COGNITIVE NEUROSCIENCE

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

T. Aflalo[1,2]*, C. Y. Zhang[1,2], E. R. Rosario[3], N. Pouratian[4], G. A. Orban[5], R. A. Andersen[1,2]

High-level sensory and motor cortical areas are activated when processing the meaning of language, but it is unknown whether, and how, words share a neural substrate with corresponding sensorimotor representations. We recorded from single neurons in human posterior parietal cortex (PPC) while participants viewed action verbs and corresponding action videos from multiple views. We find that PPC neurons exhibit a common neural substrate for action verbs and observed actions. Further, videos were encoded with mixtures of invariant and idiosyncratic responses across views. Action verbs elicited selective responses from a fraction of these invariant and idiosyncratic neurons, without preference, thus associating with a statistical sampling of the diverse sensory representations related to the corresponding action concept. Controls indicated that the results are not the product of visual imagery or arbitrary learned associations. Our results suggest that language may activate the consolidated visual experience of the reader.

INTRODUCTION

How do words get their meaning? Although the exact architecture of the semantic system is still under debate, most evidence suggests that meaning emerges from interactions between supramodal association regions that code abstracted symbolic representations and the distributed network of regions that process higher-level aspects of sensory stimuli, motor intentions, valence, and internal body state (1–5). Engagement of the distributed network is taken as evidence that the brain's representation of the physical manifestation of words is an important component of their meaning. For example, visual coding for the form of a banana, the motor act of biting into or peeling a banana, and its taste and texture would be components of meaning in addition to more symbolic, lexical aspects of meaning such as the "dictionary definition." Although this view is generally accepted, no single-unit recording evidence has demonstrated a shared neural substrate between processing the meaning of a word and its visuomotor attributes within the distributed network. To date, supporting evidence comes from lesion and functional magnetic resonance imaging (fMRI) studies establishing a rough spatial correspondence between brain areas involved in high-level sensorimotor processing and areas recruited when reading text or performing other behaviors that require access to meaning (1, 6). A lack of direct neural evidence is concerning given that neuroimaging and lesion results have been mixed and cannot establish a shared neural substrate at the level of single neurons (7, 8). Thus, how words get their meaning translates into two immediate questions with regard to single-neuron selectivity: (i) Are words and their sensorimotor representations coded within the same region of cortex? (ii) Is there a link between words and their sensorimotor representations? In this paper, linking will refer to the existence of a shared neural substrate with individual neurons exhibiting matching selectivity for both a word and the corresponding visual reality.

To complicate matters, the number of sensorimotor representations that can be described by the same basic concrete word is generally very large (e.g., the visual form of a "banana" depends on ripeness, viewing angle, lighting, and whether it is peeled or sliced), and invariance is very rarely complete in high-level sensorimotor regions [e.g., (9, 10)]. This raises a third question: If the same object is coded in different ways depending on details of presentation, how might a word link to these varied visual representations? Stated more generally: What is the neural architecture that links neuronal responses to silently reading a word and seeing varied visual presentations of what the word signifies? The answer is critical in understanding how sensorimotor representations influence our understanding of words. Do we connect the symbolic representation of a word to an abstracted invariant and, therefore, universal visual representation? To a particular canonical example? Or to the many diverse representations that comprise our varied experiences? The question applies to all concrete words that describe physical reality, including action verbs. In this study, we look at how neural coding for action verbs relates to varied visual representations of corresponding observed actions.

Last, what cognitive phenomena can account for the presence of a link between a word and its visual representation within any experimental paradigm? The link may mediate semantic memory, reflecting associations between the word and its visual representations built over a lifetime of experience. In this view, reading words activates sensorimotor representations automatically, and these representations are an intrinsic component of the meaning of the word. Second, reading a word has been hypothesized to evoke mental imagery. Responses in sensorimotor cortex may reflect such imagery, and the link could be between visual representations and mental imagery of the same stimuli, or the link may be the consequence of short-term learning such as occurs during categorization (11). Given these multiple possibilities, we address a fourth question: If a link exists, what cognitive process does the link mediate?

To address the above four questions, we recorded populations of neurons from electrode arrays implanted in two tetraplegic individuals (N.S. and E.G.S.) participating in a brain-machine interface clinical trial while the participants viewed videos of manipulative actions or silently read corresponding action verbs. The implants were placed at the anterior portion of the intraparietal sulcus (IPS; see fig. S1 for

[1]California Institute of Technology, Division of Biology and Biological Engineering, Pasadena, CA, USA. [2]Tianqiao and Chrissy Chen Brain-Machine Interface Center, California Institute of Technology, Pasadena, CA, USA. [3]Casa Colina Hospital and Centers for Healthcare, Pomona, CA, USA. [4]Department of Neurosurgery, University of California Los Angeles, Los Angeles, CA, USA. [5]Department of Medicine and Surgery, University of Parma, Parma, Italy.
*Corresponding author. Email: taflalo@caltech.edu

implant locations), a region that is part of the "action observation network" (AON) composed of the lateral occipital temporal cortex [LOTC; (12)], as well as frontal and parietal motor planning circuits (13, 14). These regions are involved in higher-order processing of observed actions (15–18), and neuroimaging and lesion evidence implicate a role in verb processing (19–25). The ability to perform invasive neural recordings provides us with the first opportunity to probe whether and how language links with corresponding visual representations at the level of single neurons in high-order sensory-motor cortex. Toward this objective, we establish four primary results relating to the four questions outlined above: First, PPC neurons show selectivity for action words and visually observed actions; second, a portion of PPC neurons link action verbs and corresponding visual representations; third, text-selective units in PPC link with all the diverse visual representations found in the neural population; and fourth, the link is not based on imagery or short-term learning and thus appears to be semantic in nature. One possible interpretation is that when reading text, we replay our visual history as part of the process of understanding and thus ground our conceptual understanding in our unique experiences.

RESULTS

Participants viewed videos of five manipulative actions presented in three visual formats (two lateral views differing in body posture and one frontal view) and a fourth format, text, requiring the subject to silently read associated action verbs (see Fig. 1A for example stimuli). Five actions were used: drag, drop, grasp, push, and rotate, for which preliminary experiments (fig. S2) had demonstrated neuronal selectivity. A total of 15 unique videos (5 distinct exemplar actions × 3 visual formats) and 5 written action verbs were presented for a total of 20 experimental conditions (5 actions × 4 formats; fig. S3). Presenting the observed actions in three formats allowed us to tease apart different models of how action verbs associate with overlapping (common to visual formats/exhibiting invariance across all formats) and distinct (idiosyncratic to given formats/not invariant or only invariant across subset of formats) features of the neural code for observed actions. This design allowed us to answer the first three questions posed in Introduction. We recorded 1586 units during 18 recording sessions in two subjects (NS: 1432 units, 13 sessions; EGS: 154 units, 5 sessions). For the first seven sessions in participant NS and all sessions for subject EGS, the participants passively watched the action videos and silently read the action verbs. To answer the fourth question, for the final six sessions in subject NS, the participant used the action verb as a prompt to "replay" the associated action video using visual imagery from either frontal (F) or lateral (L0) perspectives, thus allowing us to quantify how imagery affects verb processing. Results from silent reading (first seven sessions) and active imagery (last six sessions) were quantitatively similar in NS, and thus, data were pooled across sessions when addressing the first three questions of this paper. In addition, for question 4, we present a control study in which abstract symbols are paired with visual imagery of motor actions to better understand the effects of short-term associations.

Are human posterior parietal cortex (PPC) neurons selective for observed actions and action verbs?

Figure 1A shows the response of five representative neurons illustrating the variety of selectivity for both observed actions and action
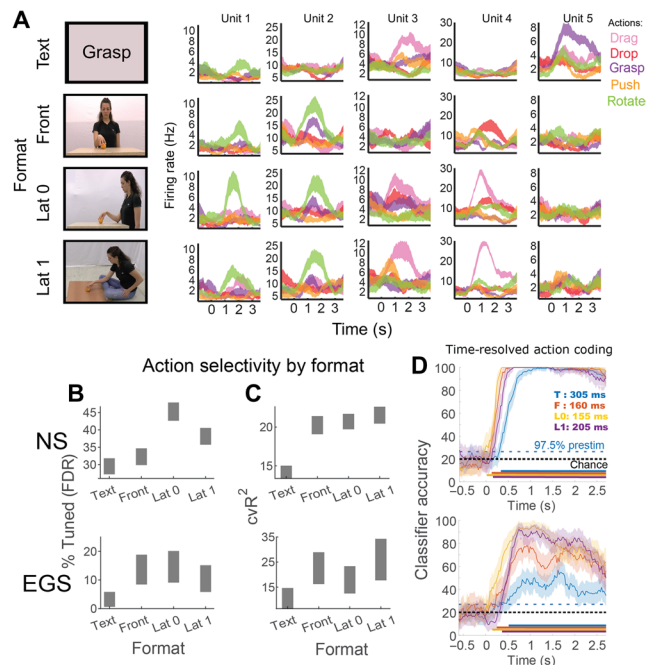

Fig. 1. Human parietal neurons are selective for observed actions and action verbs. (**A**) Example neurons illustrating diverse selectivity patterns (SPs) across formats. Left: Sample still frames depicting stimuli for one of the five action exemplars ("grasp") in each format (see fig. S3 for all action exemplars). Right: Representative units illustrating diverse neural responses to the five tested actions (color-coded) across the four tested formats. Each panel shows the firing rate (means ± SEM) through time for each action for a single format. Each column illustrates the responses of the same unit to the four formats. See fig. S1 for recording locations. Photo credit: Guy Orban, Department of Medicine and Surgery, Parma University. (**B**) Percentage of units with significant action selectivity split by format [means ± 95% confidence interval (CI), one-way ANOVA, $P < 0.05$ FDR-corrected]. Zero units were selective in each format during the 1-s window before stimulus onset (one-way ANOVA, $P < 0.05$ FDR-corrected). (**C**) Cross-validated R^2 of units with significant selectivity [units significant in (B)] split by format (means ± 95% CI). (**D**) Sliding-window within-format classification accuracy for manipulative actions. Sliding window = overlapping 300-ms windows with 10-ms increments. Classification applied to data pooled across sessions. Black horizontal dashed line = chance classification performance. Blue horizontal dashed line = 97.5th percentile of prestimulus classification accuracy for the text condition. Horizontal colored bars indicate time of significant classification. Inset displays color code for format and associated latency estimate for onset of significant decoding (see fig. S7).

verbs at the level of individual neurons. Within a format, we defined units as selective if there were significant differences in neural responses to the five actions (ANOVA, $P < 0.05$ False discovery rate corrected), to the different action identities. The percentage of cells demonstrating selective responses was significant for each format, for both subjects [χ^2 for text format, the format with the fewest selective units: NS: (1,$N = 1432$) = 503, $P < 0.001$; EGS: (1,$N = 154$) = 5.3, $P = 0.02$]. However, the percentage of selective units, as well as the consistency of the response, as measured by the cross-validated coefficient of determination (cvR^2), was smaller for text than for observed actions (Fig. 1, B and C). In addition, population classification analysis equating experimental sessions and number of units confirmed greater selectivity for participant NS than participant EGS (fig. S4). All five actions evoked significant neural responses from baseline across the four formats (fig. S5). The majority of visually

selective units were increased firing during the video presentations, as in nonhuman primate anterior intraparietal area (AIP) (18). A minority, however, were suppressed by the video and text presentations (fig. S5). The mean response strength decreased smoothly from the action evoking the maximal response to the weakest response. Individual units could show steep or more graded selectivity, and this pattern was essentially identical across formats (fig. S6). Greater selectivity for action videos relative to text was reflected in a time-resolved decode analyses (Fig. 1D). Defining the latency of action selectivity as the onset of significant classification accuracy revealed shorter latencies for the visual formats (windows starting at 155 to 205 ms depending on format) than the written word (305 ms), possibly reflecting differences in afferent pathways (fig. S7). Our results show that, all formats were encoded within the population but with greater selectivity and shorter latency for videos relative to text.

