1   Title:

2

3   **A shared neural substrate for action verbs and observed actions in human posterior**
4   **parietal cortex.**

5   Authors and affiliations:

6

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

8   [1]California Institute of Technology, Division of Biology and Biomedical Engineering, Pasadena,
9   CA, USA

10   [2] Tianqiao and Chrissy Chen Brain-Machine Interface Center, California Institute of Technology,
11   Pasadena, CA, USA

12   [3]Casa Colina Hospital and Centers for Healthcare, Pomona, CA, USA

13   [4]Department of Neurosurgery, University of California Los Angeles, Los Angeles, CA, USA

14   [5]Department of Medicine and Surgery, University of Parma, Parma Italy

15   *Correspondence to taflalo@caltech.edu

Abstract

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 the results are not the product of visual imagery nor 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 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 (*7*) and lesion results (*8-10*) have been mixed, and cannot establish a shared neural substrate at the level of single neurons (*11*). Thus how words get their meaning translates into two immediate questions with regard to single neuron selectivity: one, are words and their sensorimotor representations coded within the same region of cortex; and, two, 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 are 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 (*12-14*). 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

2

1   representation? To a particular canonical example? Or to the many diverse representations that comprise
2   our varied experiences? The question applies to all concrete words that describe physical reality,
3   including action verbs. In this study we look at how neural coding for action verbs relates to varied visual
4   representations of corresponding observed actions.

5   Finally, what cognitive phenomena can account for the presence of a link between a word and its visual
6   representation within any experimental paradigm? The link may mediate semantic memory, reflecting
7   associations between the word and its visual representations built over a lifetime of experience. In this
8   view, reading words activates sensorimotor representations automatically and these representations are an
9   intrinsic *component* of the meaning of the word. Second, reading a word has been hypothesized to evoke
10 mental imagery. Responses in sensorimotor cortex may reflect such imagery and the link could be
11 between visual representations and mental imagery of the same stimuli. Or the link may be the
12 consequence of short term learning such as occurs during categorization (*15*). Given these multiple
13 possibilities we address a fourth question: if a link exists, what cognitive process does the link mediate?

14 To address the above four questions, we recorded populations of neurons from electrode arrays implanted
15 in two tetraplegic individuals (NS and EGS) participating in a brain-machine interface (BMI) clinical trial
16 while the participants viewed videos of manipulative actions or silently read corresponding action verbs.
17 The implants were placed at the anterior portion of the intraparietal sulcus (see Fig S1 for implant
18 locations), a region that is part of the "action observation network" (AON) composed of the lateral
19 occipital temporal cortex (LOTC, (*16*)) as well as frontal and parietal motor planning circuits (*17, 18*).
20 These regions are involved in higher-order processing of observed actions (*19-22*) and neuroimaging and
21 lesion evidence implicate a role in verb processing (*23-29*). The ability to perform invasive neural
22 recordings provides us with the first opportunity to probe whether and how language links with
23 corresponding visual representations at the level of single neurons in high-order sensory-motor cortex.
24 Towards this objective we establish four primary results relating to the four questions outlined above:
25 First, PPC neurons show selectivity for action words and visually observed actions; Second, a portion of
26 PPC neurons link action verbs and corresponding visual representations; Third, text selective units in PPC
27 link with all the diverse visual representations found in the neural population; and, fourth, the link is not
28 based on imagery or short-term learning and thus appears to be semantic in nature. One possible
29 interpretation is that when reading text, we replay our visual history as part of the process of
30 understanding and thus ground our conceptual understanding in our unique experiences.

Results

32 Participants viewed videos of five manipulative actions presented in 3 visual formats (two lateral views
33 differing in body posture and one frontal view) and a fourth format, text, requiring the subject to silently
34 read associated action verbs (See Fig 1A for example stimuli). Five actions were used: drag, drop, grasp,
35 push and rotate, for which preliminary experiments (Fig S2) had demonstrated neuronal selectivity. A
36 total of 15 unique videos (five actions by three visual formats) and 5 written action verbs were presented.
37 Presenting the observed actions in three formats allowed us to tease apart different models of how action
38 verbs associate with overlapping (common to visual formats / exhibiting invariance across all formats)
39 and distinct (idiosyncratic to given formats / not invariant or only invariant across subset of formats)
40 features of the neural code for observed actions. This design allowed us to answer the first three questions
41 posed in the introduction. We recorded 1586 units during 18 recording sessions in two subjects. For the
42 first 7 sessions in subject NS, and all sessions for subject EGS, the participants passively watched the
43 action videos and silently read the action verbs. Results from silent reading and active imagery were

3

1 quantitatively similar in NS and thus data was pooled across sessions when addressing the first three
2 questions of this paper. To answer the fourth question, for the final 6 sessions in subject NS, the
3 participant used the action verb as a prompt to "replay" the associated action video using visual imagery
4 from either frontal (F) or lateral (L0) perspectives, thus allowing us to quantify how imagery affects verb
5 processing. Additionally, for question 4, we present a control study in which abstract symbols are paired
6 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?

8 Figure 1A shows the response of five representative neurons illustrating the variety of selectivity for both
9 observed actions and action verbs at the level of individual neurons. Within a format, we defined units as
10 selective if there was a significant differential response, measured as average firing rate over video
11 duration, to the different action identities. Significant selectivity across the recorded neural population
12 was found for action verbs, however, the percent of selective units as well as the consistency of the
13 response, as measured by the cross-validated coefficient of determination (cvR^2), was smaller for text
14 than for observed actions (Fig 1B, C). All five actions evoked significant neural responses from baseline
15 across the four formats (Fig S3). The majority of visually selective units were driven by the video
16 presentations, as in non-human primate anterior intraparietal area (AIP(*22*)). A minority, however, were
17 suppressed by the video and text presentations (Fig S3). The mean response strength decreased smoothly
18 from the action evoking the maximal response to the weakest response. Individual units could show steep
19 or more graded selectivity, and this pattern was essentially identical across formats (Fig S4). Greater
20 selectivity for action videos relative to text was reflected in a time-resolved decode analyses (Fig 1D).
21 Defining the latency of action selectivity as the onset of significant classification accuracy revealed
22 shorter latencies for the visual formats (windows starting at 60 -160ms depending on format) than the
23 written word (330ms), possibly reflecting differences in afferent pathways. Taken together, all formats
24 were encoded within the population, but with greater selectivity and shorter latency for videos relative to
25 text.

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

27 Having established that PPC neurons are selective for both action verbs and observed actions, we now ask
28 whether there exists a shared neural substrate, with neurons exhibiting matching selectivity for both a
29 word and the corresponding visual representation. We addressed this by using two population analyses:
30 across-format classification and across-format correlation. Leave-one out cross-validation was used to
31 train a classifier to predict action identity within-format. On each fold, the decoder was also used to
32 predict action identity from the three additional formats. This across-format generalization analysis
33 measures how well the neural population structure that defines action identity in one format generalizes to
34 other formats (Fig 2A, B). Across-format accuracy was above chance for all pairs of formats for NS and
35 above chance when pooling across visual formats to achieve adequate power for EGS (Rank-sum test,
36 $p<0.05$). In contrast, across-format accuracy was not significant when shuffling action identity between
37 formats (Shuffled Alignment, red in Fig 2A, B). This result demonstrates that the neuronal representation
38 was not random; the population is more likely to link representations across formats for the same action
39 identities. However, the results also demonstrate that the generalization is not perfect: the across-format
40 accuracy is lower than the within-format accuracy suggesting that the neural code for action identity also
41 depends on details of presentation. The strength of generalization was format dependent being near
42 perfect across body-postures (same lateral view), still high, but reduced across shifts in viewing
43 perspective (across the lateral and frontal views), and lowest when comparing observed actions with the
44 written verb.

1   Significant generalization of action representations across formats was robust to the analysis technique.
2   We correlated neural population responses across formats (Fig 2C, D). Population responses were
3   constructed by concatenating the mean response of all units to each action within-format (Fig S5). A
4   significant positive correlation was found for all format pairs while no significant positive correlation was
5   found when shuffling action identity between formats. One caveat to interpretation is that the correlation
6   between any pair of formats may be the consequence of the two formats being correlated with a third
7   format. A significant link between pairs of formats was preserved but somewhat reduced when
8   controlling for the other formats using a partial correlation analysis (Fig 2E&F). This last result indicates
9   that text links with each of the visual formats directly as the significant link is preserved when the
10 possible mediating factors of the other formats are removed.

