responses to the individual objects ($r_{12} = 0.5*(r_1+r_2)$ or $R_{12} = 0.5*(R_1+R_2)$). The blue circles in part **B**. indicate those cases where the preferred category was a human face.

C. Comparison of different ways of computing the suppression index (SI) (see **Figure 2** in the main text for the definition of the suppression index). The values in this figure indicate the mean±SD. We obtained similar suppression index values when considering only those electrodes that showed enhanced responses to human faces (blue). The suppression index values were also similar when the response was defined as the IFP signal range from 50 to 200 ms (instead of 50 to 300 ms as in the main text). We also

separately computed the suppression index for those images that contained the target category ("T", last two columns). All the other columns, as well as **Figure 2E** in the main text, only include trials where the target category was absent. For the exemplar-based analysis, we did not have enough trials to compute the suppression index for the target-present condition. The estimation of r_1, r_2 and r_{12} for the exemplar-based analyses are based on smaller numbers of trials than the corresponding estimates for the category-based analyses and, consequently, the estimation of the suppression index in the exemplar-based analyses shows a larger spread.

D-I. We considered 7 possible spatial summation models [2, 3] to fit the responses to images containing two objects from different categories (r_{12}) from the responses to the individual objects (r_1 and r_2): "max" ($\max(r_1,r_2)$; no free parameters); "average" ($0.5(r_1+r_2)$, no free parameters); "unscaled power" ($(r_1^n+r_2^n)^{1/n}$, 1 free parameter); "scaled linear" ($\alpha(r_1+r_2)$, 1 free parameter); " normalization" ($\alpha(r_1^{1/2}+r_2^{1/2})^2$, 1 free parameter); "scaled power" ($\alpha(r_1^n+r_2^n)^{1/n}$, 2 free parameters); "generalized linear" ($\alpha max(r_1,r_2)+\beta min(r_1,r_2)$, 2 free parameters). Note that essentially all of these models (except the last one) can be considered as variants of $\alpha(r_1^n+r_2^n)^{1/n}$ with parameters α, n: max=[1,∞], average=[0.5,1], unscaled power=[1,n], scaled linear=[α,1], normalization=[α,0.5], scaled power=[α,n]. We also considered models with twice as many parameters to take the position of the objects into consideration. Although the models that took into account the object position provided better fits to the data, the enhancement in the fit was too small to justify the additional parameters based on the Akaike model comparison criterion [4]. Here we show the results for the models that averaged the responses across object positions.

D-E. Two different example electrodes (electrodes from **Figure 1** and **S1C**) showing fits for each of these 7 possible model functions, $f(r_1,r_2)$, as a function of r_{12}. Each column shows a different model function and each row shows a different electrode. Each point corresponds to a separate image containing two exemplar objects from different categories. The error bars indicate one SEM for r_{12}. The black dashed line shows the diagonal ($f(r_1,r_2)=r_{12}$). The dotted blue line shows the linear fit between $f(r_1,r_2)$ and r_{12}.

Only images where we had more than 5 repetitions for each of r_1, r_2 and r_{12} were used for this analysis. The numbers in the bottom right of each subplot show the fit parameter.

F, H. Average normalized mean square error (Norm MSE). In **F**, the responses were defined by considering each exemplar separately ("exemplar-based" analysis, as in **D-E**). In **H**, the responses were defined by averaging over all exemplars within a category ("category-based" analysis). Error bars indicate one SEM. There are only 6 electrodes for which we could obtain a sufficient number of trials for the exemplar-based analysis (**F**). In contrast, the category-based analysis (**H**) considers all the visually selective electrodes (n=24).

G, I. Average parameters for each of the 5 functions that included free parameters. Error bars denote one SEM. The values are shown in separate axes because the parameter n for the unscaled/scaled power typically had very large values whereas the parameter α typically had low values.

Figure S3: Example of single-trial classification procedure for one electrode (related to Figure 3)

Five examples of *single-trial* responses for the electrode shown in **Figure 1**. The gray rectangle shows the stimulus presentation interval (100 ms). (**A**) One-object images including the preferred category; (**B**) One-object images excluding the preferred category; (**C**) Two-object images including the preferred category; (**D**) Two-object images excluding the preferred category. The gray horizontal bar (top) indicates the [50;300] ms interval used to extract three features: the range of the response (dotted lines), the time of minimum voltage (t_{min}, downward arrow) and the time of maximum voltage (t_{max}, upward arrow).

E-F. As illustrated in **A-D**, in each trial we extracted three parameters from the IFP response between 50 and 300 ms: the time of maximum IFP (t_{max}), the time of minimum IFP (t_{min}) and the IFP range (*max(IFP)-min(IFP)*). Several other ways of defining the response vector for each electrode were described previously in [1]. Here we show 2D plots for these 3 parameters for the example electrode shown in **Figure 1** for the images containing one object (**E**) or two objects (**F**). Each point represents a trial; blue points

denote trials where one of the objects belonged to the preferred category and gray points denote trials where none of the objects belonged to the preferred category.

