

Supplementary Materials for

A shared neural substrate for action verbs and observed actions in human posterior parietal cortex

T. Aflalo*, C. Y. Zhang, E. R. Rosario, N. Pouratian, G. A. Orban, R. A. Andersen

*Corresponding author. Email: taflalo@caltech.edu

Published 23 October 2020, *Sci. Adv.* **6**, eabb3984 (2020)
DOI: 10.1126/sciadv.abb3984

This PDF file includes:

Supplementary Text
Figs. S1 to S11

Supplementary Materials

Supplementary Text: Preliminary action observation experiment (Fig S2).

Experimental design:

Experimental stimuli consisted of video-clips (448x336 pixels, 50 frames/s) showing an actor, sitting at a table, viewed from the side at a distance of 1.2 m, performing seven different hand actions directed towards objects. The action videos used for this preliminary experiment are the same as (18). Each of these seven actions (drag, drop, grasp, push, roll, rotate, squeeze) was performed by a male or female, using two natural objects (a plum and a mandarin), yielding four versions of each action exemplar. The objects lay on a table or in the hand, with the arms on the table until the actions began, hence actions predominantly involved the wrist and fingers. We tried to equate videos across the actions where possible: the same two actors, wearing the same clothing performed the actions without expressing any particular emotion. Efforts were also made to equalize the scenes in which the action took place. Lighting and background were identical as well as the general organization of the scene. The actor stood on the right and kept his/her hands in the middle when performing the action. All videos measured 17.7° by 13.2° and lasted 2.6 s (the first 2 and the last 2 frames being static). Before the presentation of each video, a blurred static frame of the average all videos was shown as baseline. The edges of the videos were blurred with an elliptical mask ($14.3^\circ \times 9.6^\circ$), leaving the actor and the background of the video unchanged, but blending it gradually and smoothly into the background around the edges. A 0.2° fixation target was shown in all conditions. The fixation target was presented as close as possible to the position where the movement took place in the video, either above or below the midpoint of the trajectory of the manipulating hand. The movie stimuli are the same as those used for (18), however, presentation of stimuli was modified to use the averaged blurred static frame as baseline (described above) instead of a uniform background. Each data session consisted of 10 repetitions of each unique video for a total of 560 videos per session (2 actors x 2 objects x 2 fixation locations x 7 actions x 10 repetitions). Presentation was split into 3 runs (3 + 3 + 4 repetitions). Videos were presented in a pseudorandom manner: all conditions were randomly ordered and presented once before repetition. The subjects were instructed to attend to but otherwise passively view the videos.

Analysis:

In the preliminary experiment we manipulated 4 experimental variables that included action exemplar, fixation location, gender, and object. To establish whether these variables impacted neural responses we performed a sliding window ANOVA analyses. For each time window, we constructed an ANOVA model to explain firing rate for each unit using all factors (action, gender, object, fixation position) and their interactions. The resulting p-value for each factor and each time window was *fdr* corrected to determine significance. Window onsets started from -1.5 seconds relative to video onset with the final window chosen to be +3 seconds. Windows were stepped at 100 ms intervals. Each window included the average neural response within a 750 ms window.