11 The preceding population analyses established that text and visual representations are linked pairwise at
12 the level of the population, but the link does not perfectly generalize across formats. What is the
13 breakdown of the single units that compose the population results? To answer this question, we compared
14 the precise selectivity pattern (SP, defined as the firing rate values for each of the five actions) across
15 pairs of formats using a model selection analysis for each neuron. A linear tuning model can describe the
16 4 possible ways the SP can compare across two formats (Fig 3A). **(1)** Both formats are selective in a
17 similar manner (Fig 3A, Matched selectivity); the linear parameters ($\alpha \epsilon R^5$) for each of the five actions
18 are constrained to be identical for the two formats. **(2)** Both formats are selective, but with mismatched
19 patterns (Fig 3A, Mismatched selectivity) the linear parameters ($\alpha, \gamma \epsilon R^5$) are different between the two
20 formats. **(3, 4)** Finally, only one of the two formats may be selective (Fig 3A, Single format-1or format-2
21 selective); a constant scalar offset term is used for the non-selective format (not shown). We identified the
22 model that best described the neuronal behavior using both the Bayesian information criteria (BIC) and
23 the cross-validated coefficient of determination (cvR^2). We found that the two measures provide
24 complementary perspectives when comparing across formats (Fig S6.) In summarizing the results, we
25 used the average percentages provided by both measures. In line with our population results, we found
26 that the percent of cells with a similar selectivity patterns across formats (Fig. 3B&C, red) was format-
27 dependent, being greatest across body-postures (same lateral view), slightly reduced across shifts in
28 viewing perspective (across the lateral and frontal views), and lowest when comparing observed actions
29 with the written verb. These results not only indicate that text links with the visual formats and the visual
30 formats link with each other, but also that a percentage of the population codes the same action identities
31 in different formats with differing patterns of selectivity.

32

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

34 The preceding section demonstrated that there is a neural link between action verbs and visually observed
35 actions. Here we seek to understand the architecture of this link: to characterize how text-selective units
36 link with the varied visual presentations of the same action. As a prerequisite, we first characterized how
37 the different visual presentations were encoded with respect to each other, ignoring the text format. Just as
38 neural selectivity patterns can compare across two formats in four different ways (Fig 3A), they can
39 compare across 3 formats in 14 possible ways (see Fig. 4A, X-axis labels and examples). As above, a
40 model selection analysis was used to categorize each unit based on the model that best described the
41 selectivity patterns across the visual formats (Fig 4A). The population was heterogeneous, characterized
42 by units with matched SPs and mismatched SPs in varied combinations across the different visual
43 formats. This diversity can be seen in the individual unit examples of figure 1A; units 1 and 2 shows
44 matching patterns of selectivity across all the visual formats (Fig 4A, L0=L1=F), unit 3 shows matching

1 selectivity across two of the visual formats and no selectivity in the third (Fig 4A, L0=L1), and, unit 4
2 shows matching selectivity between two formats and mismatching selectivity in the third (L0=L1&F).
3 Thus we find that presentation details impact neural coding for action identity and that individual units
4 link action identity across formats in an assortment of ways when considering all three of the visual
5 formats at once. This result is consistent with the significant but incomplete generalization of action
6 identities across the visual formats shown in figures 2 and 3.

7 Having established that the same action is coded in different ways depending on details of visual
8 presentation, we can now look at how action verbs link to these varied visual representations. We can
9 frame our question in the following way: do action verbs link with the entire population of cells
10 demonstrating visual selectivity or specific subpopulations of cells? Figure 4B illustrates these
11 possibilities. Two primary theoretical possibilities in the literature describe how text can link with
12 subpopulations of visually selective neurons. *Overlapping* describes the architecture in which verbs link
13 specifically with the subpopulation of neurons that are invariant across the visual formats(*30*). *Exemplar*
14 describes the architecture in which verbs link with a specific prototypical exemplar, or "best example", of
15 the word (*30*). The exemplar may be of a single visual presentation or some subset of presentations.
16 Finally, we term the situation in which text links with all visually selective cells as *Available*. In this
17 architecture, the link between text and the visual representations mirror the statistics for how the visual
18 representations are encoded within the neural population independent of text. Underneath each schematic,
19 we provide a prediction for how the distribution of figure 4A should change when the smodel selection
20 analysis accounts for how text links with the visual formats.

21 We extended the model selection analysis to categorize each unit based on the model that best described
22 the selectivity patterns across all four formats (text + all visual formats.) We compared the distribution of
23 the visually selective units with a matched SP to text (Fig 4C) to the full distribution of the visually
24 selective units (Fig 4A). The distribution was essentially unchanged; the subset of visually selective units
25 that link with text reflects a random sampling of the visually selective units: a bootstrapped correlation
26 analysis comparing the empirical distribution of figure 4C with the predictions of 4B show that the
27 population best matches the Available model (correlation with invariant =0.32, exemplar=0.48,
28 available=0.97). This provides the answer to the question of architecture: the distribution of text-linked
29 units (Fig 4C) mirrors the statistics of how visual formats are encoded independent of text, or, in other
30 words, text forms links with all available visual representations. Units with a matching SP between text
31 and at least one visual format (the distribution of 4C,) represents 23% of all visually selective units (Fig.
32 4D) and 40% of all text selective units (Fig. 4E).

33

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

35 Does the link between text and the visual formats reflect a semantic association, visual imagery, or short-
36 term, learned associations that formed through the course of the experiment? Thus far, our analyses are
37 based on averaging the neural response across the video duration. This large temporal window may
38 encompass multiple cognitive processes. If neural processing for action verbs specifically reflects bottom-
39 up semantic processing, we would expect to find a shared neural response between formats very soon
40 after stimulus presentation. To address this issue we performed a dynamic, sliding-window cross-
41 validated correlation analysis to look at how the relationship within- and across formats evolves in time
42 (Fig. 5A, B.) To understand how quickly the correlation between text and the visual formats emerges, the
43 diagonal elements of the dynamic correlation matrices were extracted and plotted together for direct

1. comparison in the inset panels of figure 5A, B. These results show that the cross-modal link between text
2. and the visual formats is fast: the onset of the cross-format correlation between text and the visual formats
3. is the same as the within-format text correlation. In other words, as soon as a neural response to text
4. emerges, it immediately shares a common activation pattern with the observed actions.

5. Next, we checked whether the strength of population correlation changed over the course of the
6. experiment. If neural processing for action verbs reflects a semantic association, we would expect to find
7. the correlation between text and videos to be present from the first session throughout the course of the
8. experiment. In contrast, if the correlation between text and action videos is a product of learned
9. associations that developed over the course of the study, we would predict that the strength of correlation
10. would increase over the course of repeated exposure to the action videos and text. We found that the early
11. correlation response (cross-validated correlation over 1st second of video presentation) between text and
12. the three visual representations for each session, did not depend on session number (Figure 5C, D),
13. favoring the semantic interpretation.

14. We performed a number of control analyses and manipulations to address the possibility that associations
15. between text and observed actions reflect visual imagery. In six sessions participant NS was instructed to
16. use visual imagery to "replay" the associated action video in her mind from either the front (F) or side
17. (L0) perspectives when given the action verb prompt. If imagery were a dominant factor in establishing
18. the link between text and observed actions, the explicit manipulation of visualizing from the F or L0
19. perspective should bias the percent of cells with a matched SP in favor of F or L0. However, both the total
20. number of significant units and the population coding structure were essentially unaffected by the explicit
21. task instruction (Fig. 6 A, B). This result shows that the basic link between actions verbs and observed
22. actions is not dependent on the contents of visual imagery. To probe this result further and ensure the
23. subject followed task instructions, we split the dynamic correlation analysis between the passive and
24. active imagery sessions. Intriguingly, we found (Fig 6C, D) that the early correlation was largely
25. unaffected by the behavioral manipulation while the later (e.g. > 1sec) correlation did show significant
26. differences (paired t-test, $p<0.05$ on pixel values split between passive and imagery sessions), with active
27. imagery increasing the late period correlation strength relative to passive viewing (Fig. 6E). This result
28. suggests that the subject followed task instructions and that imagery can affect neural responses, but that
29. the early responses of neurons are independent of the contents of imagery.

30. In a final control, we collected a dataset in which four abstract symbols (snowflakes, Fig S7A) were
31. paired with visual imagery of movements for over 2 months (31 recording sessions, 114 ± 11 units per
32. session, Fig S7B). In this paradigm, subject NS was asked to visualize a movement from the first-person
33. perspective when presented with a symbol. The subject learned this task well, as we could accurately
34. decode the different symbols when the subject was instructed to use visual imagery (Fig S7C). We also
35. asked the subject to passively view the same stimuli at sporadic intervals (Fig S7B, vertical orange lines)
36. and found that the ability to decode the different symbols disappeared (Fig S7C, D). These clear
37. differences in decoding demonstrate that the subject is capable of comprehending and following task
38. instructions as they relate to passive viewing versus active visual imagery. Furthermore, the study shows
39. that not all types of visually distinct stimuli elicit a differential neural response under passive viewing.
40. Finally, it demonstrates that the recorded population does not form automatic neural responses to arbitrary
41. abstract symbols, even when the different symbols have been learned and are of direct behavioral
42. relevance.

Discussion

Our results answer the four questions raised in the introduction: PPC neurons exhibit selectivity for action verbs as well as observed actions; text links to visual representations of observed action; text links with a fraction of all available visual representations; and, the link is most consistent with being semantic in nature and not due to imagery nor 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 100ms), while text selectivity emerged nearly 200ms later. The features of the visual stimuli that determined neural selectivity remains 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, dynamics, amongst others, and 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 (see [34]). 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). Finally, 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 (*12, 31*).