Is there a link between neural representations of action verbs and observed actions in human PPC?

Having established that PPC neurons are selective for both action verbs and observed actions, we now ask whether there exists a shared neural substrate, with neurons exhibiting matching selectivity for both a word and the corresponding visual representation. We addressed this by using two population analyses: across-format classification and across-format correlation. Leave-one-out cross-validation was used to train a classifier to predict action identity within format. On each fold, the decoder was also used to predict action identity from the three additional formats. This across-format generalization analysis measures how well the neural population structure that defines action identity in one format generalizes to other formats (Fig. 2, A and B). As a control, the same values can be computed when shuffling action identity between formats [shuffled accuracy; red in Fig. 2 (A and B)]. Across-format accuracy was both above chance and shuffled accuracy for all pairs of formats for NS, for all visual pairs of EGS, and the text-visual format pairs when pooling across visual formats to achieve adequate power for EGS (rank-sum test, $P < 0.05$). This result demonstrates that the neuronal representation was not random; the population is more likely to link representations across formats for the same action identities. However, the results also demonstrate that the generalization is not perfect: The across-format accuracy is lower than the within-format accuracy, suggesting that the neural code for action identity also depends on details of presentation. The strength of generalization was format dependent being near perfect across body postures (same lateral view), still high, but reduced across shifts in viewing perspective (across the lateral and frontal views), and lowest when comparing observed actions with the written verb.

Significant generalization of action representations across formats was robust to the analysis technique. We correlated neural population responses across formats (Fig. 2, C and D). Population responses were constructed by concatenating the mean response of all units to each action within format (fig. S8). A significant positive correlation was found for all format pairs while no significant positive correlation was found when shuffling action identity between formats. One caveat to interpretation is that the correlation between any pair of formats may be the consequence of the two formats being correlated with a third format. A significant link between pairs of formats was preserved but somewhat reduced when controlling for the other formats using a partial correlation analysis (Fig. 2, E and F). This last result indicates that text links with each of the visual formats directly as the significant link is preserved when the possible mediating factors of the other formats are removed.

The preceding population analyses established that text and visual representations are linked pairwise at the level of the population, but the link does not perfectly generalize across formats. What is the breakdown of the single units that compose the population results? To answer this question, we compared the precise selectivity pattern (SP; defined as the firing rate values for each of the five actions) across pairs of formats using a model selection analysis for each neuron. A linear tuning model can describe the four possible ways that the SP can compare across two formats (Fig. 3A). (i) Both formats are selective in a similar manner (Fig. 3A; matched selectivity); the linear parameters ($\boldsymbol{\alpha}\epsilon R^5$) for each of the five actions are constrained to be identical for the two formats. (ii) Both formats are selective but with mismatched patterns (Fig. 3A; mismatched selectivity); the linear parameters ($\boldsymbol{\alpha},\boldsymbol{\gamma}\epsilon R^5$) are different between the two formats. (iii and iv) Last, only one of the two formats may be selective (Fig. 3A; single format 1 or format 2 selective); a constant scalar offset term is used for the nonselective format (scalar term not shown in equation for simplicity). We identified the model that best described the neuronal behavior using both the Bayesian information criteria (BIC) and the cvR^2. We found that the two measures provide complementary perspectives when comparing across formats (fig. S9). In summarizing the results, we used the average percentages provided by both measures. In line with our population results, we found that the percentage of cells with a similar SP across formats (Fig. 3, B and C, red) was format dependent, being greatest across body postures (same lateral view), slightly reduced across shifts in viewing perspective (across the lateral and frontal views), and lowest when comparing observed actions with the written verb. These results indicate not only that text links with the visual formats and the visual formats link with each other but also that a percentage of the population codes the same action identities in different formats with differing patterns of selectivity.

What is the architecture that links observed actions and action verbs?

The preceding section demonstrated that there is a neural link between action verbs and visually observed actions. Here, we seek to understand the architecture of this link: to characterize how text-selective units link with the varied visual presentations of the same action. As a prerequisite, we first characterized how the different visual presentations were encoded with respect to each other, ignoring the text format. Just as neural SPs can compare across two formats (Fig. 3A), they can compare across three formats in 14 possible ways (see Fig. 4A, x-axis labels and examples). As above, a model selection analysis was used to categorize each unit based on the model that best described the SPs across the visual formats (Fig. 4A). The population was heterogeneous, characterized by units with matched SPs and mismatched SPs in varied combinations across the different visual formats. This diversity can be seen in the individual unit examples of Fig. 1A; units 1 and 2 show matching patterns of selectivity across all the visual formats (Fig. 4A, L0=L1=F), unit 3 shows matching selectivity across two of the visual formats and no selectivity in the third (Fig. 4A, L0=L1), and unit 4 shows matching selectivity between two formats and mismatching selectivity in the third (L0=L1&F). Thus, we find that presentation details affect neural coding for action identity and that individual units link action identity across formats in an assortment of ways
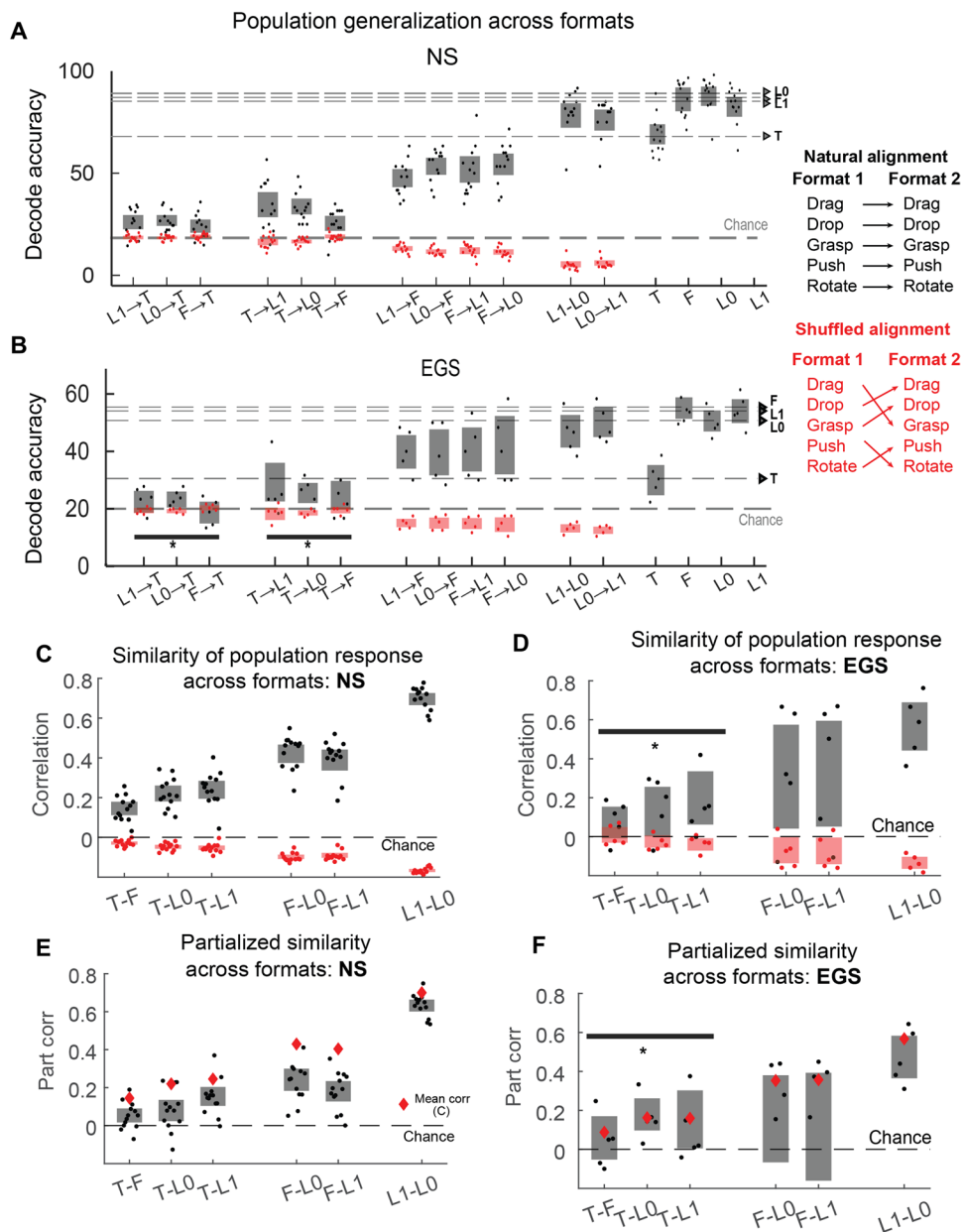
Fig. 2. Action verbs link with observed actions. (**A**) Across-format and within-format classification of manipulative actions. *x*-axis labels indicate the formats used for classifier training and testing (e.g., for across format, train→test). Dots = single-session result. Rectangle = 95% bootstrapped CI over sessions. Gray (red): values for matched (mismatched) labels across formats (see inset for definitions). Dashed horizontal lines show within-format cross-validated accuracy (mean across single-session results). All comparisons with chance performance (dashed line) or shuffled alignment reached significance (Wilcoxon rank-sum test, $P < 0.05$). (**B**) Similar to (A) but for EGS. Cross-format classification significant between all visual formats and between visual and text formats when pooling visual formats (see bar with asterisk). (**C**) Correlation of neural population responses across pairs of formats. Conventions as in (A). (**D**) Same as (C) for participant EGS (black horizontal bar indicates data that were pooled for statistical testing). (**E**) Pairwise population correlation while controlling for additional formats using partial correlation. Resulting correlations are above chance (part corr = 0) but below standard correlation values (mean = red diamonds). (**F**) Same as (E) for participant EGS.

when considering all three of the visual formats at once. This result is consistent with the significant but incomplete generalization of action identities across the visual formats shown in Figs. 2 and 3.