Supplementary Figures

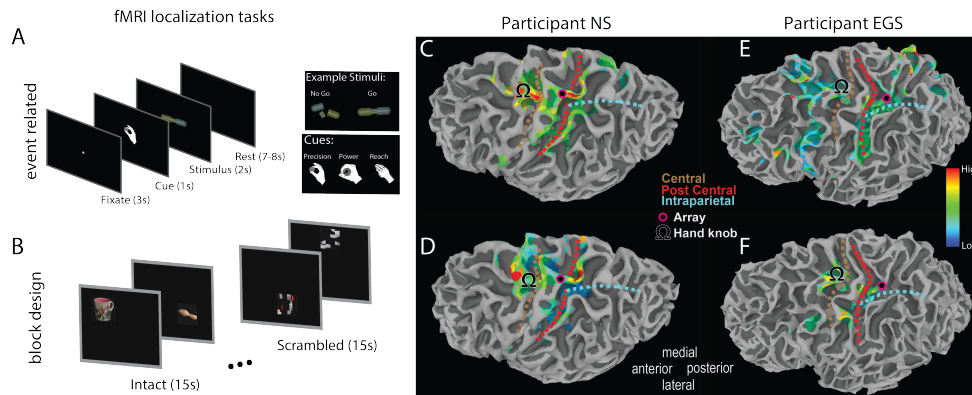


Figure S1: Functional magnetic resonance imaging (fMRI) localization of implant sites. fMRI was used to identify cortical regions that were suitable for implantation based on the BOLD response for imagined reaching and grasping actions. We performed complementary tasks to ensure robust activation across a range of paradigms and protocol types. **(A)** Event related task design. Following an intertrial interval, the subject was cued to perform a specific imagined movement (precision grasp, power grasp, or reach without hand shaping.) Following the cue, a cylindrical object was presented. If the object was intact, the subject performed the cued imagined movement. If the object was broken, the subject withheld movement. **(B)** Block task design. Eight blocks were presented for 30 seconds per run. During the first 15 seconds common objects were presented with a new image presented every three seconds at a new spatial location. The subject was instructed to either imagine pointing at, imagine reaching and grasping, or look at the object in a natural manner prior to each run. During the last 15 seconds of each block, scrambled images were presented and the subject was instructed to guess the identity of the object. **(C)** Statistical parametric map showing voxels with significant activity in the grasp “Go” versus “No-Go” condition ($p < 0.01$ fdr corrected.) Array location and cortical landmarks as depicted in the legend. **(D)** Statistical parametric map showing voxels with significant activation ($p < 0.01$ fdr corrected) for the grasping versus looking condition in the block design. **(E,F)** Same as C, D but for participant EGS.

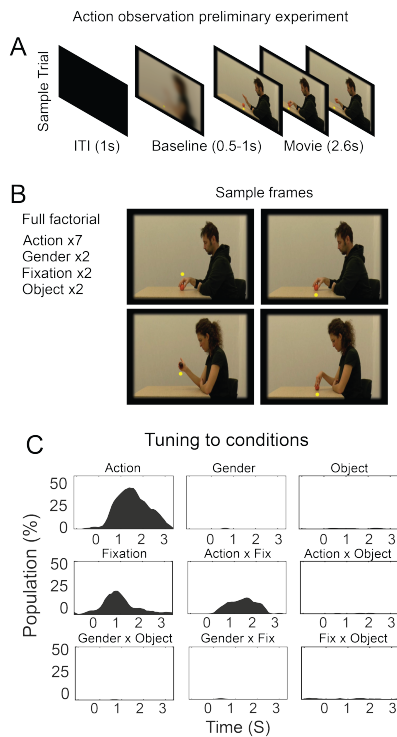


Figure S2: Preliminary study; impact of multiple factors on neural responses to observed manipulative actions. In a preliminary study we measured the impact of several conditions on neural responses from a single lateral perspective (similar to L0). **(A)** Sample trial illustrating temporal progression through first experiment. **(B)** Full factorial design used for testing neural selectivity for observed manipulative actions (OMAs). Each subject passively viewed pre-recorded videos of seven different manipulative action exemplars (dropping, dragging, pushing, rolling, rotating, grasping, and squeezing) while varying actor gender (male, female), object (orange, plum), and fixation location (above, below). Sample frames illustrate stimuli. Photo credit: Guy Orban, Department of Medicine and Surgery, Parma University. **(C)** The percent of recorded units with significant responses to the factors of interest (ANOVA $p < 0.05$, fdr corrected) through time. In addition to observed action, fixation location has a sizeable and significant impact on the neural population. Data represents the pooled population of units recorded from participant NS (3 sessions, 290 total units) and EGS (3 sessions, 111 total units).