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 (*23, 24, 26-28, 32*). However, imaging evidence can be inconsistent (*7*) and gross anatomical overlap seen in neuroimaging does not directly imply that the same neural populations support both tasks (*11*). Our evidence provides definitive evidence for a shared neural substrate by demonstrating that the precise selectivity patterns for action verbs match the selectivity patterns 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. Indeed, our results established that neural coding for action verbs depends upon the details of presentation (see Fig 4) and at least some degree of dependence upon presentation details appears to be ubiquitous throughout cortex (14). Given the diversity of neural coding, there are three likely architectures (Fig 5A), each with their 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

8

1   links to all available visual representations (e.g. Fig. 4B, "available".) If the visual representation reflects

2   the consolidation of one's experiential history with observed actions (*5, 33*) as expected for the

3   consolidation of semantic memory, then neural responses to text may be understood as the activation of

4   this consolidated visual experience. This suggests that a word's meaning is uniquely rooted in an

5   individuals' experience.

6   The comparison between the predictions for these 3 models and the data strongly favor the available

7   model. This architecture is also the easiest to implement, as a simple Hebbian mechanism will suffice and

8   would predict that acquisition of verb meaning depends on the frequency of exposure, which has indeed

9   been observed for several languages (*34*). The text response links only to a subset of the full distribution

10   of visually selective units (Fig 4D). The reason for this is unclear but may reflect inefficiencies in the

11   neural process that links verbs with visual representations and may be influenced by exposure or

12   experience. In any case, reading a word does not evoke the same perceptual experience as viewing an

13   action and thus substantial differences at the level of neural responses should be expected.

14   *Fourth: origin of the link.* Our results indicate that the link between text and the visual formats is not the

15   product of imagery or learned associations that emerge from the task. We found that experimental

16   manipulation of imagery has a minimal impact on the link between text and the visual formats. Further,

17   semantic processing should be fast and automatic, and we found that it is exactly the early component of

18   the correlation that was unaffected by the imagery manipulation (Fig 8D). In contrast, the late

19   components of the correlation systematically differentiated passive viewing and imagery sessions (Fig

20   6C-E), demonstrating that the patient followed the instructions, a view also supported by the control

21   experiment. The correlation between text and the visual formats was stable across all testing sessions (Fig

22   5). In our control experiment (Fig S7), passively viewing abstract symbols that had been paired to

23   movement imagery did not induce a neural response. Thus our results are unlikely the consequence of

24   learned associations. Yet another possibility is that neural responses to action verbs and observed actions

25   represent implicit automatic motor plans (*35*). Indeed, how we plan and execute an action is an important

26   component of meaning. However, our control study (Fig. S7) revealed no selectivity to passive

27   observation of movement-predictive cues.

28   Another possibility is that the linked selectivity patterns for observed actions and action verbs reflect a

29   population of cells that are responsive to the internal act of silent naming (i.e. generating the action verb).

30   In this view, when viewing text or videos, the participant covertly generates the same word, and thus

31   produces the same activity patterns. We think this hypothesis is unlikely. This hypothesis would predict

32   results similar to the invariant hypothesis, as generating the action verb should be consistent across the

33   different visual presentations. Instead, for simultaneously recorded neural populations, we find that text

34   responses link to the visual formats in idiosyncratic ways (e.g. text and the lateral views, but not the

35   frontal, Fig 1A, 4C). It remains possible that neurons are selective to particular cue-naming pairs,

36   however, in our prior work (*36*) in the same participants, we found no selectivity for specific cue-

37   intention pairings (e.g. response for imagined movement to the right when cued with a spatial target, but

38   not when cued with symbol.) Thus, we think the naming hypothesis is unlikely.

39   **Nature of neuronal representation of variables coded in PPC**

40   The ability of small neuronal populations to encode many variables is consistent with the mixed-

41   selectivity scheme in which distributed, non-linear, high-dimensional representations code in a

42   contextually dependent manner (*37, 38*). However, at least within the cortical locations explored in the

43   current study, we find that such encoding is not random, but systematically organized around stimulus

44   properties, a scheme referred to as partially-mixed selectivity (*39*). Indeed, neural populations coding the

9

1   same basic-level action type for different formats overlapped (e.g. Fig 3, Fig S6). Partial mixed selectivity
2   may represent a general structure for representing sensorimotor aspects of meaning within association
3   cortices, resulting in rich links between text and the diversity of overlapping and distinct components of
4   the visual formats that mirror the statistics of visual encoding independent of text (Fig 4). The partially-
5   mixed architecture may also account for the weak link between text and the visual formats (e.g. relatively
6   low population correlation and few units with matching SPs). If a cortical region encodes the many visual
7   facets of an observed action (e.g. viewpoint, posture, and other untested features) and text links with both
8   what is overlapping and distinct about action presentations, it follows that the link between text and any
9   particular presentation must be relatively weak.

10   **Cortical organization of conceptual knowledge.**

11   In understanding an action verb, we access semantic knowledge. The cortical organization of semantic
12   knowledge has been contentious. Some theories contend that conceptual knowledge is rooted in cortical
13   regions that employ supramodal symbolic processing (*11*) while other theories take the opposite
14   perspective, that semantic knowledge is encoded in the distributed sensorimotor network (*6, 40*). Most
15   recent theories posit that meaning emerges from interactions between supramodal associative areas and
16   regions directly responsible for processing sensory stimuli, motor actions, valence, and internal state (*1-*
17   *5*); Our results are consistent with these interactions models, given the longer latencies we observed for
18   text selective responses in PPC, relative to those reported for higher order language regions such as
19   superior temporal gyrus, or inferior frontal gyrus (*41*). This model comes in many versions, primarily
20   distinguished by which areas constitute the supramodal regions and the nature of the interactions. In part,
21   a deeper understanding of the organization of conceptual knowledge in the human brain has been limited
22   by the general inability to record from single neurons in humans. We know of no single unit recordings in
23   supramodal regions, but one intriguing possibility is that these areas may host neurons similar to the
24   "concept cells" of the medial temporal lobe (MTL, *42*), which respond to a preferred stimulus (e.g. a
25   particular individual) largely independently of sensory modality or presentation details (e.g. image,
26   written word, sound). While such strong invariance provides a model for neural coding mechanisms in
27   supramodal centers, much less is known about how semantically related items are encoded in the
28   distributed network. The current study contributes to this goal by providing the first demonstration of a
29   link between words and their sensorimotor representations and how the neural architecture supports this
30   link.

31   In the current paper, we have focused on how verbs are given meaning. We may also consider what our
32   results mean from the reverse direction, how the neural population may contribute to naming an observed
33   action. We find relatively high generalization across different views of the same observed action, but also
34   a degree of dependence on viewpoint and the point of fixation (Fig S2). These neural properties suggest
35   that rostral PPC neurons could play a role in creating increasingly abstracted representations that associate
36   the same actions and thus contribute to the processing needed for naming, but, given the weakness of the
37   link, that subsequent regions, potentially employing winner take all like mechanisms would be needed for
38   the final conversion to labeling the observed action.

39   The link between visual representations of actions and action verbs fits with current views of how infants
40   learn action verbs by mapping words onto conceptualizations of events (*43*). Infants can distinguish
41   action exemplars (running, marching, jumping) independently of the actors (*44*) and that this ability
42   predicts the use of action verbs at two years of age (*45*). Furthermore it provides an explanation for why
43   infants learn verbs later than nouns (*46*), as the corresponding visual representations are in different visual
44   pathways. In the PPC the development of observed action selectivity, which is originally in the service of

10

1 guiding future actions (*22*), may only occur once the infant starts moving. Indeed infants initially learn
2 verbs corresponding to their own actions (*47*).

3 **Limitations of the study**

4 *Stimuli*: We used a restricted set of observed actions and action verbs, based on the category of actions
5 that best evoke responses in neuroimaging (*48*). Recordings from the same arrays in the same subjects
6 have demonstrated selectivity for imagined reaches and saccades (*36*), simple shoulder shrugs and
7 vocalizations (*39*), as well as memory-strength and confidence (*49*). These diverse results point to
8 additional functions that may also be recruited when processing the meaning of language that will require
9 future investigations.

10 *Visual Formats*: We tested only a small number of visual formats: two postures and two viewpoints. Thus
11 the visual invariance that we established may be an overestimation, and increasing the diversity of
12 different presentations of the same action would lower the percentage of invariant cells. Hence, while it
13 remains possible that the visual invariant neurons (F=L0=L1 in fig 4) are akin to concept cells as
14 described in the MTL of humans, this is by no means established. To this point, neurons exhibiting
15 invariance in the MTL showed sparse coding (only active for a single basic level category) while the
16 invariant neurons tested in our study were broadly tuned, matching the tuning profiles of other visually
17 selective neurons (Fig S4). The small number of visual formats may also partially account for text
18 selective units with mismatched or absent visual selectivity (Fig. 4D, E) as they may link with other
19 untested visual representations of the corresponding action identity.