G-H. These 3D plots show the relationships among the 3 feature variables used for the classification task for single-object images (**G**) and two-object images (**H**). The blue circles indicate trials where the preferred category for this electrode was present and the gray circles indicate trials where the preferred category for this electrode was not present. Note that this response vector is defined for each individual trial (there is no averaging of responses across trials). Training the classifier involves finding a plane to optimally separate the blue and gray circles. The classifier approach allows us to consider each electrode independently (as in this figure) or to examine the encoding of information by an ensemble of multiple electrodes (**Figures 3** and **S5**). When considering a set of N electrodes, we assumed independence across electrodes and concatenated the responses of the electrodes in the ensemble to build the ensemble response vector: $\left[t^1_{\min}, t^1_{\max}, R^1, ..., t^N_{\min}, t^N_{\max}, R^N\right]$. Unless otherwise stated, the results shown throughout the manuscript correspond to binary classification between a given category and the other object categories (see **Figure S5F** for multiclass classification). In a binary classifier, chance corresponds to 50% (indicated in the plots by a horizontal dashed line) and perfect performance corresponds to 100%. We previously argued that performance was significantly worse when attempting to identify specific exemplars [1]. Therefore, the classifier analyses are based on the category labels. We used a support vector machine (SVM) classifier with a linear kernel to learn the map between the ensemble response vectors and the labels [5, 6]. Importantly, in all cases, the data were divided into two non-overlapping sets, a training set and a test set. We examined different ways of separating the data into a training set and a test set (see **Figure S5** and text). The classification performance was evaluated by comparing the predictions on the test data with the actual labels. Throughout the text, we report the proportion of test repetitions correctly labeled as "Classification performance" (CP). When the classifier was trained on the single-object images, we obtained the gray plane depicted in the plot. Using this plane on the two-object images, the classification performance in this example was 71%. Not all trials are correctly classified: there are also gray circles to the "left" of the plane and blue circles to the "right" of the plane. In shorthand notation, in the text we sometimes refer to

"training a classifier with image X" and "testing the classifier with image Y". What this means is: training a classifier with the response vectors and labels obtained from those trials when image X was presented and testing the classifier performance with the response vectors and labels obtained from those trials when image Y is presented. To assess the statistical significance of the CP values, we computed the distribution of CP values in 100 iterations where we randomly shuffled the object labels in each trial. The mean performance under the null hypothesis was 50% and the standard deviation was 2.3%. We considered CP to be significant if performance was more than 3 standard deviations above the null hypothesis.

Figure S4: Decoding visual information in single trials from the neural responses to two-object images (expanding on Figure 3)

A. For each visually selective electrode, we built a statistical classifier to read out category information in single trials (**Figure S3** and Experimental Procedures). We defined the responses of an electrode as "robust to two-object images" if the classification performance was more than 3 standard deviations beyond the chance level (50%) based on 100 shuffles of the object labels. Here we show the distribution of category preferences for those electrodes that showed robustness to two-object images. As reported previously (e.g. [1, 7]), there were more electrodes that showed a preference for human faces than for any of the other object categories that we tested. Yet, we also observed selectivity [1] and robustness to two-object images in electrodes that showed preferences for other object categories (see also the examples in **Figure S1B,D**). The lack of selective responses to animals in this case is likely to be due to small numbers of trials and electrodes because in an earlier study we did observe selectivity to animals [1].

B. For each of the visually selective electrodes, the x-axis indicates the classification performance in single-object images ($CP_{selectivity}$). To compute $CP_{selectivity}$ the classifier was trained on 70% of the single-object image repetitions and the classification performance was evaluated in the remaining 30% of the single-object image repetitions. The y-axis indicates the classification performance in two-object images ($_{single}CP_{2\text{-}object}$). $_{single}CP_{2\text{-}object}$ is based on training the classifier with the responses to single-object images and evaluating the classification performance using the two-

object images. Most of the points are below the diagonal line (dashed line) indicating that $_{single}CP_{2\text{-}object} < CP_{selectivity}$. The dotted line is the linear fit (slope=0.89, Pearson correlation coefficient=0.91).

C. Proportion of electrodes in each location that showed visually selective responses (Experimental Procedures). We only show those locations where there was at least one selective electrode (see **Table S1** for the entire list of recording locations). The location abbreviations are: Fsup (superior frontal cortex), MT (medial temporal lobe), ITC (inferior temporal cortex), Occ-Inf (inferior occipital gyrus), LTFus (fusiform gyrus), MTPhip (parahippocampal gyrus). The region names are derived from a human brain parcellation as described in the references in [1] (Experimental Procedures; **Table S1**). It would be of significant interest to know what the homologous locations for these regions are in the macaque brain; yet this is a not an easy question to address. We avoid direct comparisons with the macaque brain because functional homology in visual cortex between the two species remains only poorly understood.

D. Proportion of the selective electrodes that showed robustness to two-object images in each location. There were two electrodes (one in MT and one in FSup) that showed selectivity only in the presence of single-object images. Due to the very small number, we refrain from drawing strong conclusions about this observation.