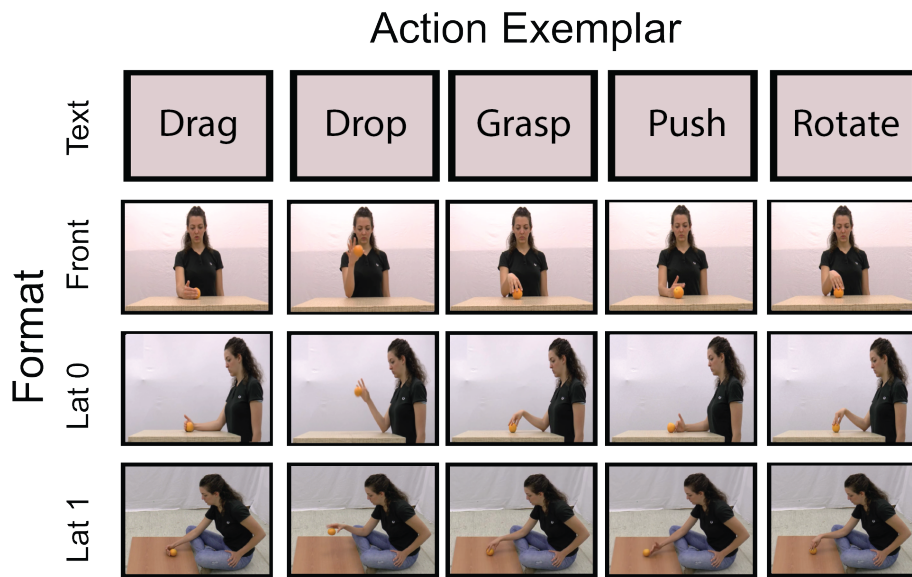


Figure S3: Representation of task stimuli. Still shots depicting stimuli used for the 5 action exemplars (drag, drop, grasp, push, and rotate) by 4 formats (text, frontal, and two lateral perspectives) experimental design. Photo credit: Guy Orban, Department of Medicine and Surgery, Parma University.

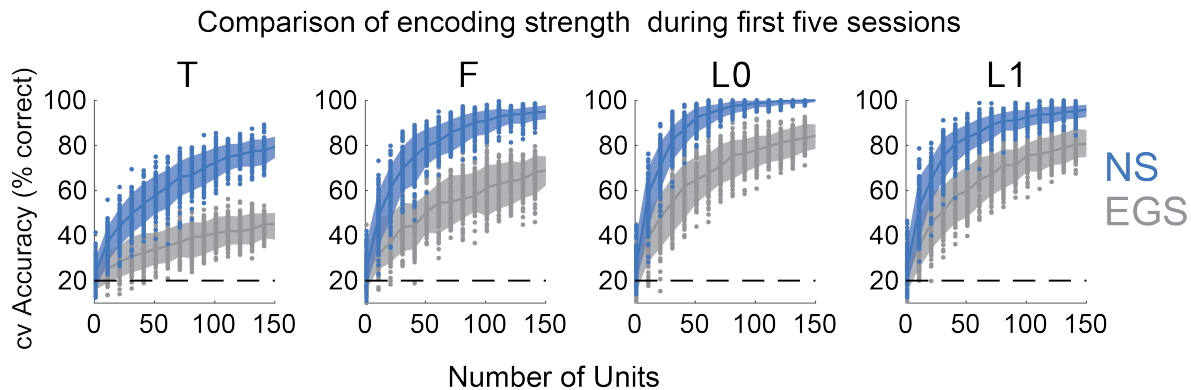


Figure S4: Neuron dropping curves to compare strength of selectivity in participants NS and EGS.

Neuron dropping curves show how cross-validated classification accuracy depends on the number of neurons included in the classification. To better enable a direct comparison between subjects, we restricted data from NS to the first five sessions, the number of sessions recorded in EGS. NS's neural population is more strongly selective, especially for action verbs, even when accounting for the number of units and sessions. This result is consistent with the smaller proportion of selective units in EGS (Fig 1B). Results shown as the mean and 95th percentile classification accuracy over 200 random samples of N neurons (x-axis) split by participant (blue=NS, grey=EGS) and format (T=text, F = front, L0=lateral standing, L1 = lateral sitting). Individual markers represent cross-validated classification accuracy for a single random sample of N neurons. Horizontal dashed line indicates chance level.