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

28 *Causality*. As with all passive neural recording studies, our study cannot determine the causal role of our
29 PPC neurons in understanding the meaning of action verbs. However, prior work has shown that damage
30 or inactivation within the fronto-parietal AON can result in specific action comprehension deficits (*24,*
31 *27-29, 32*) consistent with the idea that neurons within the AON play an important role in verb
32 comprehension. Our results provide clarity on the presence and nature of the link between neural
33 representations of action verbs and visually observed actions at the level of single units in PPC.

34 **Conclusion**

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

39

1

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 & 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 & EGS seated in their motorized wheel chair. Tasks were displayed on a 28 inch or 47-inch LCD monitor. The monitors were positioned so that the screen occupied approximately 25 degrees of visual angle. Stimulus presentation was controlled using the Psychophysics Toolbox for MATLAB.

Physiological Recordings. NS and EGS were implanted, each with one 96-channel Neuroport array on the gyrus dorsal to the junction of the intraparietal (IPS) and post-central (PCS) sulci (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 (<= 1.5mm) 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 are shown in supplementary figure 1.

Grasp related responses around the junction of the PCS and IPS have typically been described as the anterior IPS or putative human homologue of the anterior intraparietal area (phAIP) and is generally assumed to be the human homologue of macaque AIP. Macaque AIP is a region localized to the *lateral* bank of the anterior portion of the intraparietal sulcus involved in the visual control of grasping actions. However, the *medial* bank of the anterior intraparietal sulcus contains a distinct grasp field. This grasp field, described as PEip or Brodmann's area 5L (BA5L), is characterized by distinct frontoparietal connections (AIP is densely interconnected with PMv 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 homologue of AIP (2), the human homologue 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 homologues 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.

1 Neural activity was amplified, digitized, and recorded at 30KHz with the Neuroport neural signal
2 processor (NSP). The Neuroport System, comprising the arrays and NSP, has received FDA clearance for
3 <30 days acute recordings; for purposes of this study we received FDA IDE clearance (IDE #G120096,)
4 for extending the duration of the implant.

5 Single and multiunit activity was sorted using k-mediods clustering using the gap criteria to determine the
6 total number of neural clusters. Clustering was performed on the first n principal components where n
7 was selected to account for 95% of waveform variance (range of 2-4 components.) Sorting was reviewed
8 and adjusted if deemed necessary following standard practice by merging or splitting clusters as needed.

9 **Tasks and Stimuli.** Experimental stimuli consisted of five manipulative actions (drag, drop, grasp, push,
10 rotate) displayed in four different "formats": three visual video formats and one text format. In two visual
11 formats the actors were viewed from the side but the actor was either standing next to a table (Lateral 0,
12 L0) or sitting in a lotus position on the floor (Lateral 1, L1). In the third visual format, the actor was
13 standing next to the table, but was viewed from the front. Thus L1 differed from L0 only by body posture;
14 F from L0 only by viewpoint (note that two video cameras were used to simultaneously acquired videos
15 for F and L0 and thus the timing and kinematics of the movements are identical), and F from L1 both by
16 viewpoint and posture. In the text condition, the written action word (Arial at font size 80) was shown for
17 2.6s. Experimental stimuli of the visual formats consisted of video-clips (448x336 pixels, 50 frames/s)
18 showing one actor at a distance of 1.2 m performing five different hand actions (drag, drop, grasp, push,
19 rotate) directed towards an object (four versions per action-format). The objects were positioned directly
20 adjacent or within the hand such that actions predominately involved the wrist and fingers. All videos
21 measured 17.7° by 13.2° and lasted 2.6 s (the first 2 and the last 2 frames being static). The edges of the
22 videos were blurred with an elliptical mask (14.3°x9.6°), leaving the actor and the background of the
23 video unchanged, but blending it gradually and smoothly into the background around the edges. We did
24 not enforce fixation, instead asking the subject to view the actions in a naturalistic manner. The effects of
25 fixation were documented in a separate set of experimental sessions described below. Each data session
26 consisted of 12 repetitions of each unique video. Presentation was split into 3 runs (4 repetitions per run,
27 corresponding to the four versions). Videos were presented in a pseudorandom manner: all conditions
28 were randomly ordered and presented once before repetition. During baseline a highly blurred (full-width
29 half-max = 80 pixels) static frame (average of all video frames) was shown.

30 We collected this data under two different instruction sets. In the first instruction set the subject was
31 instructed to attend but otherwise passively view experimental stimuli. For the text format, we asked the
32 subject to read the word silently without any accompanying visualization. We collected 7 (subject NS)
33 and 5 (subject EGS) sessions in this passive viewing paradigm. In the second instruction set, the subject
34 NS (6 sessions) was instructed to use the text as a prompt to visualize the associated action being
35 performed from either the frontal (F, 3 sessions) or lateral standing perspective (L0, 3 sessions). Six
36 datasets in total were collected over 3 sessions. Each dataset was collected in full under the same
37 instruction. The datasets were collected over 3 experimental sessions in the following order (F, L0), (L0,
38 F), (F, L0) with () indicating that the datasets were acquired during the same session. Prior to each run,
39 the subject was reminded of the experimental condition. Following each run, the subject reported which
40 perspective she used to imagine the actions.

41 Information on the preliminary action observation task (Fig S2) to test for the presence of units selective
42 to observed actions is described in supplemental information. Additional information is found in
43 supplemental materials.

1   Statistical Analysis

2   **Within-format individual neuron analyses**

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

8   **Within-format one-way ANOVA for action identity (Fig. 1B).**
9   We defined a unit to be selective for action identity for a particular format if the unit displayed a
10   significant differential firing rate for the five action types during video presentation. Firing rate was taken
11   as the total spike count during movie presentation (2.6 seconds) starting at 0.25 seconds after onset
12   divided by the window duration. Firing rates were subjected to a one-way ANOVA with the factor of
13   action identity and significance was determined as $p<0.05$ after fdr correction. To ensure that selectivity
14   was driven by the task stimuli, we repeated the analyses using the 1 second window prior to stimulus
15   onset and found no neurons (0%) were selective in any format after fdr correction.

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

27   **Coverage (Fig S3).** We computed the p-value associated with the coefficient estimate of each action (e.g.
28   drag, drop, etc.) and found the percentage of coefficients associated with each action having a p-value less
29   than 0.05 after fdr correction. We used a bootstrap procedure for generating 95% confidence bounds on
30   the estimates of the percent selective.

31   We also wanted to know the frequency with which the different actions resulted in the peak response. To
32   test this, we first split the data into excitatory and inhibitory units. Inhibitory units were defined as units
33   for which the beta coefficient for all five actions was negative and thus suppressed relative to the baseline
34   response. We then identified the action that resulted in the largest deviation from baseline activity and
35   counted the number of units that showed a peak response for each action.

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

Selectivity curve analyses (Fig S4)

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 was 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)

We performed a 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 was performed using linear discriminate analysis with the following assumptions: one, the prior probability across the action exemplars was uniform, two, the conditional probability distribution of each unit on any given action exemplar was normal, three, only the mean firing rates differ for each action exemplar (the covariance of the normal distributions were the same for each action exemplar), and, four, 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 -.75 seconds relative to video onset with the final window chosen to be +3 seconds. 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 acquired within a single session. Mean and bootstrapped 95% confidence intervals were computed for each time bin from the cross-validated accuracy values computed for each session.

Latency analyses (Fig 1D). We used a method to identify the abrupt change in classification accuracy as the signal transitioned from chance decoding levels prior to stimulus onset to significant decoding. For each session, accuracy as a function of time was split into two regions (an early and late period) and the sum of the squared residuals (SSR) was computed after subtracting the mean. The time of transition between the early and late period was stepped through each time bin defined in the sliding window classification analyses. Latency is reported as the time of the abrupt change as defined by the transition time that resulted in the lowest SSR. This procedure resulted in 13 latency estimates for each format, (one for each of the 13 sessions in subject NS.) We used a 1x4 ANOVA to test for a main effect of format on latency followed by post-hoc testing between pairs of formats. Latency analyses was 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 2C).