Having established that the same action is coded in different ways depending on details of visual presentation, we can now look at how action verbs link to these varied visual representations. We can frame our question in the following way: Do action verbs link with the entire population of cells demonstrating visual selectivity

or specific subpopulations of cells? Figure 4B illustrates these possibilities. Two primary theoretical possibilities in the literature describe how text can link with subpopulations of visually selective neurons. Overlapping describes the architecture in which verbs link specifically with the subpopulation of neurons that are invariant across the visual formats (*5*). Exemplar describes the architecture in which verbs link with a specific prototypical exemplar or "best example" of the word (*5*). The exemplar may be of a single visual presentation or
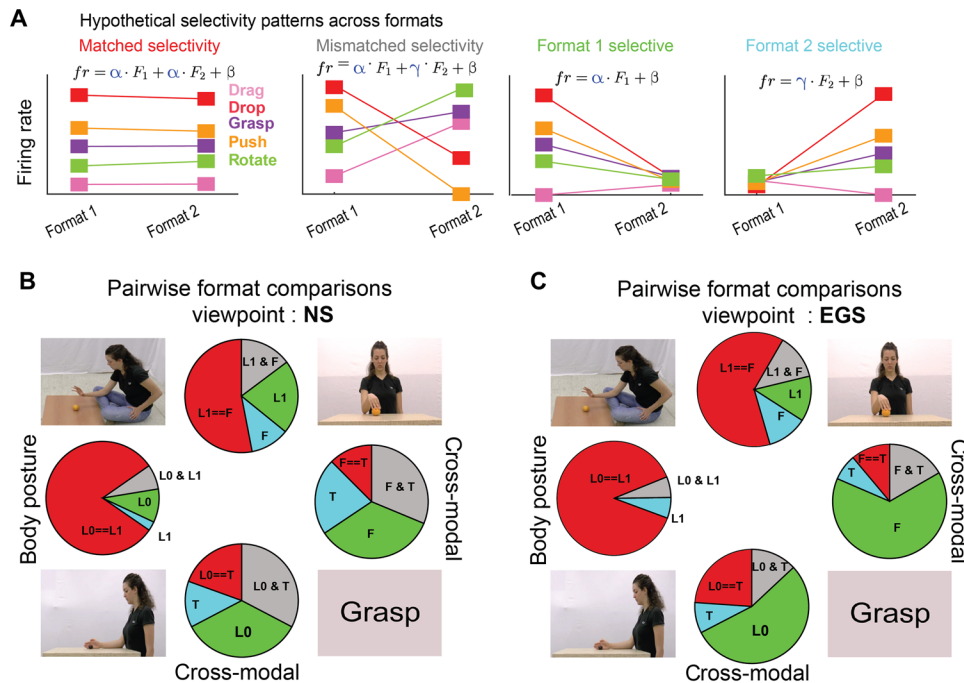
Fig. 3. Single-neuron SPs link action verbs and observed actions. (**A**) Schematic illustrating the four possible ways the SP can compare across two formats (see fig. S9 for expanded description). (**B**) Summary of SPs across pairs of formats for participant NS (see fig. S9). Red = matched SP; gray = mismatched SP; cyan and light green = selectivity for a single format only [see title colors in (A)]. Photo credit: Guy Orban, Department of Medicine and Surgery, Parma University. (**C**) Same as (B) for participant EGS. "=" indicates matched SP, and "&" denotes mismatched SP.

some subset of presentations. Last, we term the situation in which text links with all visually selective cells as Available. In this architecture, the link between text and the visual representations mirrors the statistics for how the visual representations are encoded within the neural population independent of text. Underneath each schematic, we provide a prediction for how the distribution of Fig. 4A should change when the model selection analysis accounts for how text links with the visual formats.

We extended the model selection analysis to categorize each unit based on the model that best described the SPs across all four formats (text + all visual formats). We compared the distribution of the visually selective units with a matched SP to text (Fig. 4C) to the full distribution of the visually selective units (Fig. 4A). The distribution was essentially unchanged; the subset of visually selective units that link with text reflects a random sampling of the visually selective units: A bootstrapped correlation analysis comparing the empirical distribution of Fig. 4C with the predictions of Fig. 4B shows that the population best matches the Available model (correlation with invariant = 0.32, exemplar = 0.48, available = 0.97). This provides the answer to the question of architecture: The distribution of text-linked units (Fig. 4C) mirrors the statistics of how visual formats are encoded independent of text, or, in other words, text forms links with all available visual representations. Units with a matching SP between text and at least one visual format (the distribution of 4C) represent 23% of all visually selective units (Fig. 4D) and 40% of all text-selective units (Fig. 4E).

What cognitive process does the link between action verbs and observed actions mediate?

Does the link between text and the visual formats reflect a semantic association, visual imagery, or short-term learned associations that

formed through the course of the experiment? Thus far, our analyses are based on averaging the neural response across the video duration. This large temporal window may encompass multiple cognitive processes. If neural processing for action verbs specifically reflects bottom-up semantic processing, we would expect to find a shared neural response between formats very soon after stimulus presentation. To address this issue, we performed a dynamic, sliding-window, cross-validated correlation analysis to look at how the relationship within and across formats evolves in time (Fig. 5, A and B). To understand how quickly the correlation between text and the visual formats emerges, the diagonal elements of the dynamic correlation matrices were extracted and plotted together for direct comparison in the inset panels of Fig. 5 (A and B). These results show that the cross-modal link between text and the visual formats is fast: The onset of the cross-format correlation between text and the visual formats is the same as the within-format text correlation. In other words, as soon as a neural response to text emerges, it immediately shares a common activation pattern with the observed actions.

Next, we checked whether the strength of population correlation changed over the course of the experiment. If neural processing for action verbs reflects a semantic association, we would expect to find the correlation between text and videos to be present from the first session throughout the course of the experiment. In contrast, if the correlation between text and action videos is a product of learned associations that developed over the course of the study, we would predict that the strength of correlation would increase over the course of repeated exposure to the action videos and text. We found that the early correlation response (cross-validated correlation over the first second of video presentation) between text and the three visual representations for each session did not depend on session number (Fig. 5, C and D), favoring the semantic interpretation.
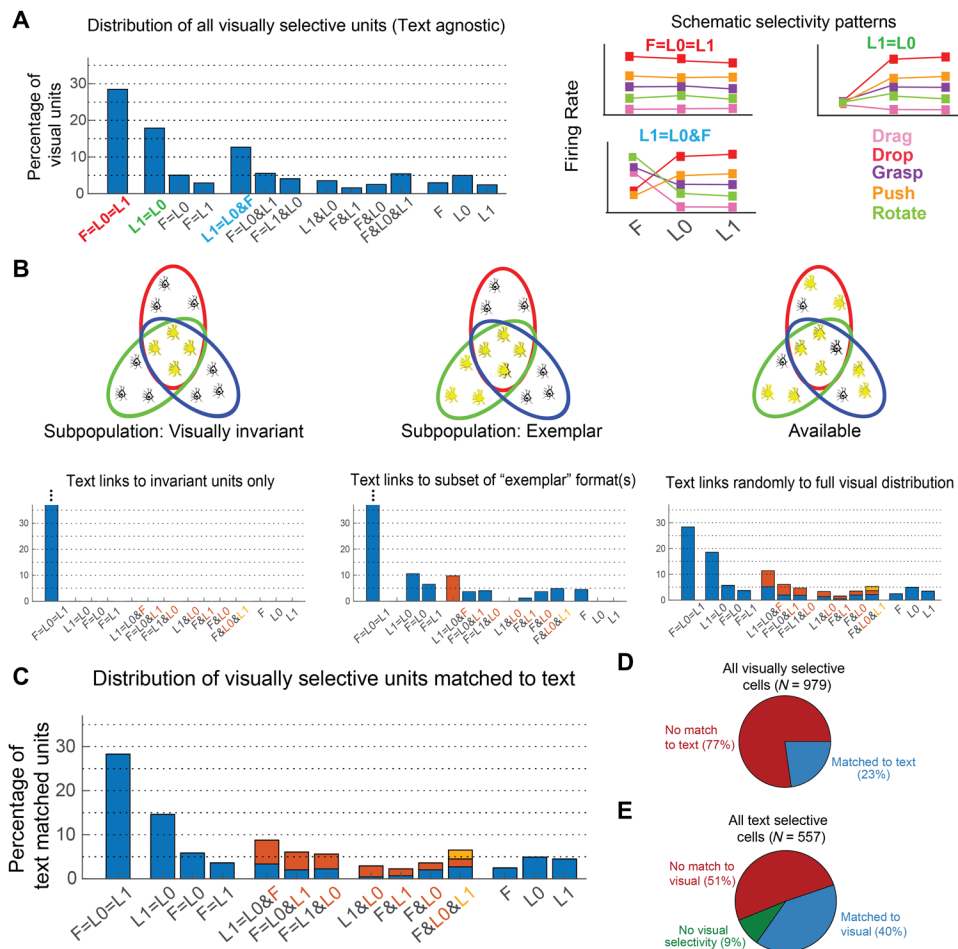
Fig. 4. Text links with all available visually selective cells. (**A**) Histogram characterizing how the population of neurons link action representations across the three visual formats (F,L0,L1). "=" indicates matched SP, and "&" denotes mismatched SP. Exclusion of a format indicates no selectivity. Three schematic SPs (right, color-coded) across the visual formats are shown to illustrate how the SPs compare across formats. (**B**) Schematic models illustrating different architectures of how text relates to three visual representations of the corresponding action. Each oval contains the population of neurons that are selective for a particular visual format. Overlap between ovals indicates matching selectivity across formats. The possible patterns of overlap between ovals may be more complicated (e.g., more overlap between two of the three ovals) but is simplified here for schematic purposes. Yellow neurons are selective for text with matching selectivity, while gray neurons are not. Underneath each schematic is a prediction for how the distribution in (A) will change when the model selection analysis filters the full distribution of (A) for units with matching text selectivity. (**C**) Similar to (A), however, the histogram is limited to the subset of visually selective units with a matched SP to text [blue subpopulation in (D)]. In cases where the units have mismatched visual SPs (e.g., L0 & F), text can have a matched SP with one of several of the visual formats. Colored segments of histogram indicate which format has matched SP with text (see x-axis labels for color code). (**D**) Percentage of visually selective units with a matched SP to text. (**E**) Percentage of text-selective units with a matched SP to at least one visual format, mismatched SP to visual formats, or without visual format selectivity.