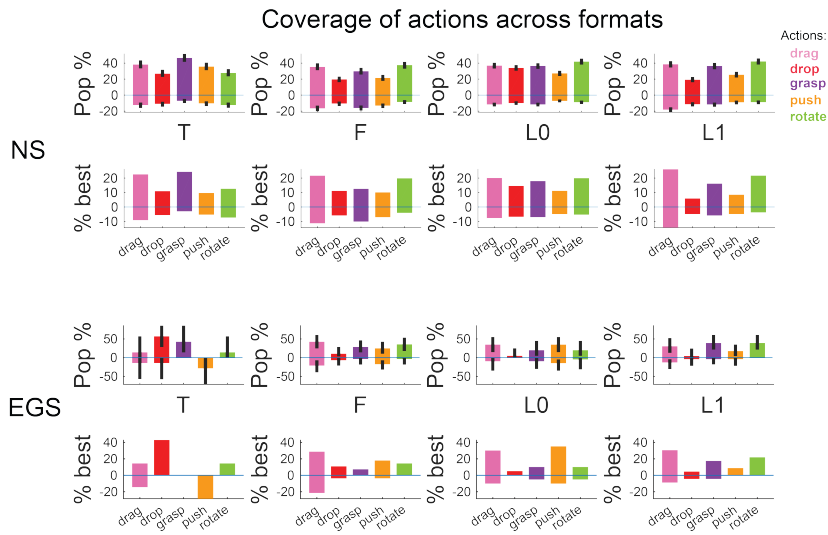


Figure S5: Percent of selective units for each action for each of the four formats. Results are shown individually for each subject. *Top:* Percent of units with significant modulation to each action type during presentation relative to baseline activity (paired t-test, $p < 0.05$ fdr corrected, mean \pm 95% CI). *Bottom:* Percent of units whose highest magnitude response belonged to each action type. Since roughly 1/3 of all recorded neurons showed purely suppressive responses (all actions suppressed activity below baseline), analysis is split between excitatory and purely suppressive units. Suppressive units shown as “negative” percentages. Color code for each action provided in inset. Formats shown in labeled columns (T=text, F = front, L0=lateral standing, L1 = lateral sitting).

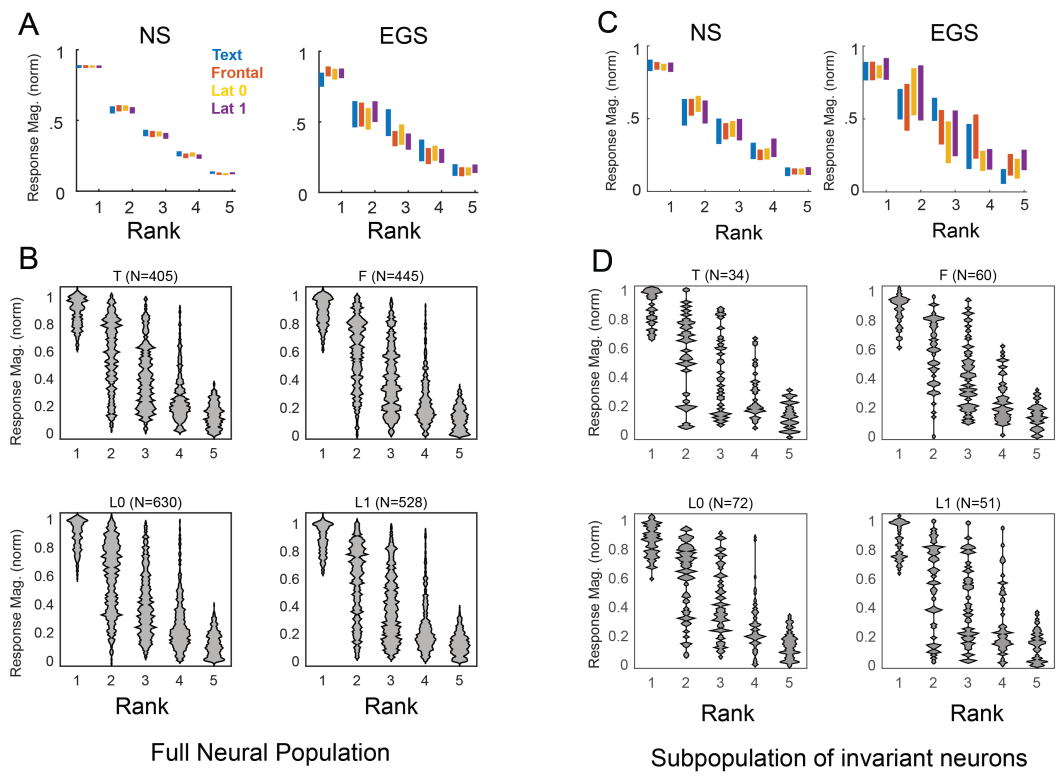


Figure S6: Shape of tuning functions for each format. (A) 95% confidence interval for normalized, cross-validated rank ordered responses for each of the four formats (see color code) and both subjects. Responses show a graded drop-off from the best action (defined as resulting in the largest magnitude response) second best action, etc. **(B)** Grey violin plot shows the probability distribution of cross-validated rank-ordered responses for subject NS. **(C)** Same as A, except for the subpopulation of cells that demonstrated invariance across all formats (invariance defined as $T=F=L0=L1$, Fig. 4D). **(D)** Same as B, except for the subpopulation of invariant neurons.