1 **Cross format individual neuron analyses.**

2 **Comparing selectivity patterns between formats. (Fig 3, Fig 4, Fig S6)**

3 How a unit codes for action identity within a format is described by its selectivity pattern (SP) defined as
4 the precise firing rate values for all five action identities (see Fig 3A). At the single unit level,
5 understanding how a unit codes action idenity across formats can be quantified by comparing SPs across
6 formats. There are four possibilities when comparing SPs across a pair of formats: **one,** the SPs are
7 similar or matched across formats, **two,** both formats are selective but the SPs are distinct across formats,
8 **three**, only the first of the two formats is selective for action identity, and, **four**, only the second format is
9 selective for action identity. Each of these possibilities can be defined mathematically within a linear
10 models framework using four models of neural coding across formats:

11 $$model\ 1: fr = \boldsymbol{\alpha} \cdot \boldsymbol{F_1} + \boldsymbol{\alpha} \cdot \boldsymbol{F_2} + \beta$$

12 $$model\ 2: fr = \boldsymbol{\alpha} \cdot \boldsymbol{F_1} + \boldsymbol{\gamma} \cdot \boldsymbol{F_2} + \beta$$

13 $$model\ 3: fr = \boldsymbol{\alpha} \cdot \boldsymbol{F_1} + c \cdot \boldsymbol{F_2} + \beta$$

14 $$model\ 4: fr = c \cdot \boldsymbol{F_1} + \boldsymbol{\alpha} \cdot \boldsymbol{F_2} + \beta$$

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

21 To determine how SPs compared across formats we fit the parameters of each of the four models using
22 standard linear regression techniques (see above) and the results were compared. Several measures are
23 commonly used to select the "best" model from a set of candidate models. We used both the Bayesian
24 information criterion (BIC) and cross-validated R^2 (cvR^2) as our model selection criteria. These two
25 methods provided slightly differing but reasonable notions of similarity. Heuristically, cvR^2 required a
26 near exact match while BIC was found to have a more qualitative notion of similarity (See Fig S6). We
27 viewed the two measures as providing something akin to upper and lower bounds on whether units were
28 similar across formats, and do not view either method as being "correct" per se. We used the arithmetic
29 mean of the percentages of units fitting each model when reporting the results in the main figures of the
30 paper.

31 This analysis was extended to 3 formats (fig 4A) requiring the creation and evaluation of 15 models for
32 all the unique ways that the SPs can be expressed across formats. The 15 models are enumerated below
33 (Models 1-15) as the set of coefficients for each format where, following the description for pairwise
34 comparisons, equivalent Greek letters indicate matched selectivity patterns across formats, a constant c
35 indicates no selectivity to action exemplar for the associated format, and multiple Greek letters indicates
36 significant but idiosyncratic selectivity patterns across formats. We performed this analyses for all
37 combinations of three formats. Finally, the analysis was extended to 4 formats requiring the creation and
38 evaluation of 51 models for all the unique ways that the SPs can be expressed across formats. These
39 models were constructed by enumerating all the possible configurations that the "fourth" format can take
40 relative to the 15 models described above. This analyses pooled data from subject NS and EGS given the
41 relative paucity of data for EGS. The 15 possibilities (3 formats) are shown below:

1

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

2

3 For 4C we explicitly restricted the distribution of cells to those that matched specific selectivity patterns
4 between text and the visual formats.

5 **Cross format neuronal population analyses**

6 **Cross-decoding analyses (Fig 2)**
7 We used two methods to measure the similarity of population responses across formats: cross-format
8 decode accuracy and population correlation.

9 Classification analyses was performed using linear discriminate analysis with assumptions and cross-
10 validation procedures as described for within-format decoding above. For cross-format results, classifiers
11 were trained within format and applied to alternate formats. More precisely, for each fold of the within-
12 format cross-validation procedure, the classifier was applied to the neural data associated with each of the
13 three other formats. All predictions across folds of the cross-validation procedure were used to compute
14 decode accuracy. This enables us to understand how well the neural representation of the different action
15 exemplars generalize to a novel format when the definitions of the actions are preserved across the two
16 formats. This approach further introduces directionality to the comparisons: e.g. how well do definitions
17 established for the text format generalize to the visual formats and vice versa. To verify that the ability to
18 generalize from one format to another required correctly aligning the action exemplars across formats, we
19 repeated the analyses but now using "mismatched" labels. In the mismatch analyses, the action identity
20 labels were swapped between action exemplars and accuracy was recomputed based on these reassigned
21 labels. For the mismatched condition, we performed all possible shuffles for which no action exemplar
22 was matched across formats (N=44).

17

Cross-format *population* correlation analyses (Fig 2, Fig 5, Fig 6)

1 To compute the population correlation measure we organized the neural response data into 4 vectors, one
2 for each format (Fig S5.) Each vector had five values per unit, one value for each of the five actions. This
3 value was computed as the mean firing rate recorded during the 2.6 seconds of stimulus presentation
4 offset by 0.25 ms, averaged across trial repetitions. The mean response across these five values was
5 subtracted and the five values per unit were then concatenated across units to create a population response
6 vector for each format. The same procedure was performed for each format ensuring that the same units
7 and actions were aligned across the format vectors. The Pearson correlation was computed across format
8 vectors to quantify the population level similarity. Note that subtraction of the mean response across the
9 five actions prior to concatenation was done to ensure that a positive correlation value across formats
10 reflected similarity in the pattern of responses to the five actions and not general offsets in the mean
11 response of the different neurons. This was necessary as some units were activated above baseline for all
12 actions and some were inhibited below baseline for all actions biasing the population towards a positive
13 correlation that was not driven by the patterns of selectivity for the different actions. To ensure that a
14 significant correlation was specifically the product of comparing the responses to the same actions across
15 formats we performed a shuffle control analysis. The population correlation was computed using the same
16 procedures except that the five values computed per unit, one for each action identity, were misaligned
17 (shuffled) between formats. The same shuffle order was applied to all units. All possible ways of
18 shuffling action identities between formats (e.g. reordering 5 values) were tested and the resulting
19 shuffled correlations were averaged in reporting the results. The fiducial and shuffled correlations were
20 computed separately for each session. Significant population correlation was determined based on the p-
21 value resulting from a one-sided t-test to determine whether the distribution of correlation values
22 computed for each session was greater than 0. The correlation values for each session are also shown
23 separately for each session with 95% confidence intervals computed using a bootstrap procedure (see
24 figure 5).

25 The correlation analysis was also performed using a sliding window approach to look at the time scale of
26 positive correlation between formats (fig 5). The approach was similar as described with the following
27 modifications: (1) Because we computed within-format correlation in addition to across-format
28 correlations, we used a cross-validation approach for computing the correlation values. The same
29 procedure as described above was performed, however, the process began with splitting trial repetitions
30 into training and testing sets (6 trial repetitions each) and concluded with computing the correlation across
31 the training and test splits. (2) The training and test sets were computed from windowed data. Windows
32 centered on time x were computed using a pseudo-Gaussian weighting function with mean=x, and
33 standard deviation – 200ms. This allowed for a relatively smooth and temporally precise measure of the
34 neural response. Correlations were computed between training and test sets for all combinations of
35 windows starting from 500ms before movie onset to 500ms after movie offset.

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

43

44 **Control analyses and tests.**

18

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 view were encoded in a distinct manner. For instance, only roughly half the units had a matched selectivity pattern across the frontal and lateral perspectives while the remaining population of selective units had a distinct selectivity pattern. If the neural responses depend on the contents of visual imagery, then visualizing from the frontal perspective should tend to activate the selectivity patterns for the frontal view while visualizing from the lateral perspective should tend to activate the selectivity patterns from the lateral perspective. To understand the impact of visual imagery we split the data based on the task instructions given to the subject prior to 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 (7 sessions) or actively imagine movements (6 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 S7)

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 prior to any motor association to test for base-line 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 8), 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 type) with windows beginning at -.5 seconds and stepping to 2 seconds.

19

1   For decode accuracy of the passive stimuli as a function of session number, we used the average firing

2   rate within a 1 second window offset by 250 ms from stimulus onset.

3

1

Supplemental materials:

3

4   Figure S1: Functional magnetic resonance imaging (fMRI) localization of implant sites.

5

6   Figure S2: Preliminary study; impact of multiple factors on neural responses to observed manipulative actions.

7

8   Figure S3: Percent of selective units for each action for each of the four formats.

9

10   Figure S4: Shape of tuning functions for each format.

11

12   Figure S5: Organization of data for population correlation method.

13

14   Figure S6: Comparison of model selection procedures.

15

16   Figure S7: Abstract symbols systematically associated with motor imagery does not evoke differential neural
17   tuning.

18

References

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. *Nature Reviews Neuroscience* **18**, 42-55 (2017).

4. K. Patterson, P. J. Nestor, T. T. Rogers, Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* **8**, 976-987 (2007).

5. Y. Miyashita, Perirhinal circuits for memory processing. *Nat Rev Neurosci* **20**, 577-592 (2019).

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. 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).

8. U. Halsband, J. Schmitt, M. Weyers, F. Binkofski, G. Grutzner, H. J. Freund, Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: a perspective on apraxia. *Neuropsychologia* **39**, 200-216 (2001).

9. G. A. Negri, R. I. Rumiati, A. Zadini, M. Ukmar, B. Z. Mahon, A. Caramazza, What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cogn Neuropsychol* **24**, 795-816 (2007).

10. C. Rosci, V. Chiesa, M. Laiacona, E. Capitani, Apraxia is not associated to a disproportionate naming impairment for manipulable objects. *Brain and Cognition* **53**, 412-415 (2003).

11. A. Caramazza, S. Anzellotti, L. Strnad, A. Lingnau, Embodied cognition and mirror neurons: a critical assessment. *Annu Rev Neurosci* **37**, 1-15 (2014).

12. 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).