Figure S5: Different ways of training and testing the classifier yielded approximately similar classification performance values (related to Figure 3)

A. There are several different ways of separating the data in this experiment into a training set and a test set. Unless otherwise stated in the text, the classifier was trained using the neural responses to single-object images and the classification performance was evaluated using the neural responses to two-object images ("$_{single}CP_{2\text{-}object}$", top). Here we illustrate different ways of separating the training and test sets with sample example images when a binary classifier is trained to recognize the "car" category. In the bottom half of the plot, the classifier is trained using the neural responses to two-object images and the classification performance is evaluated using the responses to two-object images (but always ensuring that there is no overlap between the training and test data to avoid overfitting). We

distinguish 3 different variations. "$_{all}$CP$_{2\text{-}object}$" refers to pooling all the responses to two-object images and randomly choosing 70% of the data to train the classifier and the remaining 30% of the data to evaluate the classification performance. "$_{cat}$CP$_{2\text{-}object}$" refers to a variation where the second, non-preferred, object in the two-object images belonged to different categories in the train and test sets. For example, when training the classifier to recognize cars, the $_{cat}$CP$_{2\text{-}object}$ classifier could be trained with *car+animal* pairs and *car+chair* pairs and the classification performance would be evaluated with *car+face* pairs and *car+house* pairs. "$_{eg}CP_{2\text{-}object}$" refers to a variation where the second, non-preferred, object in the two-object images was a different exemplar in the train and test sets but it could still belong to the same category. For example, when training the classifier to recognize cars, the $_{eg}CP_{2\text{-}object}$ classifier could be trained with *car1+house1* images (among other images) and the classifier performance could be tested with *car2+house2* images (among other images). In all cases, the data were subsampled so that the number of positive and negative examples was the same (ensuring that the chance level is 50%).

B-D. Comparison of $_{single}$CP$_{2\text{-}object}$, $_{cat}CP_{2\text{-}object}$ and $_{eg}CP_{2\text{-}object}$ against $_{all}CP_{2\text{-}object}$ for single electrodes. The black dashed line is the identity line. The red dashed line shows the linear fit to the data (**B**: slope=0.98, Pearson correlation = 0.96; **C**: slope=0.92, Pearson correlation = 0.85; **D**: slope = 0.98, Pearson correlation = 0.98; *n*=35 points, 24 electrodes). Error bars denote one SEM.

E. Comparison of $_{single}$CP$_{2\text{-}object}$, $_{cat}CP_{2\text{-}object}$ and $_{eg}CP_{2\text{-}object}$ against $_{all}CP_{2\text{-}object}$ for the pseudopopulation ensemble (as in **Figure 3B**). The bar colors indicate the object categories (see **Figure S1**).

F. Results of multiclass classification using the pseudopopulation ensemble (as in **Figure 3B**). Throughout the text and in other figures we used a binary classifier to distinguish a category from other categories (the chance level is 50% for the binary classifier). Here we use a multiclass classifier that addresses the question "Which category was present?" (see [1, 5] for further comparisons of binary classifiers and multiclass classifiers). In the $CP_{selectivity}$ case, there were five possible categories and therefore the chance level is 20%. In the other cases where the classifier's performance was evaluated with images containing two objects from different

categories, there were $\binom{5}{2} = 10$ possible category pairs and the chance level is 40% (each category is present in 4 out of the 10 possible category pairs).

Table S1. Distribution of electrode locations

Electrodes were localized by combining pre-surgical MR images with post-surgical CT images (Experimental Procedures). Here we show the number of electrodes in each location. For each electrode, we determined the Talairach coordinates (Experimental Procedures). We show the average coordinates for all the electrodes in each region. The Talairach coordinates are separately reported for those electrodes in the left hemisphere (L) and in the right hemisphere (R). This is provided only as a coarse indication of the locations and care should be taken when interpreting these coordinates because it is not clear that locations in Talairach space can be averaged. Talairach coordinates for each individual electrode are available upon request. The "location code" and "location" refer to one of 80 brain regions based on the human brain parcellation in ref. [8].

References

1. Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron *62*, 281-290.
2. Zoccolan, D., Cox, D.D., and DiCarlo, J.J. (2005). Multiple object response normalization in monkey inferotemporal cortex. J Neurosci *25*, 8150-8164.
3. Ghose, G.M., and Maunsell, J.H. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. J Neurosci *28*, 5115-5126.
4. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control *19*, 716-723.
5. Hung, C., Kreiman, G., Poggio, T., and DiCarlo, J. (2005). Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. Science *310*, 863-866.
6. Vapnik, V. (1995). The Nature of Statistical Learning Theory, (New York: Springer).
7. McCarthy, G., Puce, A., Belger, A., and Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. Cerebral Cortex *9*, 431-444.
8. Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968-980.