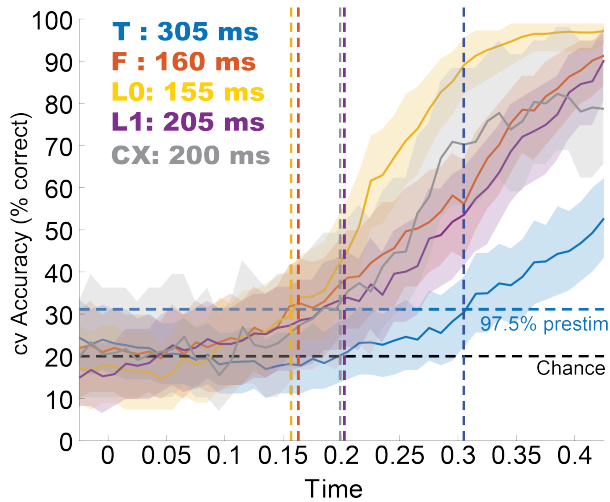


Figure S7: Latency of population encoding for all formats and control study in participant NS. A sliding window cross-validated classification analysis (5 ms step size, 25 ms full-width at half maximum smoothing) was used to track population encoding at a fine temporal scale to compare latency of selectivity across formats. Results are shown for the four formats for the main experiment (T=text, F = front, L0=lateral standing, L1 = lateral sitting) as well as the imagery condition of the supplementary experiment (CX=control experiment, see Fig S10). Latency was computed for the control study to enable a direct comparison between cognitive processing related to imagery and semantic processing. Black horizontal dashed line indicates chance performance. Blue horizontal dashed line indicates 97.5th percentile of prestimulus classification accuracy for the text condition. Vertical dashed lines provide visual representation of neural response latency provided numerically in legend.

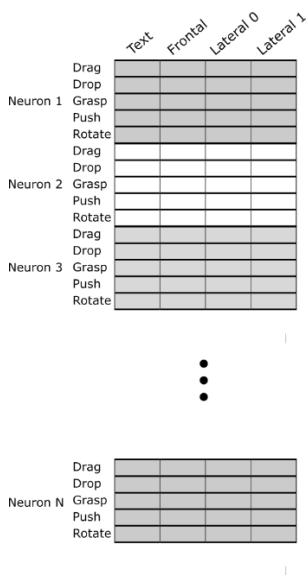


Figure S8: Organization of data for population correlation method. In order to compute a similarity measure between formats the data was organized into matrix form such that the responses for the same unit and same action identity were positioned at corresponding rows across formats (as shown in the figure). The correlation measure was computed across columns in order to provide a quantitative measure of how similar the pattern of responses for action exemplars was across formats.