13. 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).

14. 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).

15. D. J. Freedman, J. A. Assad, Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annu Rev Neurosci* **39**, 129-147 (2016).

16. A. Lingnau, P. E. Downing, The lateral occipitotemporal cortex in action. *Trends Cogn Sci* **19**, 268-277 (2015).

17. M. Graziano, The organization of behavioral repertoire in motor cortex. *Annu Rev Neurosci* **29**, 105-134 (2006).

18. R. A. Andersen, H. Cui, Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63**, 568-583 (2009).

19. 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).

20. M. F. Wurm, A. Lingnau, Decoding actions at different levels of abstraction. *J Neurosci* **35**, 7727-7735 (2015).

21. 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).

22. 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).

23. 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).

24. 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).

25. 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).

26. M. F. Wurm, A. Caramazza, Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nat Commun* **10**, 289 (2019).

27. F. Pulvermuller, O. Hauk, V. V. Nikulin, R. J. Ilmoniemi, Functional links between motor and language systems. *Eur J Neurosci* **21**, 793-797 (2005).

28. S. Kalenine, L. J. Buxbaum, H. B. Coslett, Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. *Brain* **133**, 3269-3280 (2010).

29. L. Papeo, A. Lingnau, S. Agosta, A. Pascual-Leone, L. Battelli, A. Caramazza, The origin of word-related motor activity. *Cereb Cortex* **25**, 1668-1675 (2015).

30. F. Pulvermuller, How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn Sci* **17**, 458-470 (2013).

31. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013-1028 e1014 (2017).

32. 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).

23

33. 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).

34. 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).

35. P. Cisek, J. F. Kalaska, Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* **33**, 269-298 (2010).

36. 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, Neurophysiology. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science* **348**, 906-910 (2015).

37. S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol* **37**, 66-74 (2016).

38. D. Zipser, R. A. Andersen, A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679-684 (1988).

39. 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 e694 (2017).

40. L. W. Barsalou, Perceptual symbol systems. *Behav Brain Sci* **22**, 577-609; discussion 610-560 (1999).

41. 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).

42. 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).

43. R. M. Golinkoff, K. Hirsh-Pasek, How toddlers begin to learn verbs. *Trends Cogn Sci* **12**, 397-403 (2008).

44. 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).

45. 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).

46. 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).

47. J. Huttenlocher, P. Smiley, R. Charney, Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review* **90**, 72-93 (1983).

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

3   49.   U. Rutishauser, T. Aflalo, E. R. Rosario, N. Pouratian, R. A. Andersen, Single-Neuron
4   Representation of Memory Strength and Recognition Confidence in Left Human Posterior
5   Parietal Cortex. *Neuron* **97**, 209-220 e203 (2018).

6

7

1

Acknowledgments

13

Declaration of Interests

The authors declare no competing interests

Figures:
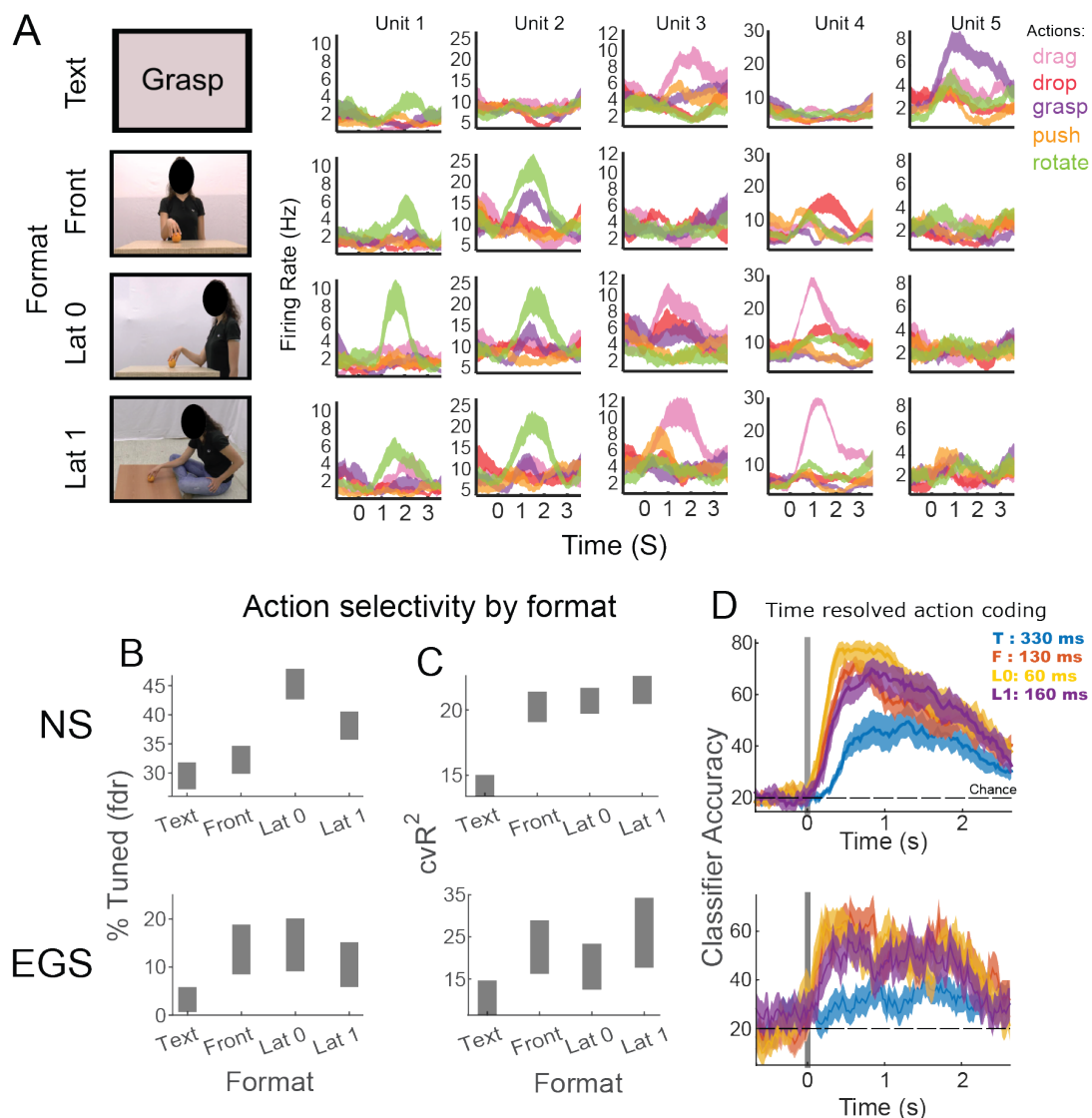
Figure 1: Human parietal neurons are selective for observed actions and action verbs. **(A)**. Example neurons illustrating diverse selectivity patterns across formats. *Left panel*: Sample still frames depicting "grasp" in each format. Face is masked to obscure identity per publisher's request. *Right panel*: Representative units illustrating diverse neural responses to the five tested actions (color coded) across the four tested formats. Each panel shows the firing rate (mean ± sem) through time for each action for a single format. Each column illustrates the responses of the same unit to the four formats. See figure S1 for recording locations. **(B)** Percent of units with significant action selectivity split by format (mean ± 95% CI, one-way ANOVA, p<0.05 fdr corrected.) Zero units were selective in each format during the 1 second window prior to 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 (mean ± 95%CI). **(D)** Sliding-window within-format classification accuracy for manipulative actions. Sliding window= Overlapping 300ms windows with 10ms increments. Inset displays color code for format and associated latency estimate for onset of significant decoding.
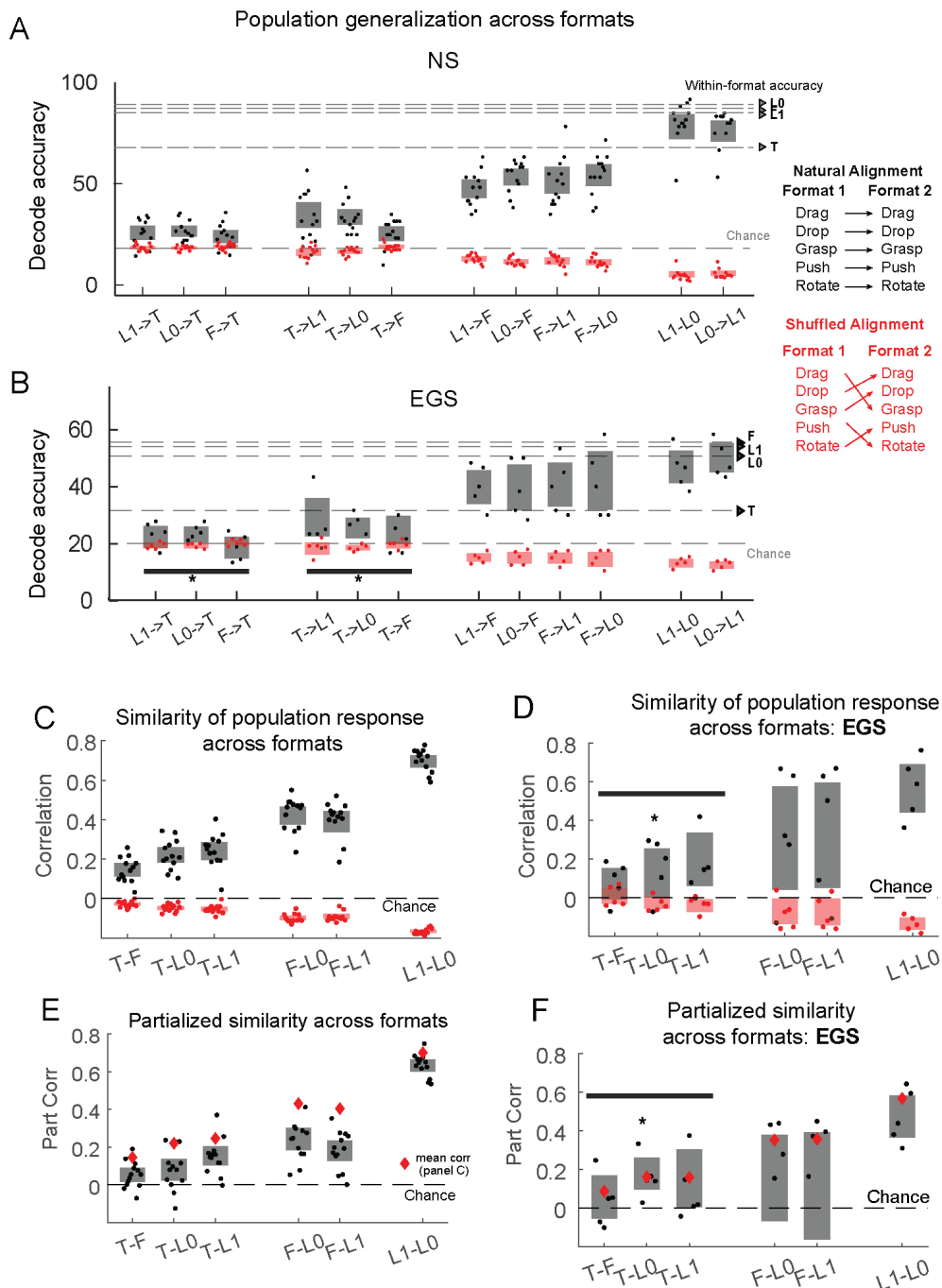
27