We performed a number of control analyses and manipulations to address the possibility that associations between text and observed actions reflect visual imagery. In six sessions, participant NS was instructed to use visual imagery to "replay" the associated action video in her mind from either the front (F) or side (L0) perspectives when given the action verb prompt. If imagery were a dominant factor in establishing the link between text and observed actions, the explicit manipulation of visualizing from the F or L0 perspective should bias the percentage of cells with a matched SP in favor of F or L0. However, both the total number of significant units and the population-coding structure were essentially unaffected by the explicit task instruction. Neither the proportion of significant units [Fig. 6A, $\chi^2(1,1432) = 2.7$, $P = 0.1$] nor the proportions of the best explanatory models [Fig. 6B, $\chi^2(1,1432) = 1.9$, $P = 0.58$] demonstrated significant differences. Further, a comparison of the per-session

population correlation did not show a significant effect of the instruction (Fig. 5C, Wilcoxon rank-sum test, $P = .43$). This result shows that the basic link between action verbs and observed actions is not dependent on the contents of visual imagery. To probe this result further and ensure the subject followed task instructions, we split the dynamic correlation analysis between the passive and active imagery sessions. We found (Fig. 6, C to E) that correlation immediately following stimulus presentation was largely unaffected by the behavioral manipulation, while correlation near the end or after stimulus presentation did show significant differences (paired t test, $P < 0.05$ on pixel values split between passive and imagery sessions). This result suggests that the subject followed task instructions and that imagery can affect neural responses, but the early responses (that are the hallmark of automatic semantic processing) are independent of the contents of imagery.
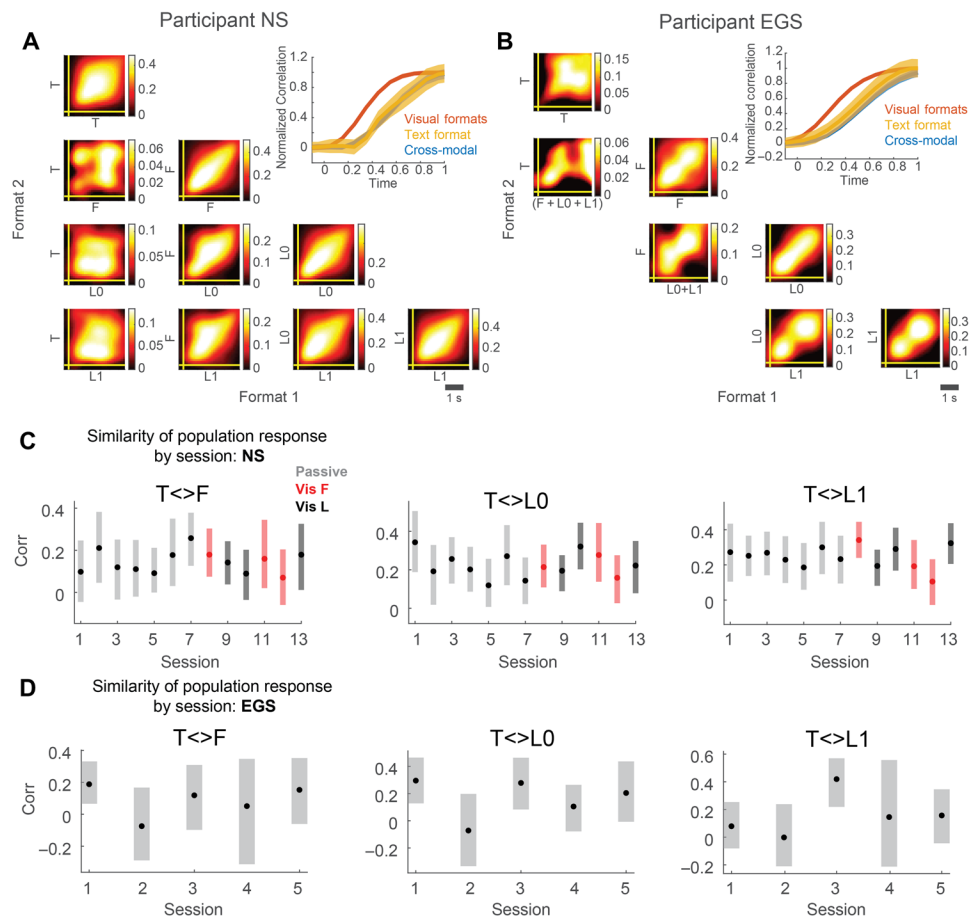
Fig. 5. Temporal features support a semantic link between verbs and observed actions. (**A** and **B**) Cross-modal match between text and visual formats occurs at low latency. (A) Dynamic cross-validated cross-correlation matrices demonstrating how the neural population response during stimulus presentation at one slice of time compares to all other slices of time, both within and across formats. Format comparisons as shown in x- and y-axis labels. Correlation magnitude as indicated by the color bar. Inset: The diagonal elements of the within- and across-format matrices were averaged into three logical groupings [(i) within-format visual, (ii) within-format text, and (iii) across-format text to visual] and normalized to a peak amplitude of 1 for comparison purposes. The temporal profile of the averaged correlations (means ± SE across sessions) is plotted to emphasize the similarity of onset timing for the within-format text and across-format text to visual population correlations. (B) Similar to (A) but for participant EGS. To compensate for the smaller number of sessions, we grouped correlation matrices for cross-modal comparisons. (**C** and **D**) Stable relationship between text and observed actions through experimental sessions. (C) Cross-format correlations for subject NS shown for text and the visual formats on a per-session basis (mean with 95% bootstrapped CI). Color code shows whether the subject was passively viewing stimuli or asked to actively imagine from the lateral or frontal perspective (see inset; Vis F = visualize from frontal perspective; Vis L = visualize from the lateral 0 perspective). (D) Same as (C) except for participant EGS (only silent reading).

In a final control, we collected a dataset in which four abstract symbols (snowflakes; fig. S10A) were paired with visual imagery of movements for over 2 months (31 recording sessions, 114 ± 11 units per session; fig. S10B). In this paradigm, subject NS was asked to visualize a movement from the first-person perspective when presented with a symbol. The subject learned this task well, as we could accurately decode the different symbols when the subject was instructed to use visual imagery (fig. S10C). We also asked the subject to passively view the same stimuli at sporadic intervals (fig. S10B, vertical orange lines) and found that the ability to decode the different symbols disappeared (fig. S10, D and E). The differences between passive viewing and active imagery when cued with experimentally defined abstract symbols in the control task provide a stark contrast to the differences between passive viewing and active imagery when viewing action verbs in the main experiment (fig. S10, D to G). The differences help to clarify several points about the main experiment.

The clear differences in classification accuracy between passive viewing and imagery in the control task demonstrate that the subject is capable of comprehending and following task instructions as they relate to passive viewing versus active visual imagery, two tasks used in the main experiment. Furthermore, the study shows that not all types of visually distinct stimuli elicit a differential neural response under passive viewing. Last, it demonstrates that the recorded population does not form automatic neural responses to arbitrary abstract symbols, even when the different symbols have been learned and are of direct behavioral relevance.

DISCUSSION

Our results answer the four questions raised in the introduction: PPC neurons exhibit selectivity for action verbs and observed actions; text links to visual representations of observed action; text links
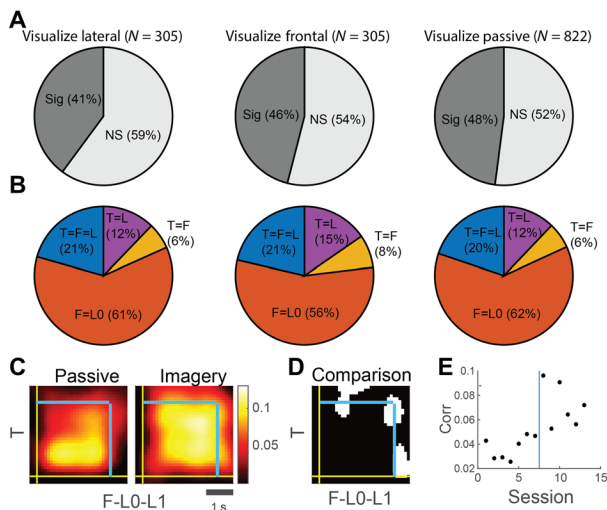
Fig. 6. The effect of explicit instruction on cross-format invariance. During the initial seven sessions, subject NS silently read action verbs. In the six subsequent runs, she explicitly visualized the frontal (F, three runs) or lateral standing (L0, three runs) perspective in response to the action verb. (**A**) The percentage of units with a significant effect of action or action-format interaction for the format by action ANOVA applied to the triplet of formats pertinent to task instruction (T,F,L0). "Sig" = significant at $P < 0.05$ FDR-corrected ("NS" otherwise). Results are split by the task instruction. Total number of sorted units shown in title. (**B**) Results for the combined (BIC + cvR^2) model selection analyses for the same triplet of actions split by task instruction. The percentage of T=L units was twice as prevalent as T=F units for passive viewing, as well as the two instructed conditions. (**C**) Mean dynamic cross-correlation between the visual formats and text split by passive viewing and active imagery in participant NS. Blue lines indicate video offset. (**D**) Pixel coordinates demonstrating a significant difference between passive viewing and active imagery (significant pixels in white, paired t test, $P < 0.05$.) Blue lines indicate video offset. (**E**) Cross-correlation value between text and the visual formats for the set of significant pixels shown in (D) as a function of session number. The blue line shows split between passive and active imagery sessions.

with a fraction of all available visual representations; and the link is most consistent with being semantic in nature and not due to imagery or learned associations.

Answers to the four questions
First: Selectivity
Both single-cell properties and within-format decoding demonstrate neuronal selectivity for action verbs and observed actions in human PPC. The visual selectivity had short latencies (about 150 ms), while text selectivity emerged nearly 150 ms later. The features of the visual stimuli that determined neural selectivity remain unclear. The term selectivity for action identity should be interpreted as a label assigned to the visual stimuli rather than coding for the basic-level type of action, e.g., "grasp." Manipulations of viewpoint or fixation point (fig. S2) changed neural coding significantly. Manipulative actions can differ in hand and arm postures, contact points with the object, and dynamics, among others; these parameters should affect neural coding to represent the behavioral complexity of natural actions. Elaborating the exact degree to which neural coding is influenced by action identity, its many parameters, or even low-level visual features needs further work. Nonetheless, the link between action verbs and observed actions demonstrates that coding of action identity cannot solely be driven by irrelevant visual features. Further, not all

visual differences are encoded by the neural population (fig. S2). Last, high-dimensional coding of both category-relevant and -irrelevant visual features is consistent with neural coding in high-level regions of the ventral visual stream (*26, 27*).

Second: Action verbs and observed actions share a common neural substrate
We demonstrate the shared substrate at the population level using cross-format decoding and population correlation between formats (Fig. 2) and showed the basis of this population link by modeling of single-cell selectivity across pairs of formats (Fig. 3). Prior neuroimaging evidence indicates a degree of anatomical overlap within the AON for processing observed actions and language (*19–22*). However, imaging evidence can be inconsistent (*8*), and gross anatomical overlap seen in neuroimaging does not directly imply that the same neural populations support both tasks (*7*). Our evidence provides definitive evidence for a shared neural substrate by demonstrating that the precise SPs for action verbs match the SPs for corresponding observed actions at the neural unit level.