Code	Location	R hem	L hem	n	<Talairach> [R]	<Talairach> [L]
6	G-frontal-inf-Opercular-part	14	5	19	[58.9 17.1 10.9]	[-4.3 46.4 53.2]
7	G-frontal-inf-Orbital-part	3	3	6	[52.4 36.5 -3.3]	[-42.1 43.4 -8.0]
8	G-frontal-inf-Triangular-part	13	9	22	[55.3 30.2 5.4]	[-48.8 39.1 4.0]
9	G-frontal-middle	12	8	20	[41.8 46.4 18.0]	[-39.5 48.7 30.4]
11	G-frontomarginal	4	7	11	[27.8 63.8 -8.5]	[-23.9 65.8 -10.0]
14	G-and-S-occipital-inferior	12	4	16	[39.8 -84.8 -10.0]	[-46.2 -53.2 4.8]
15	G-occipital-middle	5	0	5	[49.9 -74.0 35.7]	
16	G-occipital-superior	4	0	4	[22.9 -81.7 43.3]	
17	G-occipit-temp-lat-Or-fusiform	14	17	31	[39.1 -55.9 -13.8]	[-34.9 -41.8 -19.8]
18	G-occipit-temp-med-Lingual-part	5	1	6	[12.4 -78.1 -6.1]	[-10.7 -56.3 -8.6]
19	G-occipit-temp-med-Parahippocampal-	14	14	28	[22.4 -11.5 -27.2]	[-24.8 -9.4 -21.7]
20	G-orbital	28	22	50	[32.7 33.6 -15.7]	[-37.7 45.6 -12.0]
22	G-parietal-inferior-Angular-part	14	3	17	[52.9 -65.5 33.1]	[-42.3 -21.0 60.3]
23	G-parietal-inferior-Supramarginal-part	11	7	18	[66.9 -21.7 23.2]	[-58.5 -27.3 28.7]
24	G-parietal-superior	6	4	10	[13.4 -71.9 53.0]	[-29.2 -27.5 66.0]
25	G-postcentral	4	13	17	[65.7 -6.5 29.2]	[-55.6 -3.5 41.0]
26	G-precentral	11	11	22	[63.3 5.1 23.1]	[-52.9 14.9 37.1]
28	G-rectus	9	0	9	[-1.2 30.4 -20.2]	
30	G-subcentral	12	7	19	[65.9 -3.0 15.3]	[-59.1 -0.6 12.4]
31	G-temporal-inferior	63	33	96	[57.6 -34.7 -19.8]	[-52.3 -31.0 -22.2]
32	G-temporal-middle	51	31	82	[65.0 -29.5 -6.6]	[-60.1 -25.3 -6.6]
34	G-temp-sup-Lateral-aspect	50	26	76	[64.8 -12.0 0.6]	[-58.8 -5.4 3.5]
36	G-temp-sup-Planum-tempolare	4	0	4	[67.6 -42.6 21.2]	
37	G-and-S-transverse-frontopolar	5	0	5	[18.4 69.0 8.5]	
42	Pole-occipital	4	0	4	[13.1 -91.1 -14.7]	
43	Pole-temporal	41	19	60	[34.3 10.7 -33.6]	[-40.0 5.1 -34.7]
52	S-collateral-transverse-ant	1	1	2	[23.4 -79.1 -5.7]	[-27.0 -1.0 -21.0]
53	S-collateral-transverse-post	1	0	1	[23.6 -68.0 -4.0]	
60	S-occipital-anterior	4	0	4	[48.9 -81.8 1.3]	
65	S-orbital-H-shapped	2	0	2	[29.9 37.1 -14.0]	
66	S-orbital-lateral	2	0	2	[44.1 12.7 -2.9]	
79	S-temporal-inferior	0	4	4		[-62.5 -27.2 10.0]
	Total	**423**	**249**	**672**		