Figure 2: Action verbs link with observed actions. (A) Across-format classification of manipulative actions. X-axis labels indicate the formats used for classifier training and testing (e.g. train->test.) Dots = single session result. Rectangle = 95% bootstrapped CI. Grey (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) reached significance (Wilcoxson 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 D for participant EGS. **(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 F for participant EGS
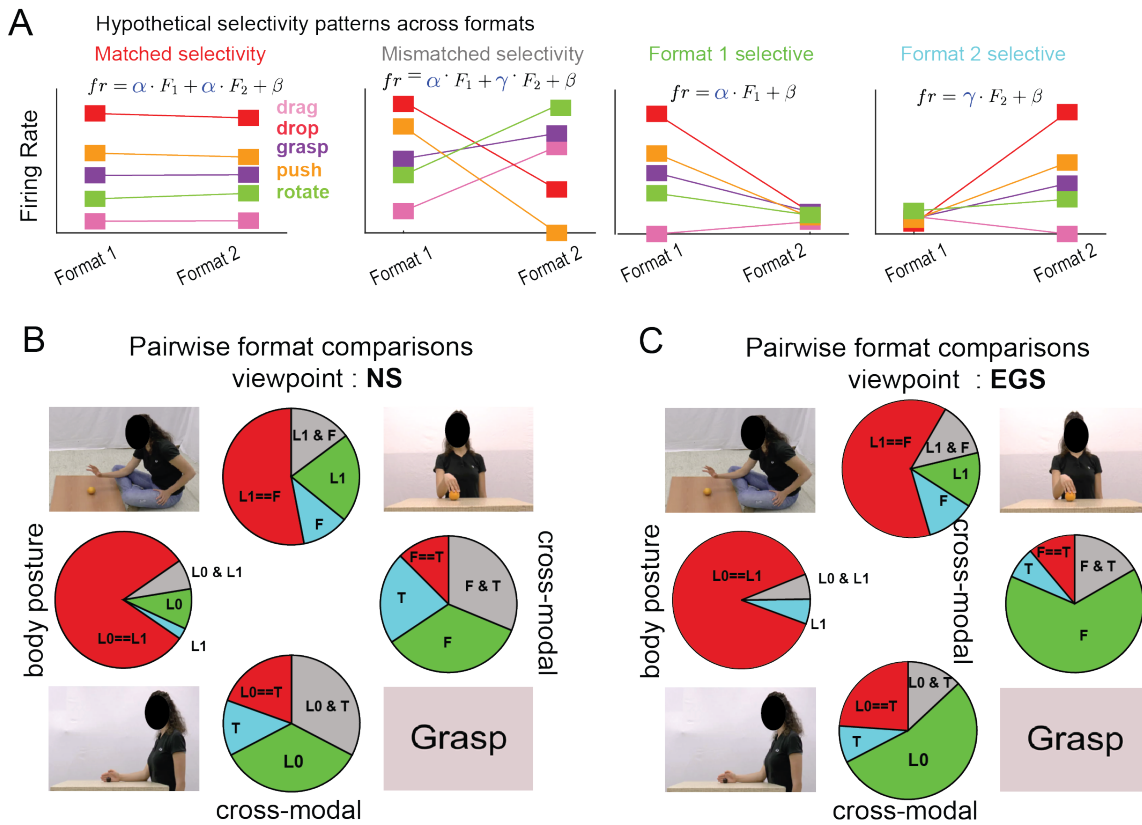
Figure 3: **Single neuron selectivity patterns link action verbs and observed actions (A)** Schematic illustrating the four possible ways the selectivity pattern (SP) can compare across two formats (see Fig S6 for expanded description). **(B)** Summary of SPs across pairs of formats for participant NS (see Fig S6). Red = matched SP, grey = mismatched SP, cyan, light green= selectivity for a single format only (see title colors in panel A). Face is masked to obscure identity per publisher's request. **(C)** Same as panel B for participant EGS.
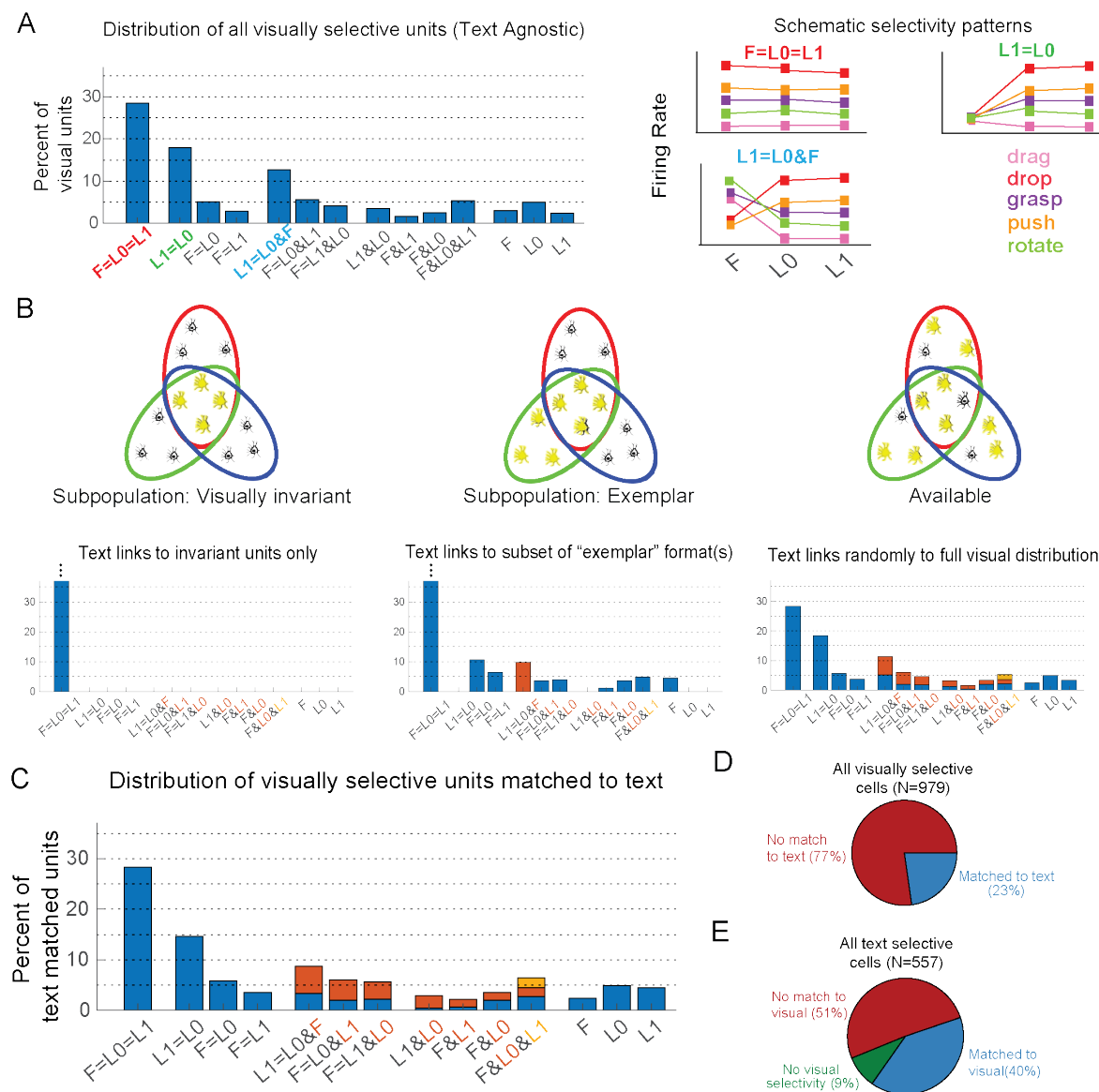
Figure 4: Text links with all available visually selective cells. (A) Histogram characterizing how the population of neurons link action representations across the three visual formats (F,L0,L1). "=" matched selectivity pattern (SP), "&" mismatched SP. Exclusion of a format indicates no selectivity. Three schematic selectivity patterns (right panel, 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 3 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 2 of the three ovals) but is simplified here for schematic purposes. Yellow neurons are selective for text with matching selectivity while grey neurons are not. Underneath each schematic, is a prediction for how the distribution in panel 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 panel 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.