Third: Architecture
We have established that, at the neural level, action verbs link with visually observed actions, suggesting that sensorimotor representations are an intrinsic component of verb meaning. The potential implications of this finding are hard to pin down without understanding the architecture of the link. There are infinite visual stimuli that could be considered a "grasp" or a "banana" or any basic category colloquially used to describe an object or action. Our results establish that neural coding for observed actions depend on presentation details (see Fig. 4), consistent with findings throughout cortex (e.g., *9, 10*). Given the diversity of neural coding, there are three likely architectures (Fig. 5B), each with its own implications for how linking is made between symbolic and visuomotor representations. Text could link exclusively to the subpopulation of cells that are visually invariant across the different visual presentations (e.g., Fig. 4B, "visually invariant"). In such a case, the aspect of "meaning" conveyed by the sensory-motor representation would be what is universal or common to all presentations. In other words, sensory-motor meaning abstracts away the details of any particular representation. Another possibility is that text could link to one or a subset of example stimuli (e.g., Fig. 4B, "exemplar"). In such a case, the aspect of "meaning" would constitute representative visual examples of the word. The third possibility is that text links to all available visual representations (e.g., Fig. 4B, "available"). If the visual representation reflects the consolidation of one's experiential history with observed actions (*4, 28*) as expected for the consolidation of semantic memory, then neural responses to text may be understood as the activation of this consolidated visual experience. This suggests that a word's meaning is uniquely rooted in an individuals' experience.

The comparison between the predictions for these three models and the data strongly favors the available model. This architecture is also the easiest to implement, as a simple Hebbian mechanism will suffice and would predict that acquisition of verb meaning depends on the frequency of exposure, which has been observed for several languages (*29*). The text response links only to a subset of the full distribution of visually selective units (Fig. 4D). The reason for this is unclear but may reflect inefficiencies in the neural process that links verbs with visual representations and may be influenced by exposure or experience. In any case, reading a word does not evoke the same perceptual experience as viewing an action, and thus,

substantial differences at the level of neural responses should be expected.

Fourth: Origin of the link

Our results indicate that the link between text and the visual formats is not the product of imagery or learned associations that emerge from the task. Our results are consistent with action verbs automatically eliciting a memory or visual/multisensory representation of an action. This could be considered a form of imagery; however, here, we use imagery to specifically refer to the effortful covert internal simulation of a movement (either of one's own body or another's body) such as might occur when a participant is explicitly asked to imagine a movement. A primary distinction between semantic memory and imagery as defined here is that semantic responses are automatic; when the action verb is read, corresponding representations in PPC are activated without conscious effort or task dependence. In the control task, passive viewing of symbols associated with actions is shown to be an ineffective stimulus to drive the neural population. There is no automatic response. Neural responses are task dependent and only found when the participant actively imagines the actions that have been associated with each cue. This is contrasted with responses to action verbs in which selectivity is found under silent reading with minimal impact from experimental manipulation of imagery. The action verbs are processed automatically, requiring nothing beyond reading to generate action-specific neural responses. Semantic processing should be fast and automatic, and we found that it is exactly the early component of the correlation that was unaffected by the imagery manipulation (Fig. 6D). In contrast, the late components of the correlation systematically differentiated passive viewing and imagery sessions (Fig. 6, C to E), demonstrating that the patient followed the instructions, a view also supported by the control experiment. From these considerations, the shared neural substrate of text and observed actions is unlikely to reflect imagery. A key signature of learned associations is the gradual strengthening of the link between text and observed actions. Yet, the correlation between text and the visual formats was stable across all testing sessions, including the very first one (Fig. 5). In our control experiment (fig. S10), passively viewing abstract symbols that had been paired to movement imagery did not induce selective neural responses. Thus, our results are unlikely the consequence of learned associations.

We consider yet two more alternatives to semantic processing. The first is that neural responses to action verbs and observed actions represent implicit automatic motor plans (30). How we plan and execute an action is an important component of meaning. However, our control study (fig. S10) revealed no selectivity to passive observation of movement-predictive cues. The second possibility is that the linked SPs for observed actions and action verbs reflect a population of cells that are responsive to the internal act of silent naming (i.e., generating the action verb). In this view, when viewing text or videos, the participant covertly generates the same word and thus produces the same activity patterns. This hypothesis would predict results similar to the invariant hypothesis, as generating the action verb should be consistent across the different visual presentations. Instead, for simultaneously recorded neural populations, we find that text responses link to the visual formats in idiosyncratic ways (e.g., text and the lateral views, but not the frontal; Figs. 1A and 4C). It remains possible that neurons are selective to particular cue-naming pairs; however, in our prior work (31) in the same participants, we found no selectivity for specific cue-intention

pairings (e.g., response for imagined movement to the right when cued with a spatial target, but not when cued with symbol). Thus, we think that the naming hypothesis is unlikely.

From the above considerations, we believe that our results are most compatible with the shared neural substrate mediating semantic memory, reflecting associations between the word and its visuomotor representations that have been built over years of experience. In this view, reading words automatically activates sensorimotor representations, and these representations are in a position to color our understanding of word meaning without our conscious effort.

Nature of neuronal representation of variables coded in PPC

The ability of small neuronal populations to encode many variables is consistent with the mixed-selectivity scheme in which distributed, nonlinear, high-dimensional representations code in a contextually dependent manner (32). However, at least within the cortical locations explored in the current study, we find that such encoding is not random, but systematically organized around stimulus properties, a scheme referred to as partially mixed selectivity (33). Neural populations coding the same basic-level action exemplar for different formats overlapped (e.g., Fig. 3 and fig. S9). Partial mixed selectivity may represent a general structure for representing sensorimotor aspects of meaning within association cortices, resulting in rich links between text and the diversity of overlapping and distinct components of the visual formats that mirror the statistics of visual encoding independent of text (Fig. 4). It is unclear whether neural overlap reported for observed and performed actions in nonhuman primate (NHP) follows similar principles of neural architecture, in part because results in NHP studies have generally been reported for responsiveness (e.g., change from baseline) to a single action (typically grasping) rather than selectivity (e.g., differential responses) for multiple distinct actions [e.g., (18)]. The partially mixed architecture may account for the weak link between text and the visual formats (e.g., relatively low-population correlation and few units with matching SPs). If a cortical region encodes the many visual facets of an observed action (e.g., viewpoint, posture, and other untested features) and text links with both what is overlapping and distinct about action presentations, it follows that the link between text and any particular presentation must be relatively weak.

Cortical organization of conceptual knowledge

In understanding an action verb, we access semantic knowledge. The cortical organization of semantic knowledge has been contentious. Some theories contend that conceptual knowledge is rooted in cortical regions that use supramodal symbolic processing (7), while other theories take the opposite perspective, that semantic knowledge is encoded in the distributed sensorimotor network (6, 34). Most recent theories posit that meaning emerges from interactions between supramodal associative areas and regions directly responsible for processing sensory stimuli, motor actions, valence, and internal state (1–5). Our results are consistent with these interaction models, given the longer latencies we observed for text-selective responses in PPC, relative to those reported for higher-order language regions such as superior temporal gyrus or inferior frontal gyrus (35). One likely possibility is that action verb activity in PPC originates from supramodal regions and automatically spreads to PPC. This interaction model comes in many versions, primarily distinguished by which areas constitute the supramodal regions and the nature of the interactions. In part, a deeper understanding of the

organization of conceptual knowledge in the human brain has been limited by the general inability to record from single neurons in humans. We know of no single-unit recordings in supramodal regions, but one intriguing possibility is that these areas may host neurons similar to the "concept cells" of the medial temporal lobe (MTL) (36), which respond to a preferred stimulus (e.g., a particular individual) largely independently of sensory modality or presentation details (e.g., image, written word, and sound). While this strong invariance provides a model for neural coding mechanisms in supramodal centers, much less is known about how semantically related items are encoded in the distributed network. The current study contributes to this goal by providing the first demonstration of a link between words and their sensorimotor representations and how the neural architecture supports this link.

In the current paper, we have focused on how verbs are given meaning. We may also consider what our results mean from the reverse direction, how the neural population may contribute to naming an observed action. We find not only relatively high generalization across different views of the same observed action but also a degree of dependence on viewpoint and the point of fixation (fig. S2). These neural properties suggest that rostral PPC neurons could play a role in creating increasingly abstracted representations that associate the same actions and thus contribute to the processing needed for naming, but, given the weakness of the link, subsequent regions, potentially using winner take all like mechanisms, would be needed for the final conversion to labeling the observed action.

The link between visual representations of actions and action verbs fits with current views of how infants learn action verbs by mapping words onto conceptualizations of events (37). Infants can distinguish action exemplars (running, marching, and jumping) independently of the actors (38) and that this ability predicts the use of action verbs at 2 years of age (39). Furthermore, it provides an explanation for why infants learn verbs later than nouns (40), as the corresponding visual representations are in different visual pathways. In the PPC, the development of observed action selectivity, which is originally in the service of guiding future actions (18), may only occur once the infant starts moving. Infants initially learn verbs corresponding to their own actions (41).

Limitations of the study
Stimuli
We used a restricted set of observed actions and action verbs, based on the category of actions that best evoke responses in neuroimaging (42). Thus, our results cannot support the conclusion that responses to written text are specific for action verbs. Neuroimaging studies have shown that brain regions exhibit some degree of domain specificity during language processing (43). Understanding domain specificity of responses to language at the single-unit level is an exciting future direction.

Visual formats
We tested only a small number of visual formats: two postures and two viewpoints. Thus, the visual invariance that we established may be an overestimation, and increasing the diversity of different presentations of the same action would lower the percentage of invariant cells. Hence, while it remains possible that the visual invariant neurons (F=L0=L1 in Fig. 4) are akin to concept cells as described in the MTL of humans, this is by no means established. To this point, neurons exhibiting invariance in the MTL showed sparse coding (only active for a single basic-level category), while the invariant neurons

tested in our study were broadly tuned, matching the tuning profiles of other visually selective neurons (fig. S6). The small number of visual formats may also partially account for text-selective units with mismatched or absent visual selectivity (Fig. 4, D and E) as they may link with other untested visual representations of the corresponding action identity.

Recording site
We tested only one region of the AON. Other regions of the AON (e.g., premotor areas or the LOTC), based on neuroimaging and lesion, likely play a role in linking language with its sensory and motor representations. Action verbs may be associated with the kinematic profiles of movement, movement dynamics, the agents typically performing the action, the objects typically subjected to the actions, the desired outcome or value of the action, and the expected sensations that accompany the action, among others. The constituent regions of the AON likely encode these movement attributes and together may form the distributed network that links action verbs with these varied aspects of meaning.

Causality
As with all passive neural recording studies, our study cannot determine the causal role of our PPC neurons in understanding the meaning of action verbs. However, prior work, using word or static picture stimuli, has shown that damage or inactivation within the frontoparietal AON, including PPC, can result in specific action comprehension deficits (23–25) consistent with the idea that neurons within the AON play a role in verb comprehension. Our results provide clarity on the presence and nature of the link between neural representations of action verbs and visually observed actions at the level of single units in PPC.