Figure S1

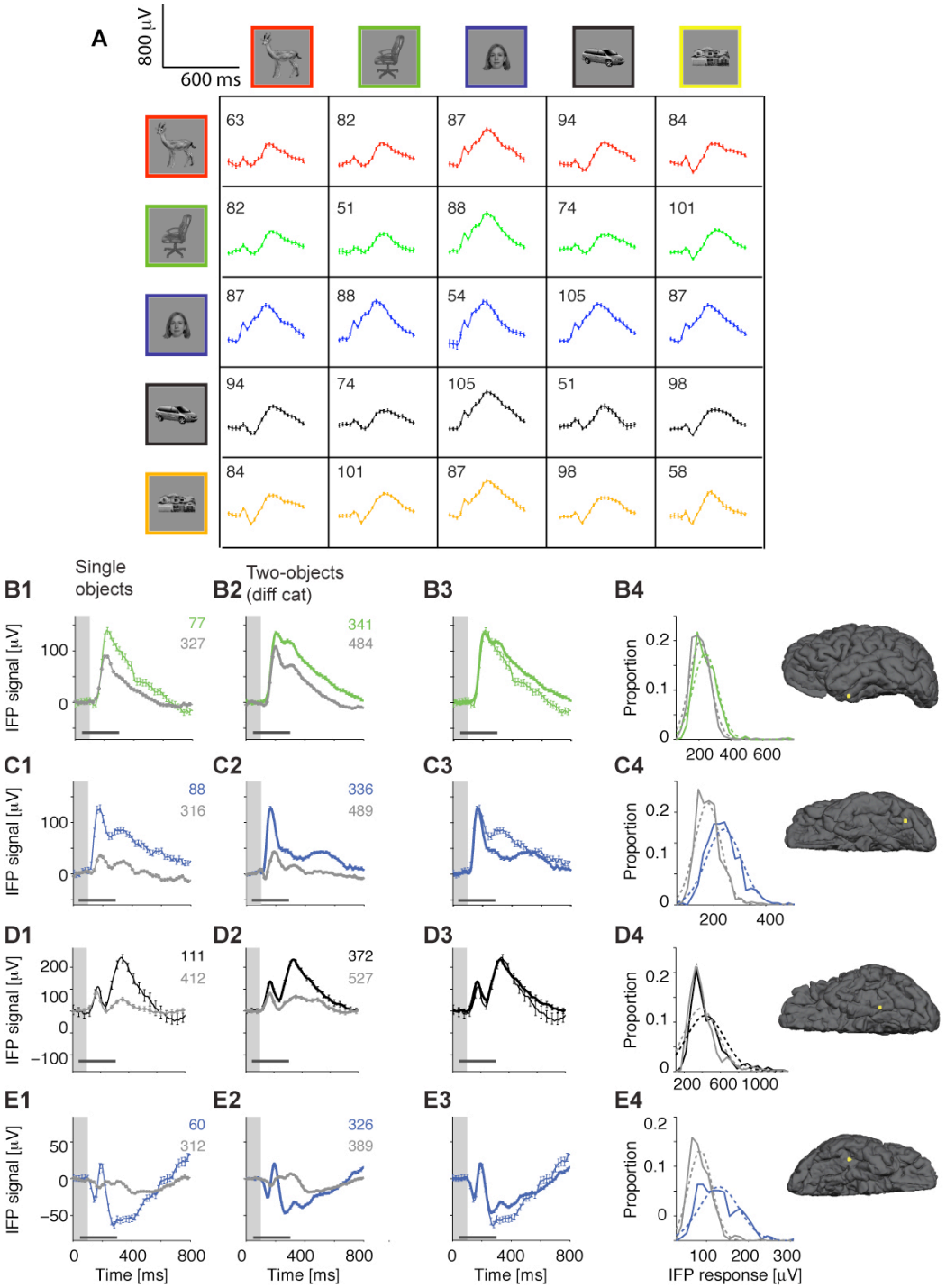

A
800 µV
600 ms
63 82 87 94 84
82 51 88 74 101
87 88 54 105 87
94 74 105 51 98
84 101 87 98 58
B1
Single objects
B2
Two-objects (diff cat)
B3
B4
77
327
341
484
IFP signal [µV]
Proportion
C1
C2
C3
C4
88
316
336
489
D1
D2
D3
D4
111
412
372
527
E1
E2
E3
E4
60
312
326
389
Time [ms]
IFP response [µV]