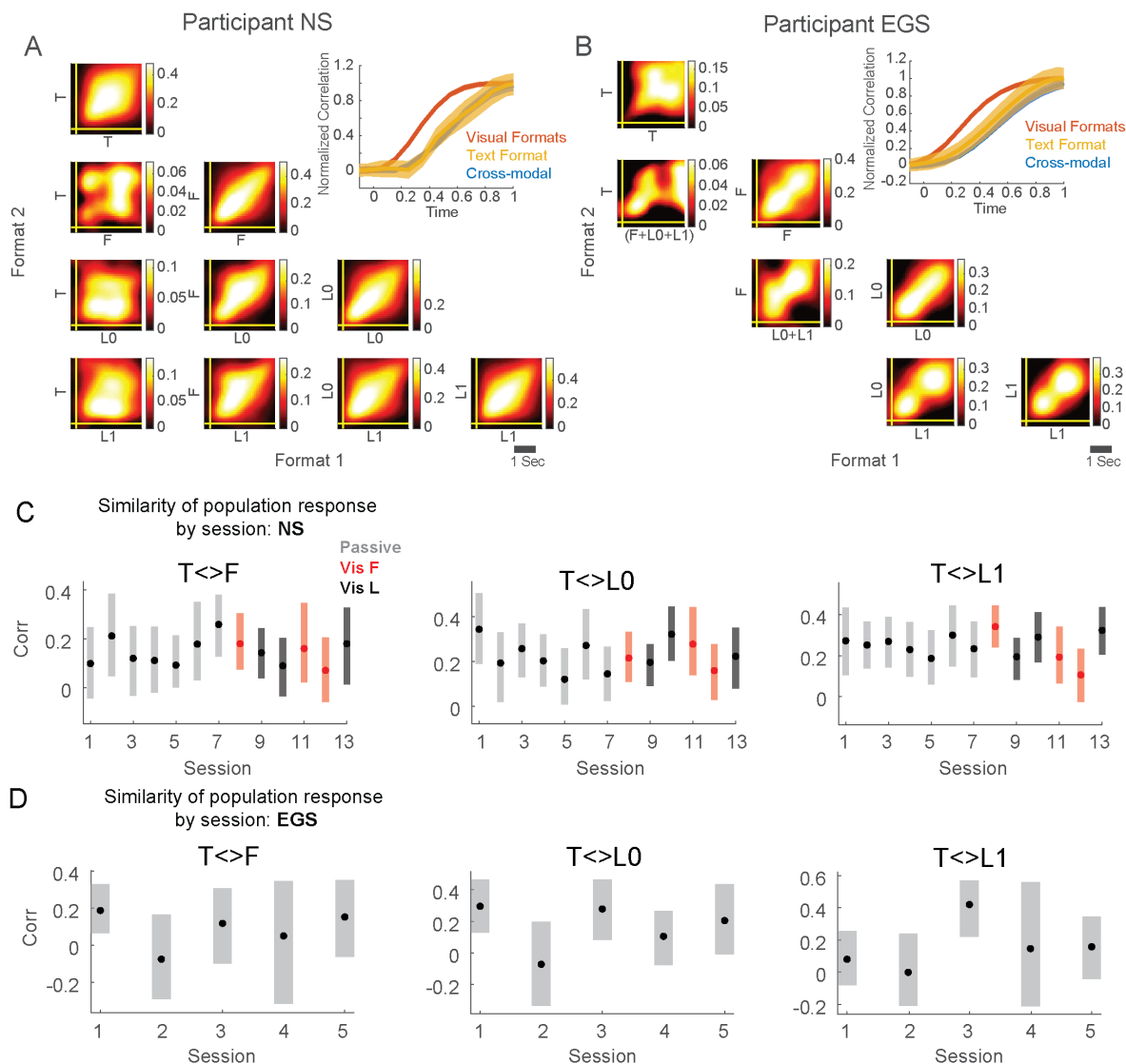
Figure 5: Temporal features support a semantic link between verbs and observed actions. (A-B) Cross-modal match between text and visual formats occurs at low latency. **(A)** Dynamic cross-validated cross-correlation matrices demonstrating how the neural population response during stimulus presentation 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 color bar. Inset: The diagonal elements of the within- and across format matrices were averaged into three logical groupings (1,within- format visual, 2,within-format text, and 3,across-format text to visual) and normalized to a peak amplitude of 1 for comparison purposes. The temporal profile of the averaged correlations (mean±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,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 panel A except for participant EGS (only silent reading).
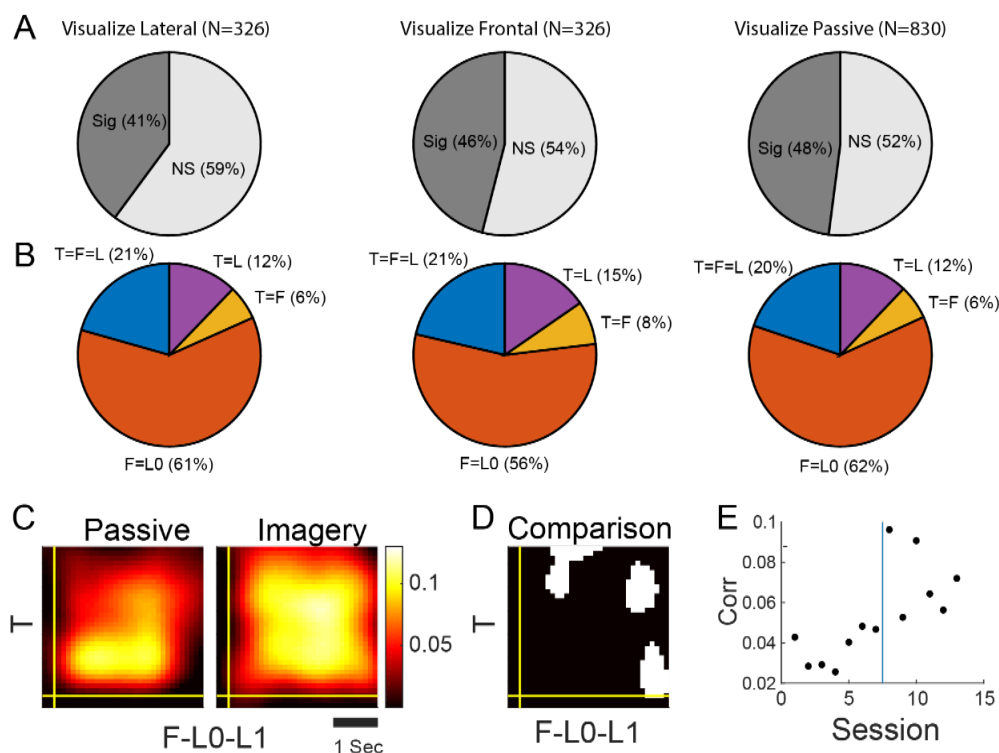
Figure 6: The effect of explicit instruction on cross-format invariance. During the initial seven sessions subject NS silently read action verbs. In the six subsequent runs she explicitly visualized the frontal (F, 3 runs) or lateral standing (L0, 3 runs) perspective in response to the action verb. **(A)** The percent 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+cvR2) 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. **(D)** Pixel coordinates demonstrating a significant difference between passive viewing and active imagery (significant pixels in white, paired t-test, $p<0.05$.) **(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. Blue line shows split between passive and active imagery sessions

1