Subjects
We investigated neural signals in two participants and thus cannot make strong conclusions about factors that influence the strength of action verb encoding. Participant NS demonstrated stronger selectivity than EGS, even when controlling for the number of neurons and sessions (fig. S4). The reason for differences are unclear but may be the product of individual differences and could include anything from the degree to which the two participants attended to stimuli on a trial-to-trial basis to the degree to which individuals intrinsically engage sensory-motor systems during semantic processing. One intriguing difference is that NS is a native English speaker, while EGS is a fluent but nonnative speaker having learned English as part of a language program in primary school. One possibility is that the time of language acquisition may affect the degree to which words engage sensory-motor systems. In addition, the recorded neurons may come from different functional regions due to either anatomical differences in implant location or high individual differences in how functional regions map to cortical anatomy. A precise functional correspondence of areas is unlikely; however, we note that functional responses were similar during functional neuroimaging (fig. S1), as well as during planning and execution epochs of motor imagery tasks at the single-unit level (31, 33).

Conclusion
The current study provides the first single-unit evidence that action verbs share a neural substrate with visually observed actions in high-level sensory-motor cortex, thus clarifying the neural organization of human conceptual knowledge. Action verbs link with all the diverse visual representations of the related concept, suggesting that language may activate the consolidated visual experience of the reader.

MATERIALS AND METHODS

Experimental design

Data acquisition

All procedures were approved by the California Institute of Technology, University of California, Los Angeles, and Casa Colina Hospital and Centers for Healthcare Institutional Review Boards. Informed consent was obtained from NS and EGS after the nature of the study and possible risks were explained. Study sessions occurred at Casa Colina Hospital and Centers for Healthcare and Rancho Los Amigos National Rehabilitation Center.

Behavioral setup

All tasks were performed with NS and EGS seated in their motorized wheelchair. Tasks were displayed on a 28- or 47-inch liquid crystal display monitor. The monitors were positioned so that the screen occupied approximately 25° of visual angle. Stimulus presentation was controlled using the Psychophysics Toolbox for MATLAB.

Physiological recordings

NS and EGS were implanted with one 96-channel NeuroPort Array on the gyrus dorsal to the junction of the IPS and postcentral sulcus (PCS; fig. S1). These locations were implanted based on three considerations: First, the NeuroPort Arrays used in the current study are not suitable for implantation within sulci given short electrode shank lengths (≤1.5 mm) and lack of long-term viability for direct implantation within sulci. Thus, implant locations must be restricted to gyri accessible on the cortical surface. Second, the cortical regions of interest are near the junction of the IPS and PCS. This consideration was included as we were targeting functional responses related to grasping, manipulation, and other behaviors that emphasize the hand. Cortical regions within and around the junction of the IPS and PCS in human neuroimaging studies have consistently shown preferential responses to hand-based actions. Third, we used functional magnetic resonance imaging within the individual participants to identify regions with a preferential response for grasping actions. We used two neuroimaging tasks suitable for paralyzed individuals to identify grasp-related responses in each individual subject. The resulting functional responses, combined with the constraints described above, determined the implant locations shown in fig. S1.

Grasp-related responses around the junction of the PCS and IPS have typically been described as the anterior IPS or putative human homolog of the anterior intraparietal area (phAIP) and is generally assumed to be the human homolog of macaque AIP. Macaque AIP is a region localized to the lateral bank of the anterior portion of the IPS involved in the visual control of grasping actions. However, the medial bank of the anterior IPS contains a distinct grasp field. This grasp field, described as PEip (intraparietal) or Brodmann's area 5L (BA5L), is characterized by distinct frontoparietal connections (AIP is densely interconnected with PMv (Premotor ventral), while PEip is connected with the rostral portion of M1), direct connections to the hand regions of the spinal cord, bilateral somatosensory responses, and functional responses related to hand and finger movements. While some progress has been made in the identification of the human homolog of AIP (17), the human homolog of PEip/BA5L has not yet been established, and it may be the case that neuroimaging results around the junction of the PCS and IPS include both the human homologs of macaque AIP and PEip/BA5L. Additional work probing the single-unit properties of the arrays in the two human subjects is needed to better understand the functional homologies of the regions investigated in the current study. In light of this uncertainty, we refer to the recording sites as parietal grasping regions.

Neural activity was amplified, digitized, and recorded at 30 kHz with the NeuroPort neural signal processor (NSP). The NeuroPort System, comprising the arrays and NSP, has received Food and Drug Administration (FDA) clearance for <30 days of acute recordings; for purposes of this study, we received FDA IDE (Investigational Device Exemption) clearance (IDE #G120096) for extending the duration of the implant.

Single-unit and multiunit activity was sorted using k-medoids clustering using the gap criteria to determine the total number of neural clusters. Clustering was performed on the first n principal components, where n was selected to account for 95% of waveform variance (range of two to four components). Sorting was reviewed and adjusted if deemed necessary following standard practice by merging or splitting clusters as needed.

Tasks and stimuli

Experimental stimuli consisted of five manipulative actions (drag, drop, grasp, push, and rotate) displayed in four different "formats": three visual video formats and one text format. In two visual formats, the actors were viewed from the side, but the actor was either standing next to a table (Lateral 0, L0) or sitting in a lotus position on the floor (Lateral 1, L1). In the third visual format, the actor was standing next to the table but was viewed from the front. Thus, L1 differed from L0 only by body posture, F differed from L0 only by viewpoint (note that two video cameras were used to simultaneously acquire videos for F and L0, and thus, the timing and kinematics of the movements are identical), and F differed from L1 by both viewpoint and posture. In the text condition, the written action word (Arial at font size 80) was shown for 2.6 s. Experimental stimuli of the visual formats consisted of video clips (448 × 336 pixels, 50 frames/s) showing one actor at a distance of 1.2 m performing five different hand actions (drag, drop, grasp, push, and rotate) directed toward an object (four versions per action format). The objects were positioned directly adjacent to or within the hand such that actions predominately involved the wrist and fingers. All videos measured 17.7° by 13.2° and lasted 2.6 s (the first two and the last two frames being static). The edges of the videos were blurred with an elliptical mask (14.3° × 9.6°), leaving the actor and the background of the video unchanged but blending it gradually and smoothly into the background around the edges. We did not enforce fixation; instead, we asked the subject to view the actions in a naturalistic manner. The effects of fixation were documented in a separate set of experimental sessions described below. Each data session consisted of 12 repetitions of each unique video. Presentation was split into three runs (four repetitions per run, corresponding to the four versions). Videos were presented in a pseudorandom manner: All conditions were randomly ordered and presented once before repetition. Video stimuli used for L0, L1, and F0 were also used in (44), which tested neural encoding of observed actions in nonhuman primates. Presentation of the videos differed, however, as before video presentation, during baseline, this study used a highly blurred (full width at half maximum = 80 pixels) static frame (average of all video frames).

We collected these data under two different instruction sets. In the first instruction set, the subject was instructed to attend to but otherwise passively view experimental stimuli. For the text format, we asked the subject to read the word silently without any accompanying visualization. We collected seven (subject NS) and five (subject EGS) sessions in this passive viewing paradigm. In the second instruction set, the subject NS (six sessions) was instructed to use the text as a prompt to visualize the associated action being performed

from either the frontal (F; three sessions) or lateral standing perspective (L0; three sessions). Six sessions in total were collected over 3 days. Each session was collected in full under the same instruction. The sessions were collected in the following order: (F, L0), (L0, F), (F, L0), with () indicating that the sessions were acquired during the same day. Before each run, the participant was reminded of the experimental condition. Following each run, the participant reported which perspective she used to imagine the actions.

Information on the preliminary action observation task (fig. S2) to test for the presence of units selective to observed actions is described in Supplementary Materials. Additional information is found in the Supplementary Materials.

Statistical analysis

Within-format individual neuron analyses

Individual unit event-related averages (Fig. 1A). For each unit, neural activity was averaged within a 750-ms window starting from −1.5 s before video onset and stepping to 4 s with 100-ms step intervals. Neural responses were grouped by the observed action exemplar, and a mean and SEM ($n = 12$) were computed for each time window and for each action.

Within-format one-way ANOVA for action identity (Fig. 1B). We defined a unit to be selective for action identity for a particular format if the unit displayed a significant differential firing rate for the five action types during video presentation. Firing rate was taken as the total spike count during movie presentation (2.6 s) starting at 0.25 s after onset divided by the window duration. Firing rates were subjected to a one-way analysis of variance (ANOVA) with the factor of action identity, and significance was determined as $P < 0.05$ after false discovery rate (FDR) correction. To ensure that selectivity was driven by the task stimuli, we repeated the analyses using the 1-s window before stimulus onset and found that no neurons (0%) were selective in any format after FDR correction.

Within-format linear fit of neural responses (percent selective, cvR^2, and coverage). We used a linear regression model for multiple analyses. For each neuron, we fit a linear regression model that explained neural firing rate during movie presentation as a function of the categorical variable manipulative action identity (five actions: drag, drop, grasp, push, and rotate). For some analyses, linear fits were performed separately for each of the four formats (e.g., written verb, frontal view, and the two lateral views). For others, such as the model selection procedure described below, we included various combinations of formats with constraints on the linear model parameters (described below). Firing rate was taken as the average unit response in a single window that extended for the duration of movie presentation (2.6 s) starting at 0.25 s after onset. The baseline response was taken as the average activity in the 1-s preceding movie onset.

Coverage (fig. S5). We computed the P value associated with the coefficient estimate of each action (e.g., drag, drop, etc.) and found the percentage of coefficients associated with each action having a P value less than 0.05 after FDR correction. We used a bootstrap procedure for generating 95% confidence bounds on the estimates of the percent selective.

We also wanted to know the frequency with which the different actions resulted in the peak response. To test this, we first split the data into excitatory and inhibitory units. Inhibitory units were defined as units for which the beta coefficient for all five actions was negative and thus suppressed relative to the baseline response. We

then identified the action that resulted in the largest deviation from baseline activity and counted the number of units that showed a peak response for each action.

Cross-validated coefficient of determination (Fig. 1C). To derive a measure of strength of selectivity, we performed a leave-one-out cross-validation procedure to estimate the cvR^2. For each fold, the single-unit regression was parameterized using all but one trial, and the resulting model was used to predict the firing rate of the held-out trial. This was repeated, and the R^2 was computed using the held-out data as 1 − (sum of squares of residuals)/(total sum of squares). The 95% confidence interval was estimated with a bootstrap procedure.