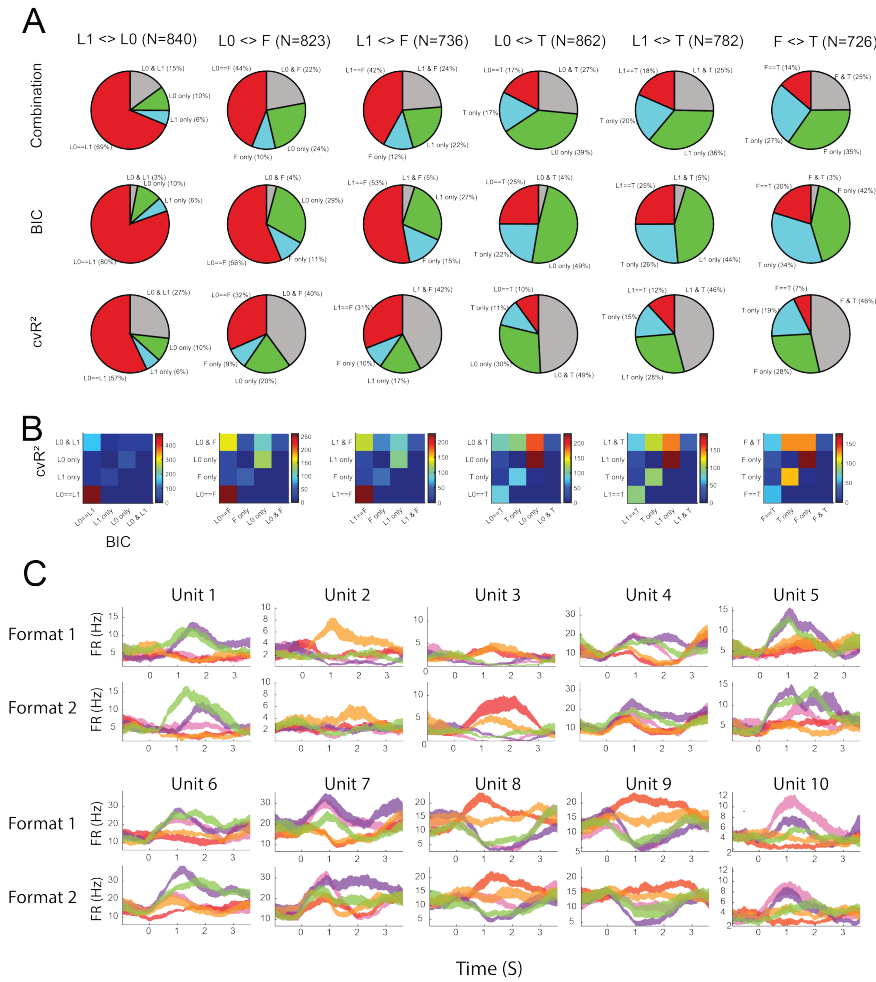


Figure S9: Comparison of model selection procedures. In main figure 3 B&C we averaged the results of using Bayesian information criteria (BIC) and cross-validated coefficient of determination (cvR^2) as the criteria to identify the four selectivity patterns across formats (red, grey, green and cyan correspond to the four panels shown in Fig 3A). **(A)** The results are extended to include all possible pairwise comparisons of formats (top row: Combination) as well as results using BIC alone (middle row: BIC) and cvR^2 alone (bottom row: cvR^2). Results are organized by the comparison metric (rows) and the formats used in the comparison (columns). The combination measure is the arithmetic mean between BIC and cvR^2 . Only action selective units were used in the model selection analyses (total number of selective units for the comparison shown as $N=xxx$ in the title). Selective units were defined as units with significant coding for action identity in a 2x5 format by action ANOVA with interactions (p -value < 0.05 fd_r corrected, either main effect or interaction). The results are shown for subject NS. Results for EGS were similar. **(B)** A “confusion matrix” summarizing how units categorized using BIC were categorized using cvR^2 across the population. For example, >400 units categorized as $L0=L1$ using BIC were also categorized as $L0=L1$ using cvR^2 while ~ 200 were categorized as $L0\&L1$ using cvR^2 (left most panel). This pattern is consistent with each matrix having an upper-left triangular structure demonstrating that cvR^2 has a greater tendency to pick models with distinct parameters across formats (e.g. $L0\&L1$). **(C)** Instructive example neurons illustrating how BIC and cvR^2 capture different notions of similarity across formats. We show the event-related firing rate (mean \pm sem) as a function of time for 10 example units across a pair of formats (different rows). For each example unit, the BIC measure labeled the pairwise comparison as having a matched pattern of selectivity while the cvR^2 measure indicated significant but mismatched patterns of selectivity. The BIC measure captures a more qualitative notion of similarity between some (but not necessarily all) of the tested actions.