Figure S2

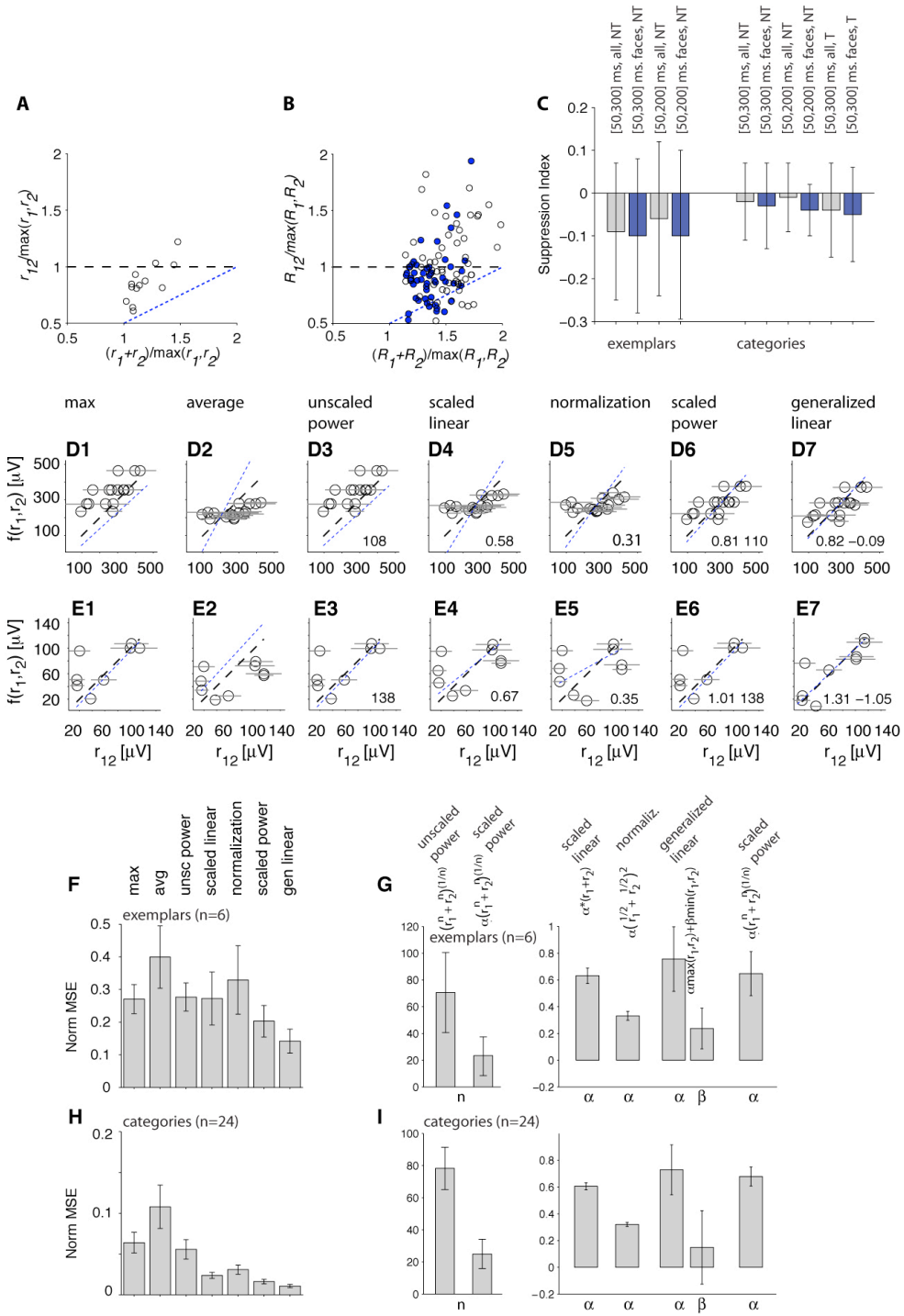

A
B
C
Suppression Index
exemplars
categories
max
average
unscaled power
scaled linear
normalization
scaled power
generalized linear
D1
D2
D3
D4
D5
D6
D7
E1
E2
E3
E4
E5
E6
E7
F
G
H
I
Norm MSE
exemplars (n=6)
categories (n=24)

Figure S3

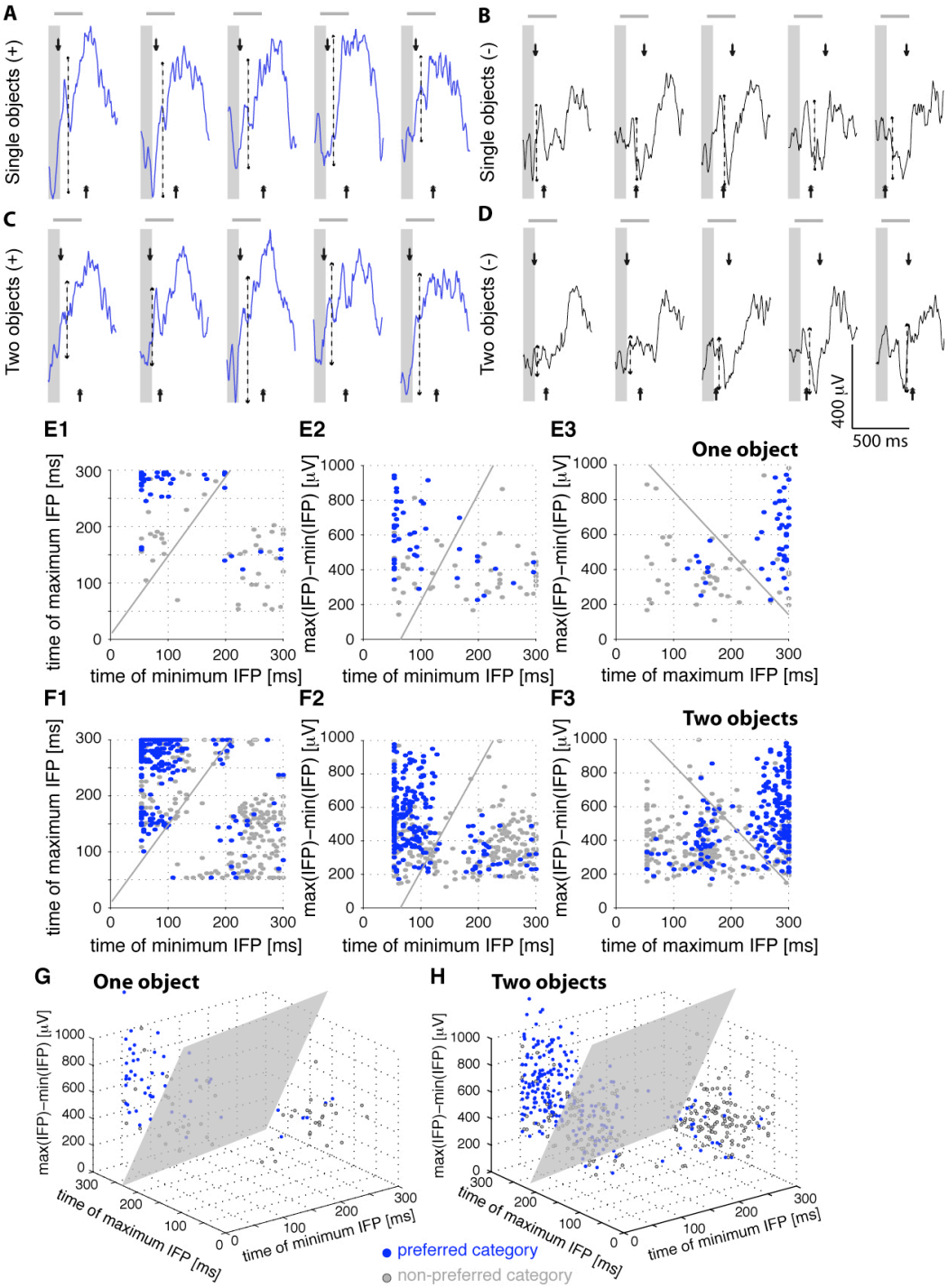

A
Single objects (+)
B
Single objects (-)
C
Two objects (+)
D
Two objects (-)
400 µV
500 ms
E1
E2
E3
One object
time of maximum IFP [ms]
time of minimum IFP [ms]
max(IFP)−min(IFP) [µV]
F1
F2
F3
Two objects
G
One object
H
Two objects
preferred category
non-preferred category

Figure S4

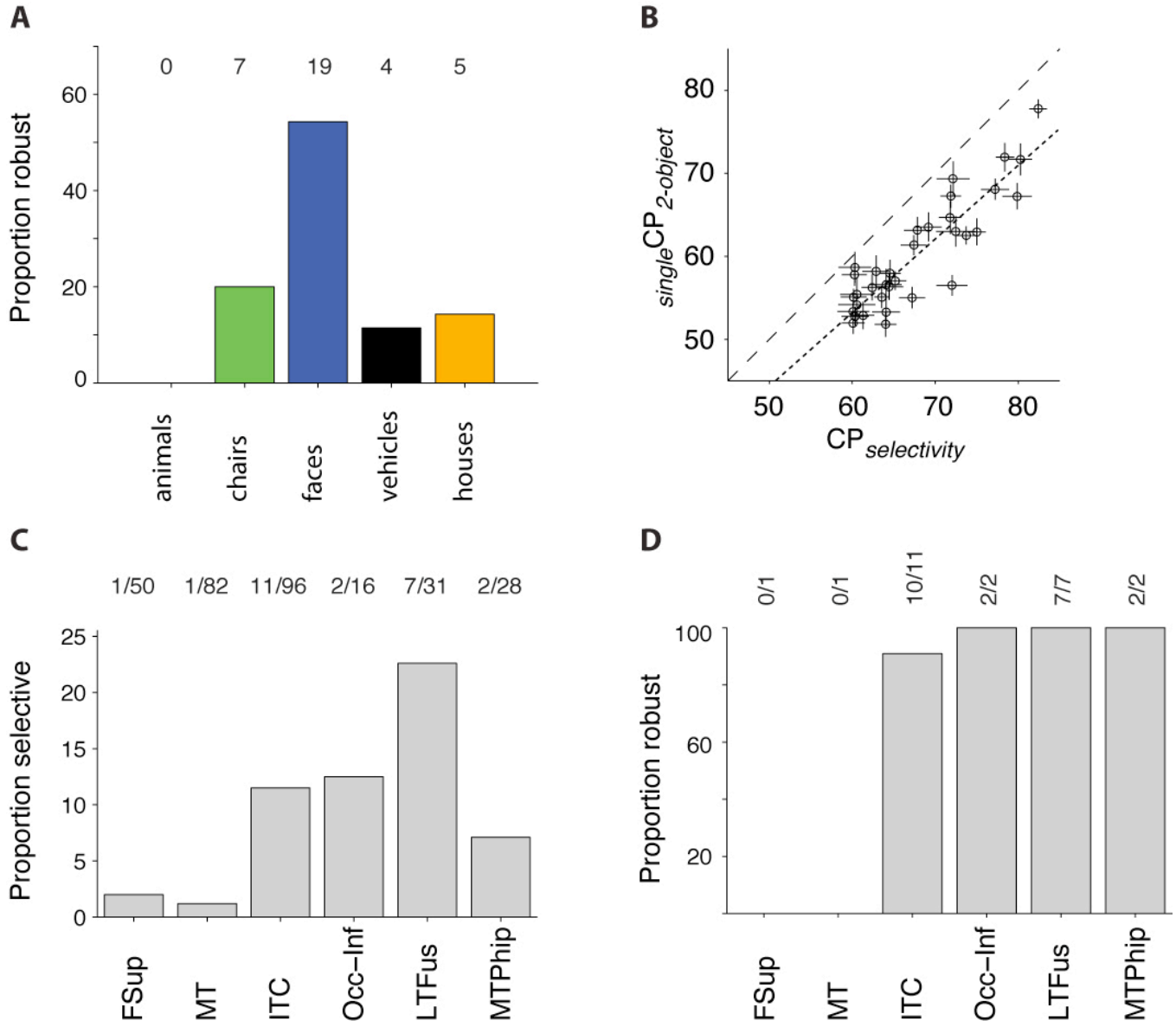
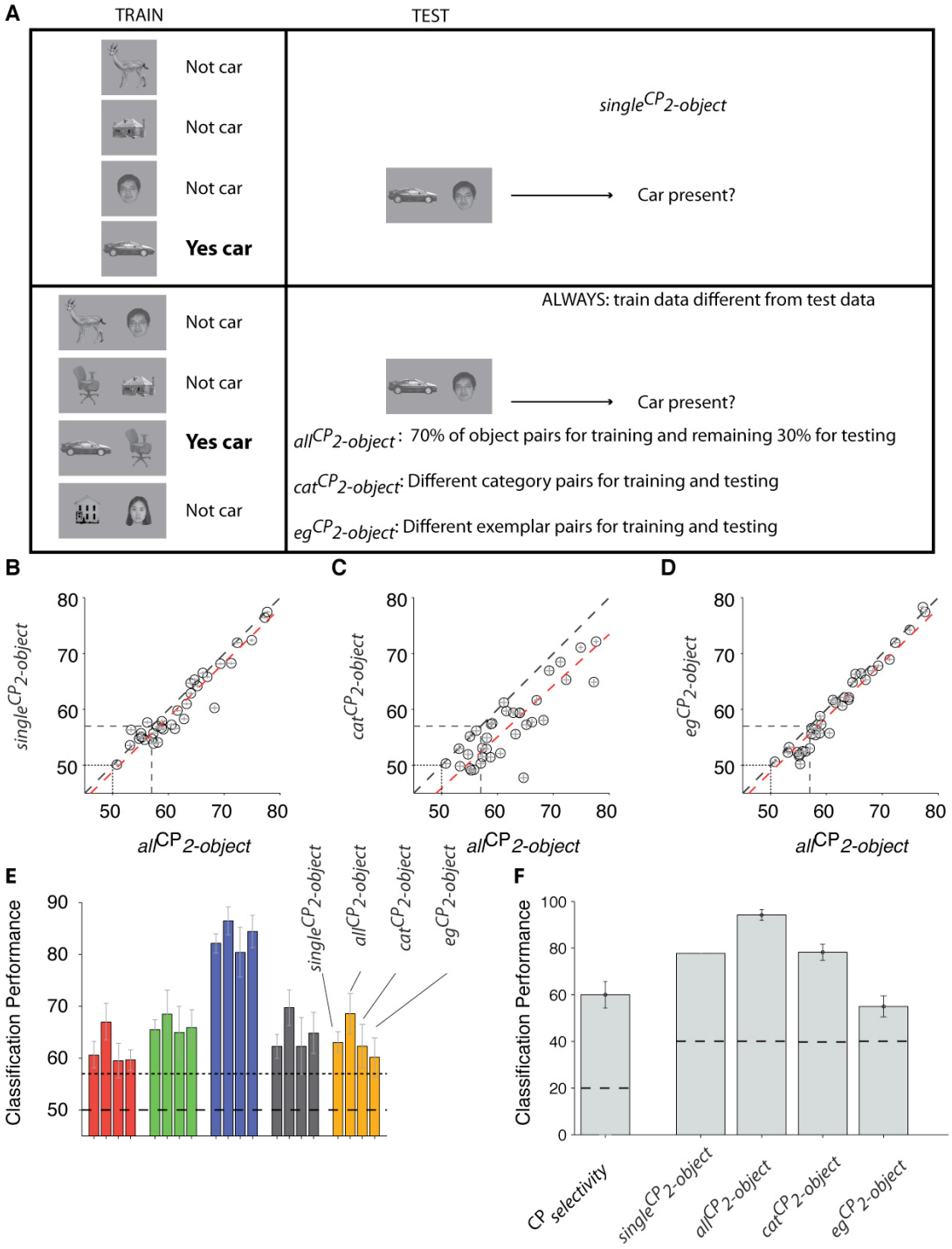

A
0
7
19
4
5
Proportion robust
60
40
20
0
animals
chairs
faces
vehicles
houses
B
80
70
60
50
single CP 2-object
50
60
70
80
CP selectivity
C
1/50
1/82
11/96
2/16
7/31
2/28
Proportion selective
25
20
15
10
5
0
FSup
MT
ITC
Occ–Inf
LTFus
MTPhip
D
0/1
0/1
10/11
2/2
7/7
2/2
Proportion robust
100
60
20
FSup
MT
ITC
Occ–Inf
LTFus
MTPhip

Figure S5


A
TRAIN
TEST
Not car
Not car
Not car
Yes car
single^CP 2-object
Car present?
Not car
Not car
Yes car
Not car
ALWAYS: train data different from test data
Car present?
all^CP 2-object : 70% of object pairs for training and remaining 30% for testing
cat^CP 2-object: Different category pairs for training and testing
eg^CP 2-object: Different exemplar pairs for training and testing
B
C
D
single^CP 2-object
cat^CP 2-object
eg^CP 2-object
all^CP 2-object
E
Classification Performance
F
CP selectivity