Selectivity curve analyses (fig. S6). Is neural coding for observed actions sharp or graded? For each selective unit (within-format ANOVA, $P < 0.05$, FDR-corrected; see above) within each format, repetitions for each observed action were split in half to create training and testing splits of the data. Repetitions were then averaged to create a single value per action for each of the training and test sets. Training set data were rank-ordered from the action resulting in the highest firing rate to the action resulting in the lowest firing rate. This computed order was used to sort the test data. This process was repeated 500 times, and the results were averaged across folds. The result is a cross-validated measure of the response of each unit as a function of rank. Responses were normalized between 0 (response to worst action computed from the training data) and 1 (response to best action computed from the training data) before averaging across the population of selective units. Confidence intervals were estimated using a bootstrap procedure. Both the mean with 95% confidence interval (estimated with a bootstrap procedure) and the full distribution (shown in a violin plot) are presented.

Within-format neuron population analyses

Time-resolved classification of action exemplars (Fig. 1D and fig. S10). We performed sliding-window classification analyses to measure the strength and latency of population coding of observed actions and action verbs. For each time window, we constructed a classifier to differentiate the observed action for each format separately. Classification analyses were performed using linear discriminate analysis (LDA) with the following assumptions: (i) the prior probability across the action exemplars was uniform; (ii) the conditional probability distribution of each unit on any given action exemplar was normal; (iii) only the mean firing rates differ for each action exemplar (the covariance of the normal distributions were the same for each action exemplar); and (iv) the firing rates of each input are independent (covariance of the normal distribution was diagonal). Relaxing these constraints (e.g., allowing a full-rank covariance matrix) generally resulted in poorer generalization performance. The classifier took as input a matrix of average firing rates for each sorted unit. We did not limit analyses to action-selective units to avoid "peeking" effects. Classification performance is reported as generalization accuracy of a stratified leave-one-out cross-validation analysis. The average neural response was calculated within 300-ms windows, stepped at 10-ms intervals. Window onsets started from −0.75 s relative to video onset with the final window chosen to be +3 s. The window size was chosen to ensure that a reasonable estimate of firing rate could be determined while still allowing temporal localization. Classification was performed on all sorted units aggregated across all sessions. Mean and bootstrapped 95% confidence intervals were computed for each time bin from the cross-validated accuracy values computed for each session. To provide an additional visual marker, we display as a horizontal line the bootstrapped

97.5th percentile averaged over all prestimulus time bins for the text condition.

Neuron-dropping curve analysis (figs. S4 and S10). Neuron-dropping curves were constructed to compare how population-level encoding of exemplar actions compared between the two subjects in a controlled manner. To construct the random neuron-dropping curves of fig. S4, we computed cross-validated decode accuracy using LDA classification (described above) for test populations of neurons ranging in size from 1 to 150 units. Sampling of neurons was performed separately for participants NS and EGS. Units from NS were restricted to the first five sessions to equate exposure (EGS sessions were limited to five) and experimental instruction (passive viewing). Thus, the neuron-dropping curve analysis controls for number of units, task exposure, and experimental instruction. Each test population was generated by randomly subselecting, without replacement, the specified number of units from the entire ensemble of recorded units. For each population size, units were randomly drawn, and cross-validated accuracy was computed 200 times to allow estimation of the variability in accuracy. As a result, the actual decoding algorithm was trained on a subset of the total number of units as determined by a per-unit significance test calculated on the training data (not including the test data) used for cross-validation analysis.

Latency analyses (fig. S7, text in Fig. 1D). Latency was estimated using a sliding-window within-format decode analysis: Classification performance was computed as generalization accuracy of a stratified leave-one-out cross-validation analysis. To better temporally resolve the signal, accuracy was computed on data stepped in 5-ms windows and smoothed with a 25-ms full width at half maximum truncated Gaussian smoothing kernel. Classification was performed on all sorted units aggregated across all sessions. For each time window, significant classification performance was determined when true cross-validated classification was greater than 97.5% of values of an empirical null distribution of classification accuracies generated by randomly shuffling labels (250 shuffles). Latency is reported as the first window with significant classification for at least 10 consecutive time bins (50 ms). Latency measured in this way was computed separately for each format of the main experiment as well as the imagery condition of the control experiment. Latency analyses were not attempted for subject EGS as decode accuracy was worse, and we had fewer sessions (5 sessions for EGS versus 13 sessions for NS), although the general time course of decode accuracy was similar (see Fig. 1D).

Cross-format individual neuron analyses

Comparing SPs between formats (Figs. 3 and 4 and figs. S9 and S11). How a unit codes for action identity within a format is described by its SP defined as the precise firing rate values for all five action identities (see Fig. 3A). At the single-unit level, understanding how a unit codes action identity across formats can be quantified by comparing SPs across formats. There are four possibilities when comparing SPs across a pair of formats: (i) the SPs are similar or matched across formats; (ii) both formats are selective, but the SPs are distinct across formats; (iii) only the first of the two formats is selective for action identity; and (iv) only the second format is selective for action identity. Each of these possibilities can be defined mathematically within a linear model framework using four models of neural coding across formats

$$\text{model 1:fr} = \boldsymbol{\alpha} \cdot \boldsymbol{F}_1 + \boldsymbol{\alpha} \cdot \boldsymbol{F}_2 + \beta$$

$$\text{model 2:fr} = \boldsymbol{\alpha} \cdot \boldsymbol{F}_1 + \boldsymbol{\gamma} \cdot \boldsymbol{F}_2 + \beta$$

$$\text{model 3:fr} = \boldsymbol{\alpha} \cdot \boldsymbol{F}_1 + c \cdot \boldsymbol{F}_2 + \beta$$

$$\text{model 4:fr} = c \cdot \boldsymbol{F}_1 + \boldsymbol{\alpha} \cdot \boldsymbol{F}_2 + \beta$$

In model 1, the linear fit is constructed with the constraint that the weight parameters ($\boldsymbol{\alpha} \epsilon R^5$) for each action exemplar are the same across the two tested formats (F_1, F_2). This model describes units with the same SP across formats. In model 2, the weight parameters ($\boldsymbol{\alpha}, \boldsymbol{\gamma} \epsilon R^5$) are allowed to be different and enable distinct SPs for action exemplars between the two formats. In models 3 and 4, one format is assumed to be unmodulated by action identity, and a single scalar value ($c \epsilon R^1$) describes the presumably equivalent response (e.g., nonselective) to all actions within the format.

To determine how SPs compared across formats, we fit the parameters of each of the four models using standard linear regression techniques (see above), and the results were compared. Several measures are commonly used to select the "best" model from a set of candidate models. We used both the BIC and cvR^2 as our model selection criteria. These two methods provided slightly differing but reasonable notions of similarity. Heuristically, cvR^2 required a near-exact match, while BIC was found to have a more qualitative notion of similarity (see fig. S9). We viewed the two measures as providing something akin to upper and lower bounds on whether units were similar across formats and do not view either method as being "correct" per se. We used the arithmetic mean of the percentages of units fitting each model when reporting the results in the main figures of the paper.

This analysis was extended to three formats (Fig. 4A) requiring the creation and evaluation of 15 models for all the unique ways that the SPs can be expressed across formats. The 15 models are enumerated in fig. S11 (models 1 to 15) as the set of coefficients for each format where, following the description for pairwise comparisons, equivalent Greek letters indicate matched SPs across formats, a constant c indicates no selectivity to action exemplar for the associated format, and multiple Greek letters indicate significant but idiosyncratic SPs across formats. We performed this analysis for all combinations of three formats. Last, the analysis was extended to four formats requiring the creation and evaluation of 51 models for all the unique ways that the SPs can be expressed across formats. These models were constructed by enumerating all the possible configurations that the "fourth" format can take relative to the 15 models described above (see fig. S11). This analysis pooled data from subjects NS and EGS given the relative paucity of data for EGS. For Fig. 4C, we explicitly restricted the distribution of neurons to those that included matching selectivity between text and the visual formats. This includes the subset of visually selective neurons shown in blue in Fig. 4D.

Cross-format neuronal population analyses

Cross-decoding analyses (Fig. 2). We used two methods to measure the similarity of population responses across formats: cross-format classification accuracy and population correlation. Classification analyses were performed using LDA with assumptions and cross-validation procedures as described for within-format decoding above. For cross-format results, classifiers were trained within format and applied to alternate formats. More precisely, for each fold of the within-format cross-validation procedure, the classifier was

applied to the neural data associated with each of the three other formats. All predictions across folds of the cross-validation procedure were used to compute decode accuracy. This enables us to understand how well the neural representation of the different action exemplars generalize to a novel format when the definitions of the actions are preserved across the two formats. This approach further introduces directionality to the comparisons: e.g., how well do definitions established for the text format generalize to the visual formats and vice versa. To verify that the ability to generalize from one format to another required correctly aligning the action exemplars across formats, we repeated the analyses but now using "mismatched" labels. In the mismatch analyses, the action identity labels were swapped between action exemplars, and accuracy was recomputed on the basis of these reassigned labels. For the mismatched condition, we performed all possible shuffles for which no action exemplar was matched across formats ($N = 44$).

Cross-format population correlation analyses (Figs. 2, 5, and 6). To compute the population correlation measure, we organized the neural response data into four vectors, one for each format (fig. S8). Each vector had five values per unit, one value for each of the five actions. This value was computed as the mean firing rate recorded during the 2.6 s of stimulus presentation offset by 0.25 ms, averaged across trial repetitions. The mean response across these five values was subtracted, and the five values per unit were then concatenated across units to create a population response vector for each format. The same procedure was performed for each format ensuring that the same units and actions were aligned across the format vectors. The Pearson correlation was computed across format vectors to quantify the population-level similarity. Note that subtraction of the mean response across the five actions before concatenation was done to ensure that a positive correlation value across formats reflected similarity in the pattern of responses to the five actions and not general offsets in the mean response of the different neurons. This was necessary as some units were activated above baseline for all actions, and some were inhibited below baseline for all actions biasing the population toward a positive correlation that was not driven by the patterns of selectivity for the different actions. To ensure that a significant correlation was specifically the product of comparing the responses to the same actions across formats, we performed a shuffle control analysis. The population correlation was computed using the same procedures except that the five values computed per unit, one for each action identity, were misaligned (shuffled) between formats. The same shuffle order was applied to all units. All possible ways of shuffling action identities between formats (e.g., reordering five values) were tested, and the resulting shuffled correlations were averaged in reporting the results. The fiducial and shuffled correlations were computed separately for each session. Significant population correlation was determined on the basis of the P value resulting from a one-sided t test to determine whether the distribution of correlation values computed for each session was greater than 0. The correlation values for each session are also shown separately for each session with 95% confidence intervals computed using a bootstrap procedure (see Fig. 5).

The correlation analysis was also performed using a sliding-window approach to look at the time scale of positive correlation between formats (Fig. 5). The approach was similar as described with the following modifications: (i) Because we computed within-format correlation in addition to across-format correlations, we used a cross-validation approach for computing the correlation values.

The same procedure as described above was performed; however, the process began with splitting trial repetitions into training and testing sets (six trial repetitions each) and concluded with computing the correlation across the training and test splits. (ii) The training and test sets were computed from windowed data. Windows centered on time x were computed using a pseudo-Gaussian weighting function with mean = x and an SD of 200 ms. This allowed for a relatively smooth and temporally precise measure of the neural response. Correlations were computed between training and test sets for all combinations of windows starting from 500 ms before movie onset to 500 ms after movie offset.