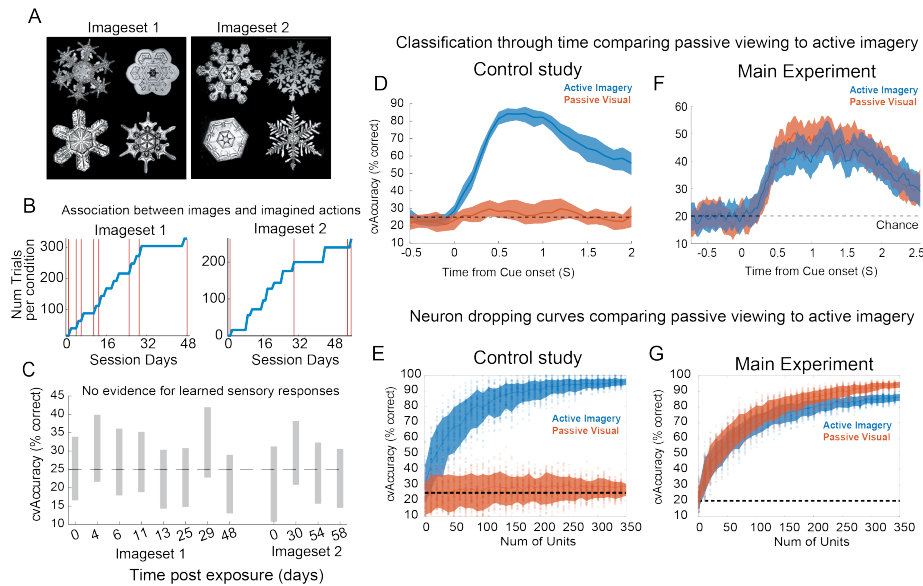


Figure S10: Unlike action verbs, abstract symbols systematically associated with motor imagery do not evoke differential neural selectivity during passive viewing. (A) Abstract geometric stimuli systematically associated with visual imagery of motor actions. (B) Cumulative count of trials for associative training during which symbols were used to cue imagery of motor actions (blue) and time of passive viewing of the same images (red lines). (C) Cross-validated classification (mean \pm 95% CI) on a per-session basis for passive viewing (dashed line indicates chance level). (D) Comparison of cross-validated classification accuracy (mean \pm 95% CI) during imagery (blue) and passive viewing (orange) of abstract symbols in the control experiment. Horizontal dashed line indicates chance level. (E) Neuron-dropping curves quantifying cross-validated accuracy as a function of number of neurons during imagery (blue) and passive viewing (orange) of abstract symbols in the control experiment. Horizontal dashed line indicates chance level. (F) Comparison of cross-validated classification accuracy (mean \pm 95% CI) during imagery (blue, last seven sessions) and passive viewing (orange, first 6 sessions) of action verbs in the main experiment. Horizontal dashed line indicates chance level. (G) Neuron-dropping curves quantifying cross-validated accuracy as a function of number of neurons during imagery (blue) and passive viewing (orange) of action verbs in the main experiment. Horizontal dashed line indicates chance level.

| Model # | Format 1 | Format 2 | Format 3 | Format 4 (possible values) |
|---------|----------|----------|----------|------------------------------|
| 1 | α | α | α | (α, γ, c) |
| 2 | α | α | γ | $(\alpha, \gamma, \beta, c)$ |
| 3 | α | α | c | (α, γ, c) |
| 4 | γ | α | α | $(\alpha, \gamma, \beta, c)$ |
| 5 | c | α | α | (α, γ, c) |
| 6 | α | γ | α | $(\alpha, \gamma, \beta, c)$ |
| 7 | α | c | α | (α, γ, c) |
| 6 | α | c | c | (α, γ, c) |
| 9 | c | α | c | (α, γ, c) |
| 10 | c | c | α | (α, γ, c) |
| 11 | α | γ | c | $(\alpha, \gamma, \beta, c)$ |
| 12 | c | α | γ | $(\alpha, \gamma, \beta, c)$ |
| 13 | α | c | γ | $(\alpha, \gamma, \beta, c)$ |
| 14 | α | γ | β | $(\alpha, \gamma, \beta, c)$ |
| 15 | c | c | c | (α, c) |

Figure S11. Complete list of the possible ways selectivity patterns can compare across 3 & 4 formats. This chart provides a visual aid for the methods section “Comparing selectivity patterns between formats” as relates to analysis presented in figure 4. In brief, the chart provides a complete enumeration of the possible linear mathematical models that describe how selectivity patterns can compare across 3 and 4 formats. Bold Greek letters indicate the vector of coefficient values describing firing rate as a function of action exemplar within a format. A constant c indicates no selectivity to action exemplar for the associated format (e.g. a single scalar offset is used to describe the response of all exemplars within a particular format). Use of the same Greek letter for multiple formats indicates identical coefficient values and thus describes neurons with matched selectivity patterns across formats. Use of different Greek letters indicates distinct coefficient values and thus significant but idiosyncratic selectivity patterns across formats.