We further asked whether the correlation between any two formats was mediated by the remaining two formats. For instance, the correlation between the text response and the frontal view could be the consequence of a text being correlated with the lateral view and the lateral view being correlated with the frontal view. To address this possibility, we performed a partial correlation analyses with the neural data from all four formats, thus looking at the correlation between two formats while regressing out the shared variance with the remaining formats.

Control analyses and tests

Understanding the effect of explicit visual imagery (Fig. 6). Can the text response be understood as the consequence of visual imagery, replaying the visual stimuli by imaging the visual sequence of events? In analyzing the visual formats, we found that the frontal and lateral views were encoded in a distinct manner. For instance, only roughly half the units had a matched SP across the frontal and lateral perspectives, while the remaining population of selective units had a distinct SP. If the neural responses depend on the contents of visual imagery, then visualizing from the frontal perspective should tend to activate the SPs for the frontal view, while visualizing from the lateral perspective should tend to activate the SPs from the lateral perspective. To understand the impact of visual imagery, we split the data based on the task instructions given to the subject before each session. In comparing the results of the model selection analysis, we compared the percentages of cells classified into each category (e.g., the percentage of cells with matched selectivity across all formats, a single format, etc.) using a chi-squared test.

In addition, we split the dynamic correlation results into sessions in which the participant was instructed to passively view (seven sessions) or actively imagine movements (six sessions) when presented with the action verb. We performed a paired t test between these two groupings at each pixel location to test for significant changes in correlation value as a function of task instruction (significance tested at the $P < 0.05$ level). The correlation values at the significant pixels were averaged and plotted to visualize the shape of temporal trends in correlation values.

Supplemental control task (fig. S10). We performed a sensory-motor association learning task to test whether repeated presentation of abstract stimuli, when paired with motor imagery of an action, would result in neural selectivity under passive viewing conditions. In the context of the current paper, this helps to constrain the interpretation of a shared neural substrate for action verbs and visually observed actions. We instructed the subject to use visual imagery to imagine finger movements when presented with fractal-like images of snowflakes. We used five images. Three of the images were associated with visual imagery of finger flexion movements of the thumb, index, and ring fingers. These movements were chosen as they resulted in especially robust neural selectivity in preliminary testing.

Two additional images were used as controls; the subject was instructed to passively view the stimuli without accompanying visualization. These two control images were used to test whether differential responses might emerge between the two images based on repeated exposure, even in the absence of any overt behavior on the part of the subject. We found that no such differential tuning emerged, and thus, for the current study, only one of the control images was used in the analysis.

The experiment began with a passive viewing session in which the subject viewed the stimuli before any motor association to test for baseline visual selectivity. Then, for the first 16 repetitions of each condition during the first two session days, the snowflakes were presented along with a key instructing which action should be performed (or no action in the case of the control images). For all other trial repetitions, the key was removed, and the subject performed reaction time and delayed imagined movements when presented with the visual stimuli. At varied intervals (fig. S10), we asked the subject to passively view the same set of stimuli. The experiment was performed twice, sequentially, with each experiment similar in structure but using a different set of visual stimuli. Experiment 1 consisted of 328 total repetitions per stimulus presented over 14 session days in a 49-day period. Experiment 2 consisted of 264 total repetitions per stimulus presented over 10 session days in a 56-day period.

We used a time-resolved classification analysis on the passive and reaction time trials separately to quantify selectivity for the cued stimuli. Classification methods were the same as described above (time-resolved classification of action exemplar) with windows beginning at −0.5 s and stepping to 2 s. For decode accuracy of the passive stimuli as a function of session number, we used the average firing rate within a 1-s window offset by 250 ms from stimulus onset.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES

1. J. R. Binder, R. H. Desai, The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
2. K. Meyer, A. Damasio, Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* **32**, 376–382 (2009).
3. M. A. Lambon Ralph, E. Jefferies, K. Patterson, T. T. Rogers, The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42–55 (2017).
4. Y. Miyashita, Perirhinal circuits for memory processing. *Nat. Rev. Neurosci.* **20**, 577–592 (2019).
5. F. Pulvermüller, How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* **17**, 458–470 (2013).
6. V. Gallese, G. Lakoff, The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* **22**, 455–479 (2005).
7. A. Caramazza, S. Anzellotti, L. Strnad, A. Lingnau, Embodied cognition and mirror neurons: A critical assessment. *Annu. Rev. Neurosci.* **37**, 1–15 (2014).
8. C. E. Watson, E. R. Cardillo, G. R. Ianni, A. Chatterjee, Action concepts in the brain: An activation likelihood estimation meta-analysis. *J. Cogn. Neurosci.* **25**, 1191–1205 (2013).
9. M. Ito, H. Tamura, I. Fujita, K. Tanaka, Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73**, 218–226 (1995).
10. V. Caggiano, L. Fogassi, G. Rizzolatti, J. K. Pomper, P. Thier, M. A. Giese, A. Casile, View-based encoding of actions in mirror neurons of area F5 in macaque premotor cortex. *Curr. Biol.* **21**, 144–148 (2011).
11. D. J. Freedman, J. A. Assad, Neuronal mechanisms of visual categorization: An abstract view on decision making. *Annu. Rev. Neurosci.* **39**, 129–147 (2016).
12. A. Lingnau, P. E. Downing, The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* **19**, 268–277 (2015).
13. M. Graziano, The organization of behavioral repertoire in motor cortex. *Annu. Rev. Neurosci.* **29**, 105–134 (2006).
14. R. A. Andersen, H. Cui, Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63**, 568–583 (2009).
15. N. N. Oosterhof, S. P. Tipper, P. E. Downing, Viewpoint (in)dependence of action representations: An MVPA study. *J. Cogn. Neurosci.* **24**, 975–989 (2012).
16. M. F. Wurm, A. Lingnau, Decoding actions at different levels of abstraction. *J. Neurosci.* **35**, 7727–7735 (2015).
17. G. A. Orban, S. Ferri, A. Platonov, The role of putative human anterior intraparietal sulcus area in observed manipulative action discrimination. *Brain Behav.* **9**, e01226 (2019).
18. M. Lanzilotto, C. G. Ferroni, A. Livi, M. Gerbella, M. Maranesi, E. Borra, L. Passarelli, M. Gamberini, L. Fogassi, L. Bonini, G. A. Orban, Anterior intraparietal area: A hub in the observed manipulative action network. *Cereb. Cortex* **29**, 1816–1833 (2019).
19. A. Raposo, H. E. Moss, E. A. Stamatakis, L. K. Tyler, Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia* **47**, 388–396 (2009).
20. M. Tettamanti, G. Buccino, M. C. Saccuman, V. Gallese, M. Danna, P. Scifo, F. Fazio, G. Rizzolatti, S. F. Cappa, D. Perani, Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* **17**, 273–281 (2005).
21. L. Aziz-Zadeh, S. M. Wilson, G. Rizzolatti, M. Iacoboni, Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* **16**, 1818–1823 (2006).
22. M. F. Wurm, A. Caramazza, Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nat. Commun.* **10**, 289 (2019).
23. S. Kalénine, L. J. Buxbaum, H. B. Coslett, Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. *Brain* **133**, 3269–3280 (2010).
24. D. Kemmerer, D. Rudrauf, K. Manzel, D. Tranel, Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* **48**, 826–848 (2012).
25. E. Binder, A. Dovern, M. D. Hesse, M. Ebke, H. Karbe, J. Saliger, G. R. Fink, P. H. Weiss, Lesion evidence for a human mirror neuron system. *Cortex* **90**, 125–137 (2017).
26. L. Chang, D. Y. Tsao, The code for facial identity in the primate brain. *Cell* **169**, 1013–1028. e14 (2017).
27. H. Hong, D. L. Yamins, N. J. Majaj, J. J. DiCarlo, Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* **19**, 613–622 (2016).
28. J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
29. W. Ma, R. M. Golinkoff, K. Hirsh-Pasek, C. McDonough, T. Tardif, Imageability predicts the age of acquisition of verbs in Chinese children. *J. Child Lang.* **36**, 405–423 (2009).
30. P. Cisek, J. F. Kalaska, Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).
31. T. Aflalo, S. Kellis, C. Klaes, B. Lee, Y. Shi, K. Pejsa, K. Shanfield, S. Hayes-Jackson, M. Aisen, C. Heck, C. Liu, R. A. Andersen, Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science* **348**, 906–910 (2015).
32. S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: High dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
33. C. Y. Zhang, T. Aflalo, B. Revechkis, E. R. Rosario, D. Ouellette, N. Pouratian, R. A. Andersen, Partially mixed selectivity in human posterior parietal association cortex. *Neuron* **95**, 697–708.e4 (2017).
34. L. W. Barsalou, Perceptual symbol systems. *Behav. Brain Sci.* **22**, 577–609; discussion 610-560 (1999).
35. N. Mainy, J. Jung, M. Baciu, P. Kahane, B. Schoendorff, L. Minotti, D. Hoffmann, O. Bertrand, J.-P. Lachaux, Cortical dynamics of word recognition. *Hum. Brain Mapp.* **29**, 1215–1230 (2008).
36. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
37. R. M. Golinkoff, K. Hirsh-Pasek, How toddlers begin to learn verbs. *Trends Cogn. Sci.* **12**, 397–403 (2008).
38. L. Song, S. M. Pruden, R. M. Golinkoff, K. Hirsh-Pasek, Prelinguistic foundations of verb learning: Infants discriminate and categorize dynamic human actions. *J. Exp. Child Psychol.* **151**, 77–95 (2016).
39. H. Konishi, A. E. Stahl, R. M. Golinkoff, K. Hirsh-Pasek, Individual differences in nonlinguistic event categorization predict later motion verb comprehension. *J. Exp. Child Psychol.* **151**, 18–32 (2016).
40. M. Imai, L. Li, E. Haryu, H. Okada, K. Hirsh-Pasek, R. M. Golinkoff, J. Shigematsu, Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Dev.* **79**, 979–1000 (2008).
41. J. Huttenlocher, P. Smiley, R. Charney, Emergence of action categories in the child: Evidence from verb meanings. *Psychol. Rev.* **90**, 72–93 (1983).

42. S. Ferri, G. Rizzolatti, G. A. Orban, The organization of the posterior parietal cortex devoted to upper limb actions: An fMRI study. *Hum. Brain Mapp.* **36**, 3845–3866 (2015).

43. A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, J. L. Gallant, Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).

44. M. Lanzilotto, M. Maranesi, A. Livi, C. G. Ferroni, G. A. Orban, L. Bonini, Stable readout of observed actions from format-dependent activity of monkey's anterior intraparietal neurons. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16596–16605 (2020).

Citation: T. Aflalo, C. Y. Zhang, E. R. Rosario, N. Pouratian, G. A. Orban, R. A. Andersen, A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. *Sci. Adv.* **6**, eabb3984 (